

VICTOR HUGO CUNHA DE MELO

Orientador: David Menotti

**UMA METODOLOGIA PARA AVALIAÇÃO
DE MÉTODOS DE CONTAGEM DE PESSOAS BASEADA
EM ANÁLISE DE VÍDEO**

Ouro Preto
Dezembro de 2011

UNIVERSIDADE FEDERAL DE OURO PRETO
INSTITUTO DE CIÊNCIAS EXATAS
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

**UMA METODOLOGIA PARA AVALIAÇÃO
DE MÉTODOS DE CONTAGEM DE PESSOAS BASEADA
EM ANÁLISE DE VÍDEO**

Monografia apresentada ao Curso de Bacharelado em Ciência da Computação da Universidade Federal de Ouro Preto como requisito parcial para a obtenção do grau de Bacharel em Ciência da Computação.

VICTOR HUGO CUNHA DE MELO

Ouro Preto
Dezembro de 2011



UNIVERSIDADE FEDERAL DE OURO PRETO

FOLHA DE APROVAÇÃO

Uma Metodologia para Avaliação
de Métodos de Contagem de Pessoas baseada em Análise de Vídeo

VICTOR HUGO CUNHA DE MELO

Monografia defendida e aprovada pela banca examinadora constituída por:

Dr. DAVID MENOTTI – Orientador
Universidade Federal de Ouro Preto

Dr. GUILLERMO CÁMARA CHÁVEZ
Universidade Federal de Ouro Preto

Dr. JOSÉ MARIA RIBEIRO NEVES
Universidade Federal de Ouro Preto

Dra. ANDREA IABRUDI TAVARES
Universidade Federal de Ouro Preto

Ouro Preto, Dezembro de 2011

Resumo

Contagem de pessoas baseada em análise de vídeo pode ser muito útil para diversas aplicações comerciais, como o monitoramento de espaços públicos ou eventos desportivos. No entanto, os métodos presentes na literatura, em geral, apenas verificam se a contagem total é correta, independente do momento em que cada contagem acontece. Neste trabalho, é proposto uma metodologia de avaliação de métodos de contagem de pessoas baseado em câmera de vídeo na posição zenital. Inicialmente, é necessário indicar manualmente em um dado vídeo quando cada pessoa entra ou sai da zona de contagem, gerando os dados de referência. A partir destes dados de referência e a saída de um método de contagem de pessoas, obtém-se o acoplamento entre as pessoas rastreadas de ambos. Este problema é modelado como um grafo bipartido completo e utiliza-se o Algoritmo Húngaro para estabelecer um acoplamento maximal. Após obtenção do acoplamento, os vértices não saturados indicam os falsos positivos e negativos da contagem do método, enquanto os vértices saturados indicam os verdadeiros positivos. Destes números, podem ser calculadas medidas padrões como *precisão*, *revocação* e *F-score*. Usando esta metodologia, é possível quantificar automaticamente a contagem de falsos positivos e negativos dos métodos e identificar a quantidade de pessoas que havia em uma cena quando um erro ocorreu. Além disso, seu uso em um método de contagem de pessoas traz benefícios referente ao ajuste de parâmetros e na comparação entre diferentes métodos de contagem de pessoas.

Abstract

People counting based on video analysis may be very useful for many commercial applications, such as in the monitoring of public spaces or in sporting events. However, the methods in the literature tend to only check if the total counting is correct, independent of where each count happens. In this work, it is proposed a methodology for the assessment of people counting methods based on video from cameras in a zenith position. Initially, it is required to manually indicate, in a given video, when each person passes into and out of the counting zone, generating the ground-truth data. From this reference data and the output of a people counting method, it is obtained the matching of the best tracked people from the reference to the output counting. For this, the Matching Problem is modeled as a complete bipartite graph $K_{n,n}$ and it is used the Hungarian Algorithm to establish a maximal matching with maximum cost. Once the matching is performed, the non-saturated vertices, *i.e.*, people, indicate false positive and negative counting of the method, while the saturated vertices indicate the true positive counting. From these figures, standard measures such as *precision*, *recall* and *F-score* can be automatically computed and may also identify where errors occur. In addition, the use of this methodology on a people counting method brings benefits for comparison purposes and adjusting parameters.

*Aos meus pais Porfírio e Ana Vera,
ao meu irmão Caio Hess,
e à minha companheira Mariana.*

Agradecimentos

A conclusão deste trabalho se deve a colaboração e intervenção de muitas pessoas. Primeiramente, eu agradeço a Deus pela força e perseverança que recebi ao longo deste curso.

Agradeço especialmente ao professor, orientador e grande amigo David Menotti, pela oportunidade de trabalhar ao seu lado. Também por não me deixar desistir e pelos puxões de orelha quando foram necessários.

A todos os professores do DECOM pela imensa contribuição ao meu conhecimento, e pelos trabalhos e provas dos quais eu tanto reclamei, mas hoje agradeço profundamente. Em especial, o professor Haroldo Gambini, me prestando grande auxílio no desenvolvimento deste trabalho, com o empréstimo de livros sobre grafos e sanando com grande paciência minhas dúvidas.

Aos meus pais e maiores exemplos Porfírio e Ana Vera, obrigado pelo amor incondicional e por sempre me incentivarem. Principalmente quando ligava para vocês falando que iria chegar só na véspera de Natal, haha.

Ao meu irmão Caio Hess, grande amigo e companheiro para todos os momentos, principalmente quando a barra está pesada. Obrigado pelas conversas edificantes.

À minha namorada Mariana, obrigado por me dedicar suas orações, mantras e energia positiva. Agradeço profundamente pelos belos momentos e pelo seu carinho e atenção.

Aos meus amigos que participaram desta jornada, os irmãos da República Molotov, Jóia Rara e do DECOM. Obrigado pelas madrugadas viradas em conjunto, por todo este tempo que passamos juntos e pelo que vocês me ensinaram.

À Adriana da Seção de Ensino da Escola de Minas e ao Pró-Reitor Adjunto de Graduação — e ilustre torcedor do América-MG — Adilson Pereira pela celeridade nos processos e imensa boa vontade.

Obrigado!

Sumário

1	Introdução	1
1.1	Definição do Problema	2
1.2	Objetivos	3
1.3	Organização do Texto	4
2	Trabalhos Relacionados	6
2.1	Metodologia de Avaliação de Resultados	6
2.2	Bases de Vídeo e Características	7
2.3	Considerações Finais	8
3	Fundamentação Teórica	9
3.1	Conceitos de Grafos	9
3.2	Acoplamentos	10
3.3	Acoplamentos em Grafos Bipartidos	10
4	A Metodologia de Avaliação	11
4.1	Geração da Referência	11
4.2	Modelagem do Problema	12
4.3	Resolução do Problema	15
4.4	O Algoritmo Húngaro	15
4.4.1	Discussão da Segunda Fase	17
4.4.2	Simulação do Algoritmo	17
4.5	Etapa Pós-Acoplamento	19
5	Experimentos	21
5.1	O Método de Contagem de Pessoas	21
5.1.1	Subtração do Fundo	21
5.1.2	Segmentação	23
5.1.3	Rastreamento	25
5.1.4	Parâmetros Ajustados	27

5.2	Propriedades dos Vídeos	27
5.3	Medidas	29
5.4	Resultados	29
5.5	Considerações Finais	31
6	Conclusões	33
6.1	Trabalhos Futuros	33
	Referências Bibliográficas	38

Lista de Figuras

1.1	As três categorias principais de métodos de contagem. Fontes: http://www.arcat.com/ e Silva (2008)	2
3.1	Exemplo de matriz de biadjacência	10
4.1	Grafo bipartido representando o Problema de Acoplamento. Cada aresta possui um peso baseado na Equação 4.1. Arestas mais escuras possuem peso maior; arestas mais claras possuem menor peso.	14
4.2	Matriz de biadjacência e sua representação em forma de grafo, antes da execução do algoritmo	17
4.3	Primeira fase do algoritmo	18
4.4	Segunda fase do algoritmo	18
4.5	Resultado da execução do algoritmo	19
5.1	Fluxograma para representação do sistema	22
5.2	Resultados demonstrativos das etapas de subtração do fundo e segmentação de pessoas. (a) quadro original. (b) Subtração do fundo e segmentação de pessoas.	24
5.3	Após segmentação pelo <i>k-means</i>	24
5.4	Exemplo de posicionamento das linhas auxiliares para contagem de pessoas	25
5.5	Exemplo de rotas identificadas pelo método de contagem. Apenas a rota em azul expandiu o suficiente para ser contada	26
5.6	Posicionamento da câmera no corredor do DECOM/ICEB - UFOP.	27
5.7	Pessoas juntas e um guarda-chuva na região de interesse no vídeo stm1	28
5.8	Pequena variação de luminosidade na região inferior da imagem, encontrada nos vídeos stm2 e stm3	28
5.9	Exemplo 2: segmentação errônea. Número real de pessoas na região de interesse: 1; Número de pessoas identificadas pelo método: 2.	32

Lista de Tabelas

4.1	Exemplo de geração da referência	12
4.2	Exemplo de possível saída de um método de contagem de pessoas	13
4.3	Exemplo de vértices da tabela de referência da Tabela 4.1 usada no Problema de Acoplamento	14
4.4	Resultado do acoplamento estabelecido entre as Tabelas 4.1 e 4.2	19
4.5	Número esperado de pessoas em cada quadro	20
5.1	Valor dos parâmetros utilizados	27
5.2	Número total de pessoas que atravessa a região de interesse em cada vídeo	28
5.3	A variação do parâmetro D_{min} e seu impacto sobre a avaliação dos resultados para diferentes vídeos	30
5.4	Número diferente de pessoas na região de interesse e número de TP, FN e FP usando $D_{min} = 50$ para o método implementado	31

Lista de Algoritmos

6.1	Algoritmo Húngaro	36
6.2	Procedimento auxiliar AUGMENT	37

Capítulo 1

Introdução

Detecção, rastreamento e contagem de pessoas podem ser de grande utilidade para diversas aplicações comerciais, como o monitoramento de espaços públicos ou eventos desportivos. Por exemplo, estações de metrô em megalópoles tem um tráfego intenso de pessoas por dia, podendo utilizar tais sistemas para medir o seu fluxo. As informações coletadas a partir do processo de contagem também auxiliam a identificar padrões de tráfego de veículos e a monitorar o público em eventos. Além disso, sistemas de vigilância podem recorrer a estes métodos para atribuir o número exato de pessoas em lugares chave de segurança, e planejar modos de evacuação eficiente.

Os métodos presentes na literatura podem ser divididos em três grandes categorias: **Métodos baseados em Contadores Mecânicos**, **Métodos baseados em Sensores** e **Métodos baseados em Visão Computacional utilizando Câmeras** (Velipasalar et al., 2006).

Nos métodos baseados em contadores mecânicos, destaca-se a catraca como seu representante (Figura 1.1a). São largamente utilizadas nos pontos de acesso de metrôs, estádios e outros ambientes com controle de entrada, mas não possibilitam a contagem de múltiplas pessoas simultaneamente. Outra desvantagem é que podem obstruir o caminho, causando congestionamento se houver um tráfego intenso de pessoas. Devido ao seu formato, são suscetíveis a efetuar subcontagens – quando as pessoas não respeitam o obstáculo saltando ou passando por baixo.

A segunda categoria é composta pelos métodos baseados em sensores, tais como os sensores de calor e o raio infravermelho (Figura 1.1b). Apesar de não obstruírem o caminho como ocorre com os métodos da primeira categoria, também estão sujeitos ao problema de subcontagem, mas devido a uma causa diferente – a sobreposição de pessoas. Também podem errar a contagem por impossibilitarem a distinção entre uma pessoa e um objeto qualquer que passa na zona de contagem.

A última categoria é composta pelos métodos baseados em visão computacional utilizando câmeras, que superam os inconvenientes da primeira categoria e permitem solucionar os problemas da segunda (Velipasalar et al., 2006; Chen et al., 2008; Huang e Chow, 2003). Com base

em um vídeo gravado por uma câmera que monitora uma dada região de interesse, utiliza-se técnicas de Processamento Digital de Imagens, Reconhecimento de Padrões e Visão por Computadores para efetuar a segmentação¹, o rastreamento e, enfim, a contagem.

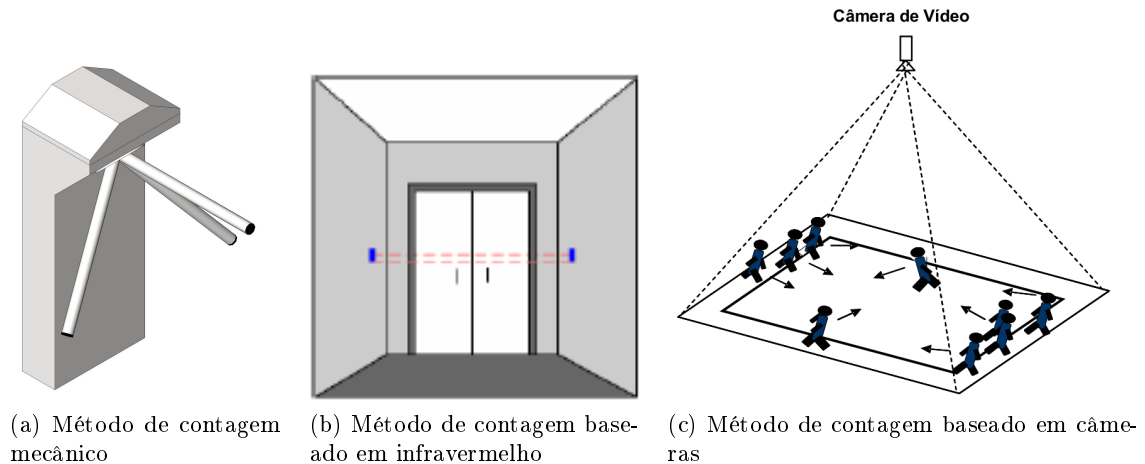


Figura 1.1: As três categorias principais de métodos de contagem. Fontes: <http://www.arcat.com/> e Silva (2008)

A primeira questão que necessita ser solucionada nestes métodos é o posicionamento da câmera. Alguns trabalhos, como visto em Kilambi et al. (2008); Elik et al. (2006), utilizam a câmera em **posição oblíqua** que, embora permita a detecção de mais características, apresenta problemas com relação a oclusões e a privacidade dos indivíduos.

O **posicionamento zenital**, por outro lado, consiste em uma câmera sobre as pessoas, rotacionada azimutalmente em 180 graus (ver Figura 1.1c). Esta posição contorna o problema de oclusão entre objetos, além de oferecer vantagens adicionais, tais como não invasão a privacidade dos indivíduos e tamanho relativamente constante dos objetos em cena (Velipasalar et al., 2006; Bozzoli e Cinque, 2007). Tendo essas vantagens em vista, este trabalho volta-se para o desenvolvimento de soluções envolvendo métodos de contagem de pessoas com câmera em posicionamento zenital.

1.1 Definição do Problema

Um dos principais problemas observado nos métodos de contagem de pessoas com câmeras em posicionamento zenital consiste na metodologia de avaliação empregada nos trabalhos publicados. Muitos autores utilizam apenas uma única medida, como a acurácia, para avaliar seus resultados. A utilização de uma única medida possui a desvantagem de não permitir o discernimento entre alarmes falsos e pessoas não reconhecidas.

¹A segmentação subdivide uma imagem em suas regiões ou objetos constituintes. O nível de subdivisão depende do problema sendo resolvido. Deste modo, a segmentação deve parar quando os objetos de interesse em uma aplicação são isolados (Gonzalez et al., 2009).

Por outro lado, o uso de um conjunto de medidas, como a *precisão*, *revocação* e *F-score*, provê uma avaliação mais detalhada. Todavia, o processo de contabilizar manualmente os falsos positivos (FP), falsos negativos (FN) e verdadeiros positivos (TP) – valores necessários para o cálculo dessas medidas – pode ser exaustivo e impraticável para vídeos maiores.

O estudo dos métodos de contagem de pessoas enfrenta outros problemas, também presente em vários outros trabalhos nesta era digital, como detalhado pelo *Digital Age Committee* (2009). Segundo este comitê, as tecnologias digitais permitiram a expansão da capacidade e do alcance da pesquisa, porém levantaram novos problemas complexos. Alguns deles incluem:

- (i) As complicações para assegurar a validade dos dados de uma determinada pesquisa;
- (ii) Normas que não acompanham a alta taxa de inovação;
- (iii) Restrições sobre o compartilhamento dos dados que reduzem a capacidade dos pesquisadores de verificar os resultados; e
- (iv) Além de uma enorme geração de novos dados, criando sérios desafios na preservação destes para uso a longo prazo.

O relatório recomenda que todos os pesquisadores recebam treinamento apropriado na gestão de dados de pesquisa e convida os pesquisadores a tornarem todos os métodos, resultados e outras informações subjacentes publicamente acessíveis em tempo hábil.

Grande parte dos autores dos métodos de visão computacional seguem parcialmente ou não seguem as recomendações do relatório, dificultando a análise com relação à reprodução de experiências e a verificação dos resultados. As informações não descritas consistem, principalmente, nos parâmetros de configuração dos métodos e características das gravações, como altura da câmera ou condições do ambiente durante gravação. Sobretudo, poucos autores disponibilizam os vídeos usados em seus trabalhos.

Diante das dificuldades apresentadas, constata-se a falta de uma metodologia de avaliação, automática e acompanhada de bases de vídeos de referência, que ofereça ferramentas para comparação confiável entre métodos de contagem de pessoas em posição zenital e permita o desenvolvimento de pesquisas futuras.

1.2 Objetivos

Neste trabalho, é proposto e desenvolvido uma nova **Metodologia para Avaliação Automática** dos métodos de contagem de pessoas. A metodologia é descrita resumidamente em quatro etapas:

1. Geração manual dos dados de referência para um determinado vídeo apenas uma única vez. São produzidos indicando quando uma pessoa entra ou sai da região de interesse;

2. Modelagem do problema como um grafo bipartido, cujas arestas representam, *grosso modo*, o grau de similaridade entre as pessoas rastreadas pelo método e as pessoas rastreadas pela referência;
3. Estabelecimento do acoplamento maximal entre os vértices;
4. Determinação automática dos falsos positivos, falsos negativos e verdadeiros positivos. Por meio desses, é possível o cálculo de medidas padrão como a *precisão*, *revocação* e *F-score*.

As principais vantagens desta metodologia automática são:

- Quantificar automaticamente a contagem de falsos positivos, negativos e verdadeiros positivos dos métodos;
- Permite uma comparação efetiva entre diferentes métodos de contagem de pessoas utilizando apenas sua saída e a referência;
- Identificar a quantidade de pessoas que havia em uma cena quando um erro de contagem ocorreu;
- De acordo com as informações fornecidas pela metodologia automática de avaliação, é possível retornar aos métodos para o reajuste de seus parâmetros e melhorar seu desempenho segundo alguma medida.

Com o objetivo de validar a metodologia de avaliação proposta neste trabalho, o método proposto por Antic et al. (2009) é implementado e executado em três vídeos de 10 minutos, gravados nos corredores do Departamento de Computação da Universidade Federal de Ouro Preto. A saída dessas execuções são avaliadas pela metodologia proposta. Em seguida, o resultado é avaliado segundo as medidas definidas em Antic et al. (2009) a partir do resultado da metodologia.

Observe que este trabalho segue as recomendações do *Digital Age Committee* (2009), compartilhando os vídeos utilizados, os dados de referência, os parâmetros de configuração utilizados na execução do método de Antic et al. (2009) e a metodologia de testes.

1.3 Organização do Texto

Esta monografia está organizada em seis capítulos, descritos a seguir:

No **Capítulo 1**, é apresentada uma breve introdução sobre métodos de contagem de pessoas utilizando câmeras em posição zenital e os principais problemas com relação ao modo em que seus resultados são avaliados pela literatura.

No **Capítulo 2**, são mostrados os trabalhos relacionados às metodologias de avaliação em métodos de contagem de pessoas.

No **Capítulo 3**, é descrita a fundamentação teórica sobre o problema de acoplamento.

No **Capítulo 4**, é decrita a metodologia de avaliação proposta, a geração da referência e a modelagem do problema.

No **Capítulo 5**, são apresentados a análise dos resultados experimentais, as medidas utilizadas e o método de contagem de pessoas implementado e avaliado.

Finalmente, no **Capítulo 6** são apontadas as conclusões e os trabalhos futuros.

Capítulo 2

Trabalhos Relacionados

Este capítulo apresenta uma revisão bibliográfica sobre os principais trabalhos relacionados ao uso de metodologias de avaliação para análise dos métodos de contagem de pessoas. Na Seção 2.1 são discutidas tais metodologias, inclusive as medidas que permitem definir. Já a Seção 2.2 relata os trabalhos que disponibilizaram suas bases de vídeo publicamente. Como poucos o fizeram, foram analisadas as características mencionadas pelos autores de cada vídeo, no intuito de oferecer características similares nos vídeos do presente trabalho.

2.1 Metodologia de Avaliação de Resultados

Nos trabalhos anteriores de métodos de contagem de pessoas, uma das principais formas de avaliação utilizada é a **acurácia**. Pode ser definida como a proporção entre o número de pessoas que o algoritmo contou com relação ao número real de pessoas no vídeo. Os autores dos seguintes trabalhos (Septian et al., 2006; Hsieh et al., 2007; Yu et al., 2007; Bozzoli e Cinque, 2007) e outros utilizam apenas esta única medida para avaliação de seus resultados. Para determiná-la, não é levada em consideração os valores de FP, FN e TP.

Assim, mesmo que seja utilizada na maioria dos artigos da área, apenas a acurácia – como qualquer outra medida utilizada individualmente – não permite determinar se os elementos contados são pessoas reais. Deste modo, é possível que um método contabilize apenas ruídos ao invés de pessoas e obtenha elevada acurácia. São necessárias outras medidas, utilizadas em conjunto a esta, que permitam assegurar que seus valores sejam pessoas verdadeiras e, deste modo, comparações mais precisas possam ser estabelecidas.

Além de avaliarem seus resultados segundo a acurácia, Velipasalar et al. (2006) incluem outras medidas na avaliação de seu método de contagem de pessoas: analisam a direção de movimento e o número de **uniões** e **separações** ocorridas no vídeo.

Outros autores (Xu et al., 2007; Bescos et al., 2003) contabilizam o número de falsos negativos e positivos detectados por seu método em conjunto a acurácia. Baseando-se nestes valores, Barandiaran et al. (2008) utilizam a matriz de confusão para estimar a *precisão* e a

revocação de um sistema. Por meio da combinação de ambas utilizando a Média Harmônica Ponderada, também propõem o uso da medida *F-score*. Antic et al. (2009) estabelecem uma comparação com seu método e o proposto por Barandiaran et al. (2008), e avaliam os resultados utilizando as já mencionadas *precisão*, *revocação* e *F-score*.

Já Chen et al. (2009) retornam ao uso da acurácia e apresentam seus resultados analisados com relação aos detalhes de velocidade e direção de movimento das pessoas, separados pelo número de pessoas na cena. Todavia, estes valores não são gerados de forma automática pelo método, e sim por contagem manual.

Os autores de Mukherjee et al. (2011) avaliam seus resultados segundo a acurácia, *precisão*, *revocação* e *F-score*. Benabbas et al. (2010) apresentam seus resultados baseados unicamente na acurácia, porém afirmam que podem ser adaptados tanto para vídeos adquiridos em posição zenital quanto oblíqua. Nenhum destes autores introduzem uma forma de calcular os valores de FP, FN e TP automaticamente.

É possível constatar que, até então, não houve propostas anteriores na literatura de contagem de pessoas para efetuar a avaliação dos métodos de contagem automaticamente ou que possibilitem identificar a quantidade de pessoas na região de interesse quando um determinado alarme falso ocorreu.

2.2 Bases de Vídeo e Características

Nos trabalhos anteriores de contagem de pessoas, poucos autores disponibilizaram suas bases de vídeo adquirida em posição zenital. Dos trabalhos analisados, apenas Antic et al. (2009); Chen et al. (2009) os compartilharam publicamente com ressalvas. Os vídeos do primeiro possuem marcações com o resultado da contagem de seu método, impossibilitando o seu uso, enquanto os vídeos do segundo não estão mais na página citada pelo artigo.

A menção mais recente sobre uma base de vídeos pública são dos autores Li et al. (2011), que citam sua importância, desenvolvem uma e a disponibilizam. Porém, os vídeos foram adquiridos em posição oblíqua, evidenciando a necessidade de uma base de vídeos pública na posição zenital.

No intuito de gerar uma base própria para este trabalho e disponibilizá-la publicamente, foram analisadas as principais características dos vídeos, mencionadas pelos seus autores. Velipasalar et al. (2006) escrevem que suas filmagens contém situações adversas envolvendo as pessoas do vídeo e incluem pessoas de mãos dadas, abraçadas ou unidas. Xu et al. (2007) informam que seus vídeos possuem 1 hora de duração e relatam a presença de muito ruído, baixo contraste e movimentos anormais dos pedestres.

Os autores Hsieh et al. (2007) escrevem que realizaram seus testes em 28 sequências de vídeo, todavia não relataram nenhum detalhe de suas gravações. Yu et al. (2007) informam que seus vídeos foram adquiridos durante o dia e a noite.

Bozzoli e Cinque (2007) relatam que utilizaram dois vídeos longos, um com 1,58 horas de duração, e o outro com 46 minutos. O primeiro vídeo foi adquirido à noite e o segundo durante a manhã para tratar diferentes condições de iluminação.

Yu et al. (2008) utilizaram vídeos próprios encontrados na *internet*. Usaram situações de movimento em uma única direção e em duas direções. Dentre as situações adversas, são listadas pessoas juntas e de mãos dadas, e pessoas que andam rápido e devagar.

2.3 Considerações Finais

Os trabalhos relacionados apresentados neste capítulo mostram que não houve, anteriormente, proposições para se efetuar a avaliação dos métodos de contagem de pessoas automaticamente ou que possibilitem identificar a quantidade de pessoas na região de interesse quando um determinado alarme falso ocorreu. Além disso, nota-se a necessidade de uma base de vídeos pública, adquirida em posição zenital. Nos capítulos seguintes, serão descritas a proposta da metodologia de avaliação automática de métodos de contagem de pessoas e a base de vídeos adquirida para execução dos testes.

Capítulo 3

Fundamentação Teórica

A metodologia de avaliação apresentada neste trabalho propõe, em sua etapa final, a correspondência automática entre os dados de referência com os dados obtidos por um método de contagem de pessoas. Este problema pode ser caracterizado como um **Problema de Acomodamento** em **Teoria dos Grafos** (Boaventura Netto, 2006). Neste capítulo é apresentado a fundamentação teórica para o problema.

3.1 Conceitos de Grafos

Um **grafo** não orientado $G = (V, E)$ consiste de um conjunto finito V não vazio e um conjunto E , em que cada elemento $e \in E$ é definido em função dos elementos $v \in V$. Os elementos $v \in V$ são chamados **vértices** e os elementos $e \in E$ são denominados **arestas**. Dois vértices que participam de uma ligação são ditos **adjacentes** (Jungnickel, 2007).

Se o conjunto V de vértices de um grafo $G = (V, E)$ é particionado em dois subconjuntos disjuntos X e Y tal que cada aresta em E é estabelecida entre um vértice em X e outro vértice em Y , é denominado **grafo bipartido** e é denotado por $G = (X \cup Y, E)$ (Balakrishnan, 1995).

Teorema 1 *Um grafo com três ou mais vértices é bipartido se e somente se ele não possui ciclos ímpares.*

Um grafo bipartido $G = (V \cup W, E)$ é completo se há uma aresta entre cada vértice $v \in V$ e $w \in W$. Um grafo completo com n vértices é denotado por K_n . Se $G = (V \cup W, E)$ é um **grafo bipartido completo** com m vértices em V e n vértices em W , então G é indicado por $K_{m,n}$ (Balakrishnan, 1995).

Uma das formas de se representar um grafo bipartido $G = (V \cup W, E)$ é por meio de uma **matriz de biadjacência**. Consiste de uma submatriz (V, W) de valores, em que cada linha corresponde a um vértice de V e cada coluna a um vértice de W (Figura 3.1).

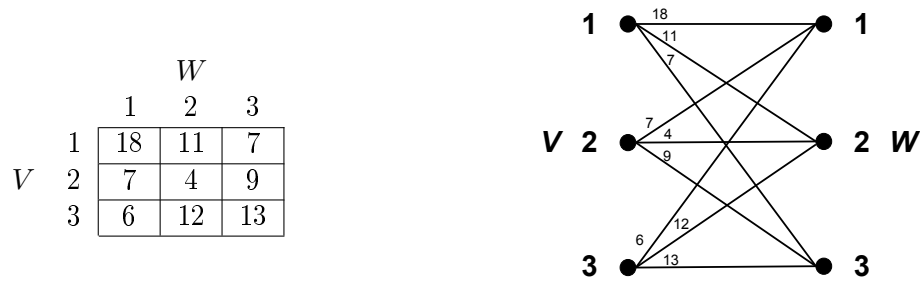


Figura 3.1: Exemplo de matriz de biadjacência

3.2 Acoplamentos

Segundo Balakrishnan (1995), o Problema de Acoplamento pode ser definido como:

Definição 1 Um **acoplamento** em um grafo não-direcionado $G = (V, E)$ é um conjunto M de arestas tal que duas arestas em M não possuem um vértice em comum.

O problema de encontrar o acoplamento em um dado grafo com o maior número de arestas possível é conhecido como o **Problema de Cardinalidade Máxima**. Se cada aresta do grafo recebe um peso, o **Problema de Acoplamento Valorado** – do inglês *Weighted Matching Problem* – tem-se o problema de encontrar um acoplamento tal que a soma dos pesos de todas as arestas em um acoplamento seja ótima. A seguir, a Definição 2 apresenta um problema correlacionado.

Definição 2 Em um grafo $G = (V, E)$, um conjunto $P \subset V$ é chamado uma **cobertura de vértices** se todo $e \in E$ tiver ao menos uma extremidade em P .

O Problema da Cobertura de Vértices trata de um conceito paralelo ao de dominância, porém se refere as arestas. Em um conjunto dominante, para um grafo não-orientado, cada vértice externo deve ter ao menos uma aresta em comum com um vértice do dominante.

3.3 Acoplamentos em Grafos Bipartidos

O subproblema de Acoplamentos em Grafos Bipartidos, chamado de **Problema de Alocação** (*assignment problem*), é um caso particular dentre os problemas lineares. Pode ser entendido como o da busca de um acoplamento de cardinalidade máxima e valor ótimo em um grafo bipartido valorado (Boaventura Netto, 2006).

Capítulo 4

A Metodologia de Avaliação

Como referido anteriormente, a metodologia proposta permite aferir automaticamente os resultados de um método de contagem de pessoas – doravante denominado método alvo – em relação a dados de referência gerados manualmente. Ela pode ser dividida nas etapas seguintes:

1. Geração dos dados de referência;
2. Modelagem do problema como um grafo bipartido completo $K_{n,n}$, em que os vértices $v \in V$ representam as pessoas rastreadas pela referência e os vértices $w \in W$ as pessoas rastreadas pela saída do método alvo, enquanto suas arestas representam o grau de similaridade entre as rotas de cada pessoa;
3. Acoplamento maximal entre os vértices $v \in V$ e $w \in W$, a partir das arestas estabelecidas na modelagem;
4. Determinação dos verdadeiros positivos (TP), falsos positivos (FP) e falsos negativos (FN) que podem ser usados para definir medidas, como a *precisão*, *revocação* e *F-score*.

A geração da referência é documentada na Seção 4.1. Na Seção 4.2 é indicado o Problema do Acoplamento, suas possíveis soluções discutidas na Seção 4.3 e a solução utilizada na Seção 4.4.

4.1 Geração da Referência

A referência é gerada através da análise de todo o vídeo, quadro-a-quadro. As anotações são feitas apenas para os quadros onde ocorre algum evento. Um evento é considerado como uma pessoa que entra ou sai da região de interesse.

Uma vez que um evento acontece, ele é inserido em uma tabela de referência da seguinte forma. Se o evento consiste de uma pessoa entrando na região de interesse, um identificador

numérico único é adicionado conjuntamente a um **senal positivo** na coluna de direção correspondente, *i.e.*, *Up* ou *Down*. Caso contrário, se consiste em alguém saindo da região de interesse, a mesma identificação que foi criada para esta pessoa ao entrar é utilizada, porém com um **senal negativo** e na coluna de direção correspondente.

Note-se que esta convenção *Up* e *Down* pode ser adaptada para vídeos onde as pessoas locomovem-se para a esquerda/direita. É preciso destacar que, nesta convenção, duas ou mais pessoas podem entrar ou sair simultaneamente da região de interesse em um mesmo quadro, sem perda de generalidade. Um exemplo de geração de referência para 4 pessoas é mostrado na Tabela 4.1.

Tabela 4.1: Exemplo de geração da referência

Quadro	Up	Down
1004	0	1
1019	2	0
1083	-1	0
1113	0	-2
2058	3	0
2067	4	0
2114	-3	0
2150	0	-4
2202	5	0
2280	-5	0

Na Tabela 4.1, a pessoa cujo identificador corresponde ao **1**, entrou pela parte inferior da zona de contagem no quadro 1004, e saiu pela parte superior da zona de contagem no quadro 1083, determinando um movimento **unidirecional**. Um caso particular desse processo de geração de referência é representado pela pessoa **3**, que adentra a zona de contagem por cima e retorna pela mesma direção. Este caso representa um movimento **bidirecional**.

4.2 Modelagem do Problema

Após a geração da referência e em posse da saída do método alvo segundo as convenções estipuladas na seção anterior, deseja-se saber quais pessoas rastreadas pelo método alvo correspondem, isto é, mais se assemelham, às pessoas rastreadas pelos dados de referência.

O rastreamento das pessoas por um típico método de contagem não segue um comportamento regular, de modo que o método alvo pode identificar a entrada ou a saída de uma pessoa com uma latência maior ou menor. Ou seja, ele pode se precipitar ou tardar a identificar uma pessoa no vídeo.

A Tabela 4.2 contém um exemplo de uma possível saída de um método de contagem para o mesmo vídeo dos dados de referência da Tabela 4.1. Note que a pessoa cujo identificador

Tabela 4.2: Exemplo de possível saída de um método de contagem de pessoas

Quadro	Up	Down
1006	0	1
1025	2	0
1090	-1	0
1109	0	-2
2055	3	0
2070	4	0
2110	-3	0
2153	0	-4
2155	0	5
2192	-5	0

corresponde ao **1** entrou no quadro 1006 – um atraso de dois quadros no reconhecimento. Além disso, observe que o método detectou uma pessoa a mais que o indicado pela referência. Como efetuar o acoplamento entre a referência e a saída do método alvo considerando que as pessoas rastreadas não são, na maioria das vezes, idênticas?

Neste trabalho, o Problema de Acoplamento entre as pessoas da referência com relação à saída do método alvo é modelado do seguinte modo. Primeiramente, extrai-se da referência e da saída do método alvo uma representação simples. Cada pessoa rastreada é representada como uma tripla composta por seu identificador único (ID^i) e os números dos quadros onde ela entra e sai da região de interesse – Q_{in}^i e Q_{out}^i respectivamente. Juntamente a estes quadros, são acrescentadas as informações de direção.

Portanto, a i -ésima pessoa rastreada dos dados de referência R pode ser instanciada como a tripla $(RID^i, RQ_{in}^i, RQ_{out}^i)$, enquanto a j -ésima pessoa rastreada pelo método M , como $(MID^j, MQ_{in}^j, MQ_{out}^j)$. A Tabela 4.3 ilustra esta representação, originária da Tabela 4.1. É importante ressaltar que a saída do método alvo respeita as mesmas convenções impostas para a geração da referência, porém a cardinalidade dos dois conjuntos pode ser diferente.

Cada conjunto de pessoas rastreadas para a referência R e o método M podem ser vistos como conjuntos disjuntos, onde o peso de conexão entre os elementos desses conjuntos são proporcionais às suas sobreposições no domínio do tempo¹. O problema de acoplar as pessoas entre os dados de referência e dados de saída do método alvo passa a pertencer ao domínio de Teoria dos Grafos, em Problema de Acoplamento. Para calcular os pesos das arestas, P^{ij} , entre RID^i e MID^j propõe-se considerar a interseção e união de seus intervalos de tempo, *i.e.*,

$$P^{ij} = \frac{|RInt^i \cap MInt^j|}{|RInt^i \cup MInt^j|} \quad (4.1)$$

¹O domínio do tempo de uma pessoa é o número de quadros que ela aparece no vídeo, *i.e.*, $RQ_{out}^i - RQ_{in}^i + 1$ ou $MQ_{out}^j - MQ_{in}^j + 1$.

Tabela 4.3: Exemplo de vértices da tabela de referência da Tabela 4.1 usada no Problema de Acoplamento

RID	RQ_{in}	RQ_{out}
1	-1004	+1083
2	+1019	-1113
3	+2058	+2114
4	+2067	-2150
5	+2202	-2280

em que $RInt^i$ e $MInt^j$ correspondem ao intervalo de tempo de RID^i e MID^j das pessoas rastreadas, respectivamente, e $0 \leq P^{ij} \leq 1$. Quando a direção do movimento das pessoas rastreadas RID^i e MID^j é diferente, atribui-se a sua aresta peso nulo.

A partir desta definição, temos que quanto maior o intervalo de tempo entre a interseção RID^i e MID^j , maior é o peso da aresta. Em contraste, quanto maior é a união dos intervalos de tempo entre RID^i e MID^j , menor é o peso da aresta. A Figura 4.1 ilustra o resultado deste procedimento, onde as arestas mais escuras representam um peso maior enquanto as arestas mais claras representam um peso menor.

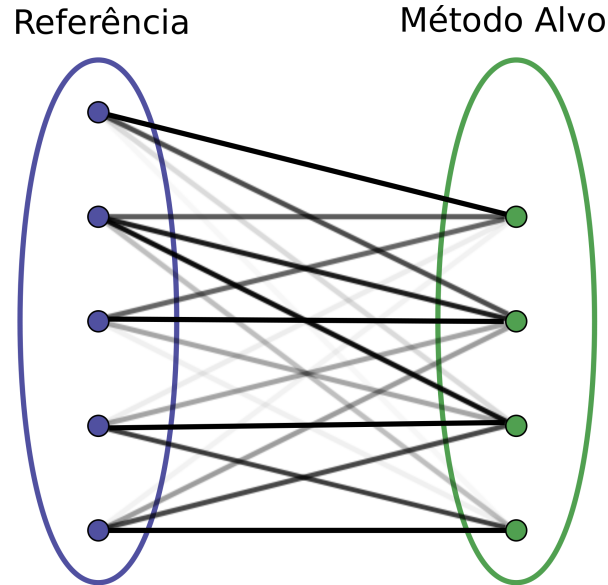


Figura 4.1: Grafo bipartido representando o Problema de Acoplamento. Cada aresta possui um peso baseado na Equação 4.1. Arestas mais escuras possuem peso maior; arestas mais claras possuem menor peso.

4.3 Resolução do Problema

No caso de grafos bipartidos, o Problema de Acoplamento é conhecido como o **Problema de Alocação**, um caso especial do Problema de Transporte (*Transportation Problem*) que é, por sua vez, um subproblema de Problemas de Fluxo de Custo Mínimo (*Minimum-cost Flow Problem*). Portanto, alguns dos algoritmos que se aplicam a estes conjuntos de problemas podem ser aplicados na resolução do Problema de Acoplamento, como **Programação Linear** ou o algoritmo **Simplex de Rede** (*Network Simplex*), considerando o grafo como uma rede incapacitada (Balakrishnan, 1995).

Jungnickel (2007) comprova que a melhor ordem de complexidade conhecida para a resolução do problema é $O(n^3)$ em grafos bipartidos completos $K_{n,n}$ com funções de custo positivas. Dentre os algoritmos conhecidos que se enquadram nesta ordem de complexidade, lista-se o **Algoritmo Húngaro**, um dos modos de se resolver o Problema de Acoplamento especificado neste trabalho.

Por outro lado, para grafos bipartidos não-completos, existem outros algoritmos mais eficientes, como o descrito por Fredman e Tarjan (1987) cuja ordem de complexidade é de $O(|V||E| + |V|^2 \log |V|)$. O Problema de Acoplamento deste trabalho não precisa necessariamente ser modelado como um grafo bipartido completo, possibilitando o uso desta solução. Todavia, optou-se por utilizar o Algoritmo Húngaro devido a sua simplicidade (Boaventura Netto, 2006; Balakrishnan, 1995) e sua importância histórica entre os algoritmos (Jungnickel, 2007).

4.4 O Algoritmo Húngaro

Elaborado por Kuhn (1955), seu nome se deve a vinculação as ideias dos húngaros König e Egerváry. Visa encontrar um acoplamento de cardinalidade máxima e valor mínimo, em um grafo bipartido valorado, com custos não negativos (Boaventura Netto, 2006). Como o problema deste trabalho requer que seja calculado o valor máximo, é necessário aplicar a Equação 4.2 para inverter o peso das arestas e assim reduzir ao problema de custo mínimo.

$$P^{ij} = INF - P^{ij}, \quad (4.2)$$

em que P^{ij} é o peso de uma aresta entre os vértices i e j , e INF um número suficientemente grande, definido como

$$INF = \max_{(i,j) \in E} P^{ij}, \quad (4.3)$$

em que $0 \leq P^{ij} \leq 1$ e, portanto, utilizou-se $INF = 1$.

O Algoritmo Húngaro trabalha com uma matriz de biadjacência, procurando obter um conjunto de valores nulos independentes através da subtração de constantes de cada linha e

coluna. Boaventura Netto (2006) afirma que é possível provar que o conjunto de soluções ótimas é o mesmo, para a matriz original e para todas as matrizes dela obtidas por esse processo.

Considere o conjunto RID^i e MID^j , um acoplamento corresponde a um conjunto de posições independentes na matriz (V, W) . O algoritmo procura o menor número de linhas e colunas que contenha todos os zeros obtidos pelas subtrações, equivalente a obter uma cobertura de vértices para o grafo parcial correspondente às arestas zeradas. Se o acoplamento não for perfeito, torna-se possível zerar ao menos uma nova aresta, o que permitirá aumentá-lo.

O algoritmo também pode ser usado quando $|V| \neq |W|$. Para isso, é necessário completar com vértices fictícios o conjunto de menor cardinalidade, unindo esses vértices aos do outro conjunto com arestas de custo nulo. Neste caso, não existe um acoplamento perfeito para o grafo original, em que pelo menos um vértice não participará do acoplamento maximal a ser obtido. Ele pode ser dividido em duas fases. A seguir, o passo-a-passo, extraído de Boaventura Netto (2006):

Primeira fase do algoritmo

1. Subtrair de todos os elementos de cada linha o mínimo da linha.
2. Se um zero foi obtido em cada coluna, ir para o **passo 4**; se foi obtido **exatamente** um zero em cada coluna, **fim** (solução ótima).
3. Subtrair de todos os elementos de cada coluna sem zeros o mínimo da coluna; ir para o **passo 2**.
4. Marcar um zero na primeira linha (fazendo assim uma alocação) e inabilitar a linha e a coluna desse zero; repetir para as demais linhas, até que não haja mais zeros disponíveis.
5. Se se obteve um acoplamento perfeito, **fim** (solução ótima); caso contrário, passar à segunda fase, a qual deve ser iniciada com um acoplamento maximal.

A **segunda fase** do algoritmo determina uma cobertura de vértices para os zeros e procura obter ao menos um zero dentre os elementos não cobertos. Utilizam-se duas operações: **marcar** e **riscar**. A última corresponde à construção de uma cobertura de vértices.

Segunda fase do algoritmo

1. **Marcar** as linhas que não receberam alocações no **passo 4** da primeira fase.
2. **Marcar** as colunas não marcadas que possuem **zeros** em linhas marcadas.
3. **Marcar** as linhas não marcadas que receberam **alocações** em colunas marcadas.

4. Repetir os passos 2 e 3 até que não ocorram novas marcações.
5. Riscar todas as linhas não marcadas e todas as colunas marcadas.
6. Subtrair de todos os elementos não riscados o menor deles e somá-lo aos elementos que tiverem sido riscados duas vezes (em linha e coluna).
7. Voltar ao passo 4 da primeira fase.

4.4.1 Discussão da Segunda Fase

No final do algoritmo, uma linha não marcada sem alocações é riscada, equivalente a cobrir seus zeros não utilizados com o vértice $v \in V$ a ela associado. Deste modo, uma coluna marcada é riscada e, portanto, todos os seus zeros são cobertos por vértices $w \in W$. Assim, para todas as arestas zeradas, o processo gera uma cobertura de vértices.

Marca-se cada linha não marcada possuindo uma alocação em uma coluna marcada – portanto, esta linha não é marcada pela **Regra 1**. Logo, a aresta alocada será coberta por um vértice $v \in V$.

A **etapa 6** gera ao menos um novo zero que é equivalente a subtrair o menor elemento não riscado de toda a matriz e voltar a somá-lo a cada linha e cada coluna riscadas. A **etapa 7** remete à criação de um novo acoplamento.

Se a segunda fase resultar em um número de linhas maior do que o número de alocações obtidas, deve existir um acoplamento de maior cardinalidade que o já construído. Isto pode ser encontrado no exemplo da Seção 4.4.2.

4.4.2 Simulação do Algoritmo

A simulação deste algoritmo será realizada sobre o grafo apresentado na Figura 4.2, extraído de Boaventura Netto (2006). O peso de suas arestas é representado pela matriz de biadjacência ao seu lado.

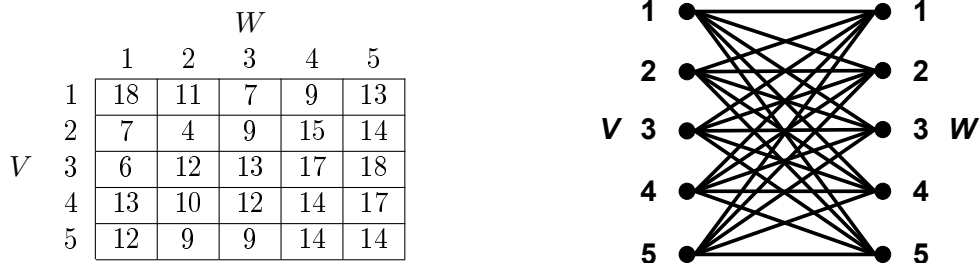


Figura 4.2: Matriz de biadjacência e sua representação em forma de grafo, antes da execução do algoritmo

Na execução da primeira fase do Algoritmo Húngaro não obteve-se um acoplamento perfeito, haja vista que a coluna 4 não possui um zero independente (Figura 4.3).

		W					
		1	2	3	4	5	
V	1	18	11	7	9	13	(-7)
	2	7	4	9	15	14	(-4)
	3	6	12	13	17	18	(-6)
	4	13	10	12	14	17	(-10)
	5	12	9	9	14	14	(-9)

$\Rightarrow$

11	4	0	0	1
3	0	5	9	5
0	6	7	9	7
3	0	2	2	2
3	0	0	3	0

(-2) (-5)

Figura 4.3: Primeira fase do algoritmo

A Figura 4.4 ilustra o acoplamento maximal obtido nesta etapa, com custo total de 43. Os zeros em negrito correspondem às arestas cheias e os demais zeros correspondem às arestas pontilhadas.

Ainda na Figura 4.4, os números na parte superior (**2**) e na lateral direita da tabela (**3** e **1**) correspondem às etapas do algoritmo durante a marcação da segunda fase. Não há necessidade da etapa 4 (repetição). As posições não riscadas (4,3) e (4,4) indicam o número mínimo (2) a ser subtraído de todas as posições não riscadas e somado aos cruzamentos que aparecem na coluna 2.

		2					
		↓					
		1	2	3	4	5	
V	1	11	4	0	0	1	
	2	3	0	5	9	5	$\Leftarrow 3$
	3	0	6	7	9	7	
	4	3	0	2	2	2	$\Leftarrow 1$
	5	3	0	0	3	0	

W

Figura 4.4: Segunda fase do algoritmo

Após a execução da segunda fase, obteve-se a matriz apresentada a seguir, com acoplamento perfeito de custo 45, obtido por nova primeira fase, etapas 4 e 5 (Figura 4.5):

Mais detalhes sobre a implementação do algoritmo húngaro pode ser encontrada no Apêndice.

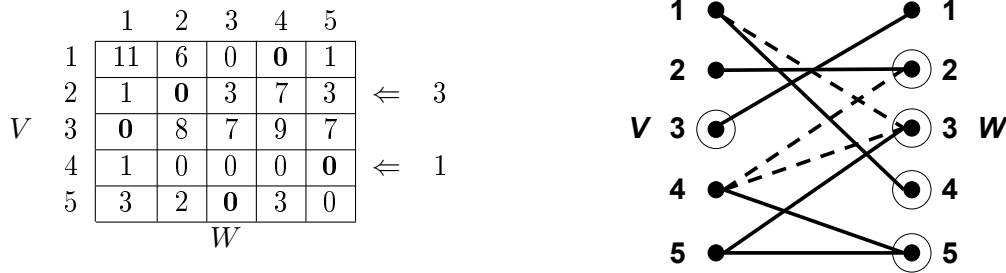


Figura 4.5: Resultado da execução do algoritmo

4.5 Etapa Pós-Acoplamento

Após efetuar-se o acoplamento entre a referência RID^i e a saída do método de contagem MID^j por meio do Algoritmo Húngaro, é possível quantificar os valores de TP, FP, e FN. Suponha que foi estabelecido um acoplamento final entre as Tabelas 4.1 e 4.2, apresentadas no início deste capítulo. O resultado é descrito na Tabela 4.4.

Tabela 4.4: Resultado do acoplamento estabelecido entre as Tabelas 4.1 e 4.2

RID^i	MID^j
1	1
2	2
3	3
4	4
5	x
x	5

Verdadeiros Positivos Todo acoplamento estabelecido entre um vértice da referência RID^i e um vértice do método MID^j cujo peso $P^{ij} \neq 0$, é considerado um verdadeiro positivo (TP). Sendo assim, do acoplamento da Tabela 4.4 obtém-se $TP = 4$ (vértices 1-1, 2-2, 3-3, 4-4);

Falsos Negativos Consistem nos elementos de RID^i não acoplados ou acoplados com custo $P^{ij} = 0$. Para o acoplamento deste exemplo, obtém-se $FN = 1$ (vértice 5 sem um par);

Falsos Positivos Consistem no inverso dos falsos negativos: os elementos de MID^j tal que não foram estabelecidos acoplamentos ou obteve-se um acoplamento de custo $P^{ij} = 0$. Para este exemplo, coincidentemente $FP = 1$ (vértice 5 sem par).

Por meio desta metodologia, também é possível identificar automaticamente onde ocorrem situações de erro a partir dos dados referenciais. Em primeiro lugar deve-se calcular o número

esperado de pessoas na zona de contagem em cada quadro. Esta informação pode ser facilmente estimada acumulando o tempo de intervalo de cada pessoa controladas em um vetor de tempo total. A Tabela 4.5 exemplifica este procedimento a partir dos dados de referência da Tabela 4.1.

Tabela 4.5: Número esperado de pessoas em cada quadro

Quadro	Num. Pessoas
1003	0
1004	1
1005	1
...	1
1018	1
1019	2
1020	2
...	2
1083	1
...	1
1113	0
1114	0

Uma vez que este vetor de tempo total é calculado, é possível obter o número esperado de pessoas para cada pessoa rastreada na referência de dados e na saída do método. Para isso, utilizando a Tabela 4.5, são efetuadas consultas sobre o número máximo de pessoas em um intervalo $[Q_{in}^i, Q_{out}^i]$, tanto para a referência R , como para a saída do método M e obtidos por uma função $mVR(x)$ e $mVM(x)$, respectivamente, em que x é o valor de FN, FP ou TP. Como os conjuntos FN e FP são conhecidos, para a construção do histograma é suficiente acumular cada erro em sua posição correspondente em que cada uma é dada por $mVR(FN^i)$ e $mVM(FP^i)$, respectivamente. Semelhante processo pode ser executado para obter a situações onde as pessoas TPs rastreadas acontecem ($mVR(TP^i)$).

Por exemplo, seja $FN^1 = 5$ e o máximo de pessoas em seu intervalo $mVR(5) = 2$. Em seu histograma, na coluna de quantidade de pessoas 2, será acumulado o erro obtido por esta função. Portanto, $hFN(2) = hFN(2) + 1$.

Capítulo 5

Experimentos

Para efetuar a validação da metodologia de avaliação automática proposta neste trabalho, implementou-se em MATLAB um método de contagem de pessoas, descrito na Seção 5.1. Este método foi executado utilizando três vídeos distintos como entrada, com diferentes configurações de parâmetros. Para cada saída, o resultado obtido foi analisado segundo a metodologia proposta. É demonstrado como a metodologia pode ser utilizada para identificar erros na contagem. As características dos vídeos utilizados nos testes são detalhadas na Seção 5.2; as medidas, definidas a partir do número de FP, FN e TP extraídos, são definidas na Seção 5.3; e a discussão dos resultados na Seção 5.4.

5.1 O Método de Contagem de Pessoas

Esta seção descreve em maiores detalhes o método para contagem de pessoas utilizado na validação da metodologia de avaliação proposta neste trabalho. O método que foi implementado é dividido em: a partir da captura do vídeo, efetua-se a subtração do fundo, segmentação, rastreamento e contagem de pessoas (ver Figura 5.1). As operações nos quadros do vídeo são feitas em blocos de *pixels*, o que reduz a quantidade de computações e o efeito é o mesmo obtido caso essas operações fossem feitas *pixel-a-pixel*. O tamanho padrão para os blocos é 8×8 .

5.1.1 Subtração do Fundo

A primeira etapa do método é a subtração do fundo. Essa operação é essencial para a detecção das pessoas que será realizada posteriormente, através da comparação dos blocos do quadro atual com os blocos do quadro pertencente ao fundo. As imagens ou quadros que pertencem ao fundo do vídeo são obtidas através do seguinte filtro

$$F^{t+1} = (1 - \alpha) \cdot F^t + \alpha \cdot I^t \quad (5.1)$$

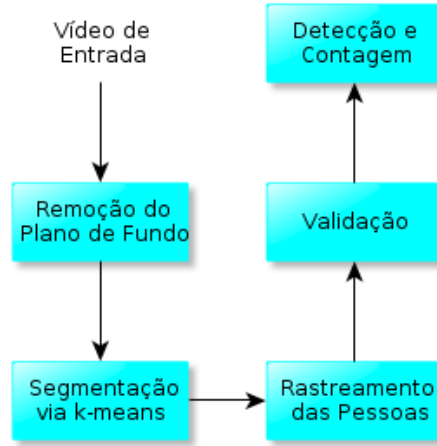


Figura 5.1: Fluxograma para representação do sistema

em que F e I representam, respectivamente, os quadros de fundo e os quadros do vídeo original; t é o número do quadro; e α é uma taxa de aprendizado que pode variar entre 0.01 e 0.1. Essa taxa deve ser ajustada de acordo com as características do vídeo. Para os vídeos deste trabalho, observou-se que os ruídos foram minimizados ao utilizar 0.01 como taxa de aprendizado. O filtro é aplicado sobre todos os quadros, em todos os seus canais RGB ¹.

O algoritmo utiliza fatores multiplicativos $\beta_{m,n,p}^t$, determinados através da **Estimativa de Verossimilhança Máxima** (*Maximum-likelihood Estimation*, ou **MLE**). Observe que é um método estatístico utilizado para ajustar os dados a um modelo e fornecer estimativas para os seus parâmetros, *i.e.*,

$$\beta_{m,n,p}^t = \frac{\sum I_{m,n,p}^t \cdot F_{m,n,p}^t}{\sum (F_{m,n,p}^t)^2} \quad (5.2)$$

em que os índices (m,n) referem-se às coordenadas dos blocos e p aos canais da imagem – RGB .

A detecção de pessoas nos quadros é realizada através da diferença entre os fatores multiplicativos máximo e mínimo. Para cada quadro, são calculados o maior e o menor β entre os canais da imagem e sua diferença é armazenada em $\delta\beta^t$.

$$\delta\beta^t = \max_p \beta_{m,n,p}^t - \min_p \beta_{m,n,p}^t \quad (5.3)$$

Sejam os blocos referentes às pessoas definidos por P^t . Os fatores multiplicativos dos blocos do fundo tem valor aproximado de 1. Assim, os *pixels* P^t referentes as pessoas são determinados da seguinte forma

¹ **Red**, **Green** e **Blue**. Vermelho, verde e azul, respectivamente

$$P^t = \begin{cases} 1, & \text{se } \delta\beta^t > T_1 \vee |\beta_{m,n,p}^t| > T_2, \\ 0, & \text{caso contrário} \end{cases} \quad (5.4)$$

em que P é a imagem com as pessoas e T_1 e T_2 são limiares entre $[0, 1; 0, 2]$ e $[0, 3; 0, 6]$, respectivamente. Esses parâmetros também devem ser ajustados através de experimentos para cada situação específica. Para os vídeos utilizados neste trabalho, o melhor ajuste dos parâmetros consistiu em $T_1 = 0,2$ e $T_2 = 0,6$, por minimizar o ruído. Nesse momento, obteve-se uma imagem P para cada quadro, possuindo os blocos referentes às pessoas. O próximo passo do algoritmo é a sua segmentação.

5.1.2 Segmentação

A segmentação é um problema difícil em Análise de Imagem, devido às várias características que representam uma pessoa. Como nos vídeos em questão aparece apenas a parte superior das pessoas, esse problema é reduzido. Assim as pessoas passam a ser vistas como formas geométricas (Figura 5.2), podendo extrair o número de pessoas dessa figura por meio de técnicas tradicionais de *clustering*, como o *k-means*. A Figura 5.2 ilustra o processo de subtração de fundo do método. As imagens da primeira coluna são os quadros originais e as imagens da segunda coluna os blocos de pessoas.

No *k-means* (Duda et al., 2000), existem k centróides, um para cada grupo – ou *cluster*. Cada indivíduo é associado ao centróide mais próximo, e os centróides são recalculados como a média dos elementos classificados. Novamente, é feita um agrupamento dos elementos baseado neste novo centróide, e este processo é repetido até convergir. O método de *k-means* recebe como parâmetro um valor k que é o número de agrupamentos que desejados naquele quadro.

O número de grupos desejado é justamente o número de pessoas na região de interesse. No entanto, este valor não é conhecido *a priori*. O valor de k é estimado como o número máximo de *clusters* em que a distância entre os *clusters* é maior do que uma distância mínima D_{min} . Essa constante corresponde ao tamanho médio de uma pessoa na cena, e deve ser estabelecida através de experimentos. Em uma imagem com k *clusters*, cujos centróides são $C_i, i = 1, 2, \dots, k$, a distância mínima entre os *clusters* é definida como

$$d_{min}^k = \min_{1 \leq i < j \leq k} \|C_i - C_j\| \quad (5.5)$$

No caso de apenas um *cluster*, definimos formalmente $d_{min}^1 = \infty$. O número ótimo de *clusters* k^* é então estimado como o máximo número de *clusters* que possuem a distância mínima dentro do *cluster* d_{min}^k maior que D_{min} , *i.e.*,

$$k^* = \max\{k | d_{min}^k \geq D_{min} \wedge d_{min}^{k+1} < D_{min}\} \quad (5.6)$$

No *k-means*, a inicialização dos centróides é muito importante pois pode-se melhorar a

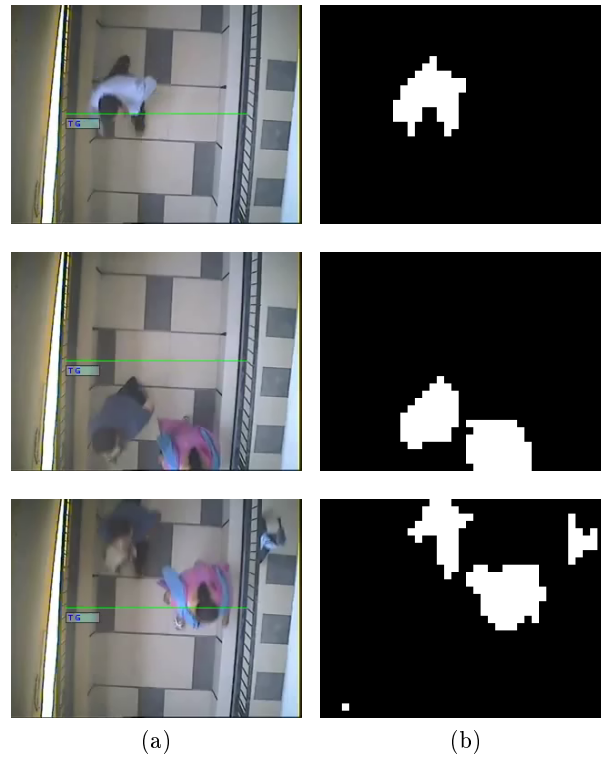


Figura 5.2: Resultados demonstrativos das etapas de subtração do fundo e segmentação de pessoas. (a) quadro original. (b) Subtração do fundo e segmentação de pessoas.

convergência do algoritmo. Sempre que possível, os centróides são inicializados com a posição dos centróides encontrados no quadro anterior. Dessa forma, eles são inicializados em uma posição mais próxima de ser a melhor para os *clusters*, pois o deslocamento de uma pessoa no vídeo é pequeno.

A Figura 5.3 apresenta o resultado da segmentação das pessoas através do *k-means*, onde o número de *clusters* é automaticamente determinado usando a distância mínima inter-cluster. Nesse caso, o método segmentou corretamente encontrando o valor de $k = 2$, *i.e.*, duas pessoas.

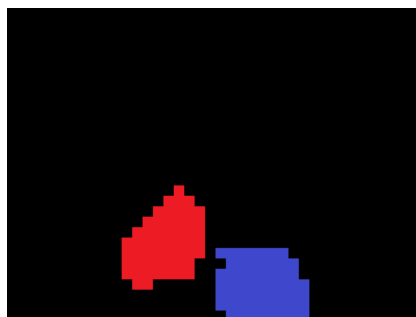


Figura 5.3: Após segmentação pelo *k-means*

Nesse ponto do algoritmo são conhecidas as pessoas em cada quadro do vídeo. A próxima parte consiste em realizar o rastreamento dessas pessoas, ou seja, descobrir se a mesma pessoa está em vários quadros consecutivos para então contá-las. Esse passo é implementado por uma **Estratégia Gulosa**, analisando dois quadros consecutivos por vez.

5.1.3 Rastreamento

O algoritmo encontra os *clusters* correspondentes em dois quadros consecutivos Q^i e Q^j que possuam a menor Distância Euclidiana. Para tanto, é calculado a distância de todo centróide pertencente ao quadro Q^i para todo centróide pertencente ao quadro Q^j . Em seguida, por meio da Estratégia Gulosa supracitada, a menor distância encontrada equivale a dois *clusters* correspondentes entre os quadros Q^i e Q^j . Este processo é repetido para as distâncias restantes até que todos os *clusters* de um quadro estejam pareados. Deste modo, a rota descrita por cada pessoa é armazenada, ou seja, o centróide de cada *cluster* correspondente a uma mesma pessoa.

Após obter a rota de cada pessoa pela zona de contagem, são posicionadas arbitrariamente duas linhas auxiliares LIN1 e LIN2. Apenas as rotas das pessoas que expandirem mais que a metade do tamanho da região de interesse serão contadas. Esta convenção evita que ruído ou má segmentação interfiram na contagem, pois geralmente são menores que o tamanho necessário.

A Figura 5.4 ilustra esta etapa. Para fins didáticos, foram introduzidas retas paralelas pontilhadas ao longo do quadro. Elas determinam a largura de 1 *pixel* na imagem. A região delimitada pelas linhas LIN1 e LIN2 consiste na região de interesse. A área hachurada corresponde a largura que as rotas precisam possuir para que sejam contabilizadas. Ou seja, neste exemplo as rotas precisam ser maior que 5.

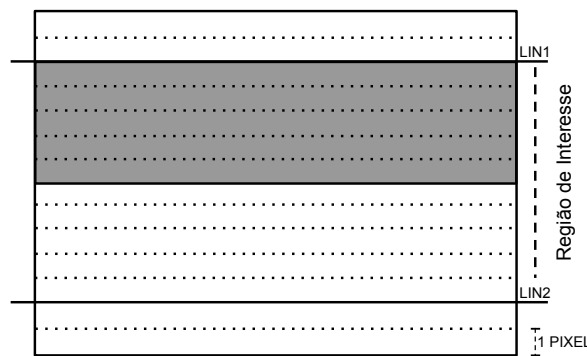


Figura 5.4: Exemplo de posicionamento das linhas auxiliares para contagem de pessoas

A Figura 5.5 apresenta, a título de exemplo, três rotas identificadas pelo método de contagem. Cada sequência de quadrados de uma cor equivale a rota de uma pessoa identificada pelo método. Deste modo, o método iria contabilizar apenas a rota azul por expandir em

8 *pixels*, enquanto as rotas em vermelho e verde não seriam contabilizadas por não possuir tamanho suficiente (3 e 1, respectivamente). Portanto, o método contará apenas uma pessoa neste vídeo exemplo.

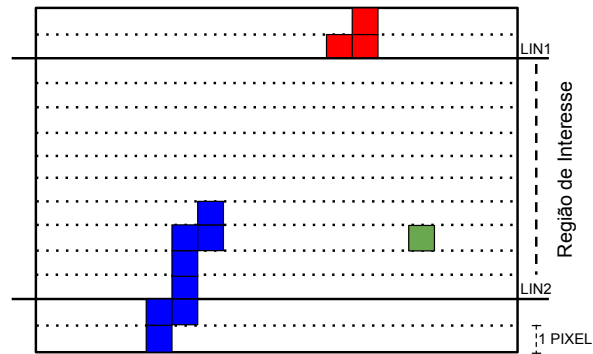


Figura 5.5: Exemplo de rotas identificadas pelo método de contagem. Apenas a rota em azul expandiu o suficiente para ser contada

5.1.4 Parâmetros Ajustados

A Tabela 5.1 apresenta os parâmetros finais utilizados para este método em todos os três vídeos. O parâmetro que requer maior atenção é o tamanho mínimo de uma pessoa (D_{min}), sendo o principal responsável por segmentações incorretas. As linhas superior e inferior representam a delimitação da zona de contagem no vídeo, utilizada no processo de contagem pelo método.

Tabela 5.1: Valor dos parâmetros utilizados

Parâmetro	Valor
α	0,01
Bloco	8×8
T_1	0,2
T_2	0,6
D_{min}	$\{50, 65, 80\}$ <i>pixels</i>
LIN1	30
LIN2	450

5.2 Propriedades dos Vídeos

Os vídeos foram adquiridos em posição zenital (Figura 5.6) no corredor do Departamento de Computação do Instituto de Ciências Exatas e Biológicas, da Universidade Federal de Ouro Preto (DECOM/ICEB - UFOP).



Figura 5.6: Posicionamento da câmera no corredor do DECOM/ICEB - UFOP.

São três vídeos nomeados **stm1**, **stm2** e **stm3**. Possuem 10 minutos de duração cada, 30 quadros por segundo (*fps*) e resolução de 640×480 *pixels*.

O vídeo **stm1** possui situações adversas, como pessoas que caminham muito próximas e objetos na região de interesse (ver Figura 5.7).



Figura 5.7: Pessoas juntas e um guarda-chuva na região de interesse no vídeo **stm1**

A principal adversidade observada nos vídeos **stm2** e **stm3** são as variações de luminosidade, visível na Figura 5.8.



Figura 5.8: Pequena variação de luminosidade na região inferior da imagem, encontrada nos vídeos **stm2** e **stm3**

Finalmente, a Tabela 5.2 apresenta resumidamente a quantidade de pessoas total que atravessa a região de interesse para cada um dos vídeos.

Tabela 5.2: Número total de pessoas que atravessa a região de interesse em cada vídeo

Vídeo	Qtd. pessoas
stm1	33
stm2	16
stm3	22

Os vídeos, juntamente com os dados de referência, estão disponíveis para *download* na

página do Laboratório de Processamento Digital de Imagens².

5.3 Medidas

Ao final do processo da seção anterior, é estabelecido um acoplamento de cardinalidade máxima e peso máximo. As alocações cujo peso $P^{ij} = 0$ são desconsideradas na contagem do total de alocações.

A partir do número de vértices alocados de R para M (verdadeiros positivos ou **TP**), o número de vértices não alocados em R (falsos negativos ou **FN**) e em M (falsos positivos ou **FP**), três medidas podem ser utilizadas, isto é, *precisão*, *revocação*, e *F-score*.

$$precisão = \frac{TP}{TP + FP}, \quad (5.7)$$

$$revocação = \frac{TP}{TP + FN}, \quad (5.8)$$

e

$$F\text{-score} = \frac{2 \times precisão \times revocação}{precisão + revocação}. \quad (5.9)$$

Esta não é a primeira vez na literatura de contagem de pessoas com base em métodos de análise de vídeo que essas medidas são utilizadas para avaliar os resultados de contagem. No entanto, devido a metodologia de avaliação proposta, é possível determinar automaticamente em quais situações, segundo o número de pessoas presentes na zona de contagem, estes erros acontecem como foi descrito na metodologia.

5.4 Resultados

Além de avaliar automaticamente os resultados de um método de contagem de pessoas aplicado a um dado vídeo, a metodologia proposta pode ser utilizada no ajuste de parâmetros de um método de contagem de pessoas. Para ilustrar esta capacidade, a saída gerada pelo método alvo foi avaliada nos três vídeos variando os valores do principal parâmetro do método, isto é, a mínima distância *intercluster* permitida – D_{min} – que é relacionada ao tamanho médio de uma pessoa, em *pixels*.

Três valores foram utilizados para o parâmetro $D_{min} = \{50, 65, 80\}$ no experimento. A Tabela 5.3 apresenta, em detalhes, o número de TPs, FNs e FPs e as medidas determinadas para estes parâmetros, automaticamente geradas pela metodologia de avaliação proposta.

Note que a soma dos valores TP e FN é constante para cada vídeo, haja vista que representa o número real/esperado que atravessaram a região de interesse, ou seja, os dados de referência.

²<http://www.iceb.ufop.br/decom/lapdi/peoplecounting/>

Por outro lado, a soma dos valores TP e FP variam para cada vídeo, dependendo do valor de D_{min} , pois representam o número de pessoas contabilizadas pelo método alvo.

Tabela 5.3: A variação do parâmetro D_{min} e seu impacto sobre a avaliação dos resultados para diferentes vídeos

		Número de Pessoas			Medidas (%)		
Vídeo	D_{min}	TP	FP	FN	<i>precisão</i>	<i>revocação</i>	<i>F-score</i>
stm1	50	29	6	4	82,9	87,9	85,3
	65	28	4	5	87,5	84,8	86,2
	80	26	0	7	100,0	78,8	88,1
stm2	50	16	25	0	39,0	100,0	56,1
	65	15	17	1	46,9	93,8	62,5
	80	15	15	1	50,0	93,8	65,2
stm3	50	21	25	1	45,7	95,5	61,8
	65	21	18	1	53,8	95,5	68,9
	80	21	10	1	67,7	95,5	79,2

Analisando os valores da Tabela 5.3, observa-se que os valores de *precisão* – que depende diretamente dos valores de FP – e consequentemente a medida *F-score*, são menores para os vídeos **stm2** e **stm3** em comparação ao vídeo **stm1**, enquanto os valores de *revocação* – que depende diretamente dos valores de FN – não são tão sensíveis para os três vídeos assim como a *precisão*. Este resultado pode ser justificado pela súbita variação de iluminação que ocorre com maior frequência nos vídeos **stm2** e **stm3** e causam um maior número de alarmes falsos na etapa de segmentação. Nota-se que a implementação do método alvo deste trabalho é sensível a mudança de fundo.

Apesar disso, pode-se constatar, para os três vídeos, que os valores de TP e FN diminuem de acordo com o aumento do parâmetro D_{min} , enquanto FP aumenta conforme D_{min} diminui. Estas informações relacionadas ao método de contagem podem ser utilizadas para ajustar seus parâmetros de acordo com a necessidade do sistema do usuário. Vale ressaltar que estas informações foram obtidas por meio da análise dos valores na Tabela 5.3, automaticamente geradas pela metodologia de avaliação proposta, sem a necessidade de efetuar para cada vídeo a recontagem manual. Além disso, a metodologia de avaliação proposta pode ser usada para identificar o momento em que ocorreu um erro de contagem. Isto é, quantas pessoas eram esperadas em um quadro quando ocorreu um FP ou FN.

Para ilustrar este benefício, é apresentado na Tabela 5.4 o número de TP, FN e FP para diferentes quantidades de pessoas nos vídeos, tal que o parâmetro $D_{min} = 50$. No intervalo de tempo (conjunto de quadros) onde não constam pessoas (0 pessoas), não é possível ter nenhum TP ou FN, portanto foram preenchidas por um 'x'. Analisando os valores de FP e FN, é possível verificar onde os erros ocorrem e tentar adaptar a implementação do método de contagem de pessoas para tornar-se mais robusto com relação a essas dificuldades. Baseando-se

nessa tabela, segue alguns exemplos de como isso pode ser realizado:

Tabela 5.4: Número diferente de pessoas na região de interesse e número de TP, FN e FP usando $D_{min} = 50$ para o método implementado

Vídeo		Número de Pessoas				
		0	1	2	3	Total
stm1	TP	x	15	12	2	29
	FP	0	4	2	0	6
	FN	x	1	2	1	4
stm2	TP	x	14	2	0	16
	FP	5	20	0	0	25
	FN	x	0	0	0	0
stm3	TP	x	14	7	0	21
	FP	0	21	4	0	25
	FN	x	0	1	0	1

Exemplo 1: Para o vídeo de entrada **stm1**, o método não identificou nenhuma pessoa quando os quadros eram vazios (Tabela 5.4, Número de pessoas: 0), indicando que o método não acrescentou nenhum ruído a contagem final;

Exemplo 2: Ainda para o vídeo de entrada **stm1**, quando havia 1 ou 2 pessoas na região de interesse (Tabela 5.4, Número de pessoas: 1/2), o método contou equivocadamente algumas pessoas a mais, isto é, ele segmentou de forma incorreta uma pessoa como duas ou mais. Cada cor na Figura 5.9 corresponde a uma pessoa identificada pelo método. Essa posição no vídeo foi detectada automaticamente usando a metodologia proposta;

Exemplo 3: Para a entrada **stm2**, esta implementação do método de contagem de pessoas contabilizou uma pessoa quando não havia ninguém na região de interesse (Tabela 5.4, Número de pessoas: 0). Portanto, é possível inferir que necessita-se correções na implementação para aumentar sua robustez a ruídos ou aplicar filtros para diminuí-los.

5.5 Considerações Finais

Neste capítulo foi apresentada a implementação de um método de contagem de pessoas, as propriedades dos vídeos de gravação, as definições das medidas e, finalmente, a validação da metodologia de avaliação de métodos de contagem de pessoas. Como foi demonstrado, por meio da metodologia de avaliação, é possível quantificar automaticamente o número de FP, FN e TP e também identificar onde ocorreram os erros de contagem para modificar a implementação de um método. Foi possível observar que, após modificar um parâmetro

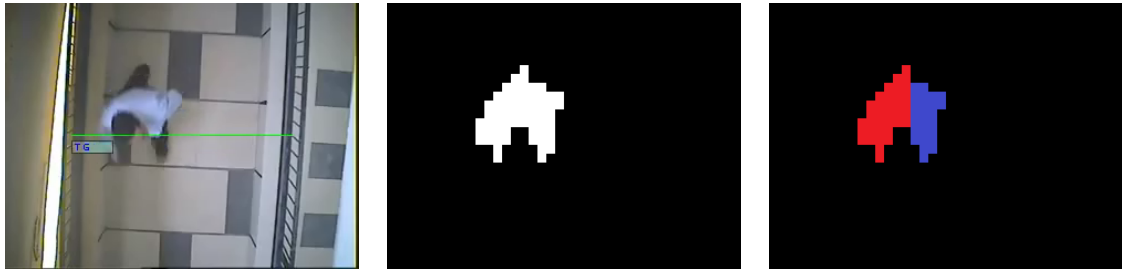


Figura 5.9: Exemplo 2: segmentação errônea. Número real de pessoas na região de interesse: 1; Número de pessoas identificadas pelo método: 2.

sensível do método de contagem de pessoas indicado pela análise da saída da metodologia, o desempenho do método aumentou segundo a medida F -score.

Capítulo 6

Conclusões

Neste trabalho, foi proposta uma metodologia de avaliação automática para métodos de contagem de pessoas baseados em câmeras com posicionamento zenital. Por esta metodologia de avaliação tornou-se possível quantificar automaticamente a contagem de verdadeiros positivos, falsos positivos, e falsos negativos, também permitindo identificar quando estes erros ocorrem (isto é, o número de pessoas presente na região de interesse).

A metodologia de avaliação proposta permite avaliar automaticamente situações em que as pessoas do vídeo variam de velocidade, entre normal, rápida e abrupta, por meio de limiares. Além disso, é possível estabelecer situações onde movimentos unidirecionais e bidirecionais ocorrem. Entretanto, esta metodologia não permite avaliar situações onde as pessoas passam separadas ou se juntam durante o trajeto, pois não são levadas em conta as posições reais das pessoas durante sua passagem.

Um método de contagem de pessoas foi avaliado segundo esta metodologia. Demonstrou-se a capacidade da metodologia de identificar quando os erros ocorreram, de modo que seja possível adaptar a implementação do método de contagem de pessoas de acordo com as necessidades do sistema do usuário. O ajuste do parâmetro D_{min} apresentou, em geral, uma melhora no desempenho com relação a medida $F-score$. E, finalmente, foi disponibilizada a base de vídeos utilizada e os dados de referência na página do Laboratório de Processamento Digital de Imagens (LaPDI).

6.1 Trabalhos Futuros

Como futuras direções para este trabalho, é necessário superar a desvantagem dessa metodologia de não poder detectar pessoas que transitam separadamente ou casos de junção. Uma solução seria efetuar a segmentação perfeita de cada pessoa durante a sua passagem através da região de interesse, quadro a quadro, ou indicar o centro de massa durante a sua passagem. É perceptível que o tempo de geração dos dados de referência pode aumentar consideravelmente.

Também pretende-se concluir a implementação de outros dois métodos de contagem de

pessoas (Barandiaran et al., 2008; Yu et al., 2008) e implementar outros três (Mukherjee et al., 2011; Gasparini et al., 2011; Chen et al., 2009), a fim de realizar uma avaliação extensa dos métodos usando a metodologia aqui proposta.

Além dos métodos de contagem, serão disponibilizadas futuramente uma maior quantidade de vídeos, com maior tempo de duração e coletados em locais diferentes, usando câmeras e condições distintas de iluminação. Além disso, é necessário coletar vídeos variando o número de pessoas na zona de contagem. Isto é, desde os momentos com alto tráfego, onde é possível obter facilmente cinco ou mais pessoas na zona de contagem, até os momentos de baixa atividade.

Apêndice

Neste capítulo é apresentado o pseudocódigo do Algoritmo Húngaro descrito na Seção 4.4, extraído de Jungnickel (2007). O Algoritmo 6.1 descreve sua parte principal e o Algoritmo 6.2 uma função auxiliar.

Algoritmo 6.1: Algoritmo Húngaro

Entrada: Seja $G = (V \cup W, E)$ um grafo bipartido completo, onde $V = \{1, 2, \dots, n\}$ e $W = \{1', 2', \dots, n'\}$ e onde cada aresta ij' de G possui um peso não-negativo associado p_{ij} .

Saída: O algoritmo determina uma alocação ótima em G descrita por um vetor de adjacência $mate$.

// Inicialização

para $v \in V$ **faça**

└ $mate(v) \leftarrow 0$;

para $i = 1$ **até** n **faça**

└ $u_i \leftarrow \max\{w_{ij} : j = 1, 2, \dots, n\}$;

└ $v_i \leftarrow 0$;

$nrex \leftarrow n$;

enquanto $nrex \neq 0$ **faça**

└ **para** $i = 1$ **até** n **faça** $m(i) \leftarrow \text{false}$; $p(i) \leftarrow 0$; $\delta_i \leftarrow \infty$;

└ $aug \leftarrow \text{false}$; $Q \leftarrow \{i \in S : mate(i) = 0\}$;

└ **repita**

└└ remova um vértice arbitrário i de Q ; $m(i) \leftarrow \text{true}$; $j \leftarrow 1$;

└└ **enquanto** $aug = \text{false}$ e $j \leq n$ **faça**

└└└ **se** $mate(i) \neq j'$ **então**

└└└└ **se** $u_i + v_j - w_{ij} < \delta_j$ **então**

└└└└└ $\delta_j \leftarrow u_i + v_j - w_{ij}$; $p(j) \leftarrow i$;

└└└└ **se** $\delta_i = 0$ **então**

└└└└└ **se** $mate(j') = 0$ **então**

└└└└└└ AUGMENT($mate, p, j'$; $mate$);

└└└└└└ $aug \leftarrow \text{true}$; $nrex \leftarrow nrex - 1$;

└└└└ **senão**

└└└└└ $Q \leftarrow Q \cup \{mate(j')\}$

└└ $j \leftarrow j + 1$;

└└ **se** $aug = \text{false}$ e $Q = \emptyset$ **então**

└└└ $J \leftarrow \{i \in S : m(i) = \text{true}\}$; $K \leftarrow \{j' \in T : \delta_j = 0\}$;

└└└ $\delta \leftarrow \min\{\delta_j : j' \in T \setminus K\}$;

└└└ **para** $i \in J$ **faça** $u_i \leftarrow u_i - \delta$;

└└└ **para** $j' \in K$ **faça** $v_j \leftarrow v_j + \delta$;

└└└ **para** $j' \in T \setminus K$ **faça** $\delta_j \leftarrow \delta_j - \delta$;

└└└ $X \leftarrow \{j' \in T \setminus K : \delta_j = 0\}$;

└└└ **se** $mate(j') \neq 0$ para todo $j' \in X$ **então**

└└└└ **para** $j' \in X$ **faça** $Q \leftarrow Q \cup \{mate(j')\}$;

└└└ **senão**

└└└└ escolha $j' \in X$ com $mate(j') = 0$;

└└└└ AUGMENT($mate, p, j'$; $mate$);

└└└└ $aug \leftarrow \text{true}$; $nrex \leftarrow nrex - 1$;

└ **até** $aug = \text{true}$;

Algoritmo 6.2: Procedimento auxiliar AUGMENT

```
repita
     $i \leftarrow p(j);$ 
     $\text{mate}(j') \leftarrow i;$ 
     $\text{next} \leftarrow \text{mate}(i);$ 
     $\text{mate}(i) \leftarrow j';$ 
    se  $\text{next} \neq 0$  então  $j' \leftarrow \text{next}$ 
até  $\text{next} = 0;$ 
```

Referências Bibliográficas

- Antic, B.; Letic, D.; Culibrk, D. e Crnojevic, V. (2009). K-means based segmentation for real-time zenithal people counting. In *IEEE International Conference on Image Processing (ICIP)*, pp. 2565–2568.
- Balakrishnan, V. K. (1995). *Network Optimization*. Chapman & Hall, primeira edição.
- Barandiaran, J.; Murguia, B. e Boto, F. (2008). Real-time people counting using multiple lines. In *International Workshop on Image Analysis for Multimedia Interactive Services (IAMIS)*, pp. 159–162.
- Benabbas, Y.; Ihaddadene, N.; Yahiaoui, T.; Urruty, T. e Djeraba, C. (2010). Spatio-temporal optical flow analysis for people counting. In *IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 212–217.
- Bescos, J.; Menendez, J. M. e Garcia, N. (2003). DCT based segmentation applied to a scalable zenithal people counter. In *IEEE International Conference on Image Processing (ICIP)*, volume 3, pp. 1005–1008.
- Boaventura Netto, P. O. (2006). *Grafos: Teoria, Modelos, Algoritmos*. Editora Edgard Blücher, quarta edição.
- Bozzoli, M. e Cinque, L. (2007). A statistical method for people counting in crowded environments. In *IEEE International Conference on Image Analysis and Processing (ICIAP)*, pp. 506–511.
- Chen, C.-H.; Chang, Y.-C.; Chen, T.-Y. e Wang, D.-J. (2008). People counting system for getting in/out of a bus based on video processing. In *International Conference on Intelligent Systems Design and Applications (ISDA)*, pp. 565–569.
- Chen, T.-Y.; Chen, T.-H. e Wang, D.-J. (2009). A cost-effective people-counter for passing through a gate based on image processing. *International Journal of Innovative Computing, Information and Control (ICIC International)*, 5(3):785–800.
- Duda, R. O.; Hart, P. E. e Stork, D. G. (2000). *Pattern Classification*. Wiley-Interscience, segunda edição.

- Elik, H.; Hanjalic, A. e Hendriks, E. (2006). Towards a robust solution to people counting. In *IEEE International Conference on Image Processing (ICIP)*, pp. 2401–2404.
- Digital Age Committee (2009). *Ensuring the Integrity, Accessibility and Stewardship of Research Data in the Digital Age*. National Academy Press.
- Fredman, M. L. e Tarjan, R. E. (1987). Fibonacci heaps and their uses in improved network optimization algorithms. *Journal of the ACM (JACM)*, 34(3):596–615.
- Gasparini, L.; Gottardi, M.; Massari, N.; Petri, D. e Manduchi, R. (2011). FPGA implementation of a people counter for an ultra-low-power wireless camera network node. In *Conference on Ph.D. Research in Microelectronics and Electronics (PRIME)*, pp. 169–172.
- Gonzalez, R. C.; Woods, R. E. e Eddins, S. L. (2009). *Digital Image Processing Using MATLAB, 2nd ed.* Gatesmark Publishing, Upper Saddle River, NJ, USA, segunda edição.
- Hsieh, J.-W.; Peng, C.-S. e Fan, K.-C. (2007). Grid-based template matching for people counting. In *IEEE Workshop on Multimedia Signal Processing (MMSP)*, pp. 316–319.
- Huang, D. e Chow, T. W. S. (2003). A People-Counting System Using a Hybrid RBF Neural Network. *Neural Processing Letters*, 18:97–113.
- Jungnickel, D. (2007). *Graphs, Networks and Algorithms*. Springer Publishing Company, Incorporated, terceira edição.
- Kilambi, P.; Ribnick, E.; Joshi, A. J.; Masoud, O. e Papanikolopoulos, N. (2008). Estimating pedestrian counts in groups. *Computer Vision and Image Understanding*, 110(1):43–59.
- Kuhn, H. W. (1955). The Hungarian Method for the assignment problem. *Naval Research Logistics Quarterly*, 2:83–97.
- Li, J.; Huang, L. e Liu, C. (2011). Robust people counting in video surveillance: Dataset and system. In *2011 8th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*, pp. 54–59.
- Mukherjee, S.; Saha, B.; Jamal, I.; Leclerc, R. e Ray, N. (2011). A novel framework for automatic passenger counting. In *IEEE International Conference on Image Processing (ICIP)*, pp. 2969–2972.
- Septian, H.; Tao, J. e Tan, Y.-P. (2006). People counting by video segmentation and tracking. In *International Conference on Control, Automation, Robotics and Vision (ICARCV)*, pp. 1–4.

- Silva, L. S. (2008). Sistema computacional para contagem automática de pessoas baseado em análise de sequências de imagens. Dissertação de mestrado, pg. 21, Programa de Pós-Graduação do Centro Federal de Educação Tecnológica de Minas Gerais.
- Velipasalar, S.; Tian, Y.-L. e Hampapur, A. (2006). Automatic counting of interacting people by using a single uncalibrated camera. In *IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1265–1268.
- Xu, X.-W.; Wang, Z.-Y.; Liang, Y.-H. e Zhang, Y.-Q. (2007). A rapid method for passing people counting in monocular video sequences. In *IEEE International Conference on Machine Learning and Cybernetics (ICMLC)*, volume 3, pp. 1657–1662.
- Yu, H.; He, Z. e Liu, J. (2007). A vision-based method to estimate passenger flow in bus. In *IEEE International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, pp. 654–657.
- Yu, S.; Chen, X.; Sun, W. e Xie, D. (2008). A robust method for detecting and counting people. In *IEEE International Conference on Audio, Language and Image Processing (ICALIP)*, pp. 1545–1549.