

3

Solução Proposta

A solução proposta para atender os objetivos desta dissertação consiste do seguinte conjunto de etapas seqüenciais.

- Captura da imagem e pré-processamento da imagem.
- Separação do fundo / primeiro plano.
- Atualização da estimativa do fundo.
- Localização / contagem de pessoas.
- Detecção de atitudes suspeitas.

Cada um destes passos está explicado nas seções a seguir.

3.1.

Captura e Pré-Processamento da Imagem

Para que o método proposto possa funcionar corretamente é necessário que as imagens a serem processadas sejam coloridas e estejam no sistema de coordenadas de cores RGB. Fora isto, outras exigências em relação à câmera a ser utilizada são dispensáveis, uma vez que todo o processo foi avaliado com imagens obtidas a partir de uma webcam convencional com captura de apenas 4 a 5 quadros por segundo. Ou seja, dentro desta realidade, qualquer câmera de baixo custo colorida com capacidade maior ou igual a 4 quadros por segundo pode ser utilizada.

Em relação ao pré-processamento das imagens capturadas, sabemos que existem vários sistemas de cores e que tais sistemas possuem características próprias e apresentam as informações de cores de uma forma particular que os torna mais ou menos convenientes, dependendo da aplicação.

Alguns sistemas de cores reúnem numa única componente a informação de brilho, deixando para as demais componentes a informação de cromaticidade (como, por exemplo, os sistemas de coordenadas de cores NTSC, HSV, YCbCr). Esta característica é explorada neste trabalho no tratamento de sombras. De fato, verificou-se experimentalmente, que vários destes sistemas de cores, se adequariam aos propósitos do trabalho.

Durante o desenvolvimento do sistema a análise de desempenho de alguns algoritmos foi feita inicialmente de maneira visual, usando a percepção humana para avaliar os resultados intermediários, portanto de forma subjetiva.

Conforme Forsyth e Ponce [FORSYTH 03], *“um espaço de cores é uniforme, quando as distâncias nas coordenadas do espaço é uma boa medida da significância da diferença entre duas cores – neste espaço; se a distância nas coordenadas do espaço estiver abaixo de certo limiar, um observador humano não seria capaz de dizer que se tratam de cores distintas”*.

De acordo com esta definição, percebe-se que, uma vez que a análise de desempenho na etapa de escolha do sistema de cores foi visual, um sistema de cores uniforme seria a melhor opção a se adotar. Com base nisto, a escolha do espaço de cores recaiu sobre o sistema CIE Lab [FORSYTH 03], por, atualmente, se tratar do sistema ou espaço de cores uniforme mais popular, além de concentrar a informação de brilho numa única componente.

Independente deste fato, foi feita uma comparação dos resultados obtidos nas etapas iniciais com três sistemas de cores diferentes, foram eles: o NTSC, o YCbCr e o CIE Lab. Aliando a tendência descrita anteriormente com o fato de que os melhores resultados foram observados com o sistema CIE Lab, este foi o sistema escolhido. Uma descrição mais detalhada do CIE Lab é apresentada no Apêndice A.

Por fim, após mudança para o sistema de coordenadas de cores CIE Lab, um filtro mediana 3x3 bi-dimensional é aplicado sobre cada uma das componentes de cores das imagens capturadas para redução de ruídos.

3.2. Separação do Fundo / Primeiro Plano

Esta etapa consiste em distinguir na imagem, quais pixels representam o fundo e quais pixels representam o primeiro plano, sendo o primeiro plano considerado como tudo aquilo que não faz parte do fundo. Esta etapa produz uma máscara binária em que o fundo é representado pelo bit “verdadeiro” e o primeiro plano é representado pelo bit “falso”.

A dificuldade para separar o primeiro plano do fundo decorre de dois problemas principais.

- i) Sombras.
- ii) Ruídos.

O método proposto trata destas dificuldades, conforme descrito a seguir.

3.2.1. Tratamento de Sombras

Antes de descrever o método implementado para o tratamento de sombras, é importante observar que não bastaria, por exemplo, agregá-las às pessoas, uma vez que, não necessariamente, os pixels relativos às sombras estarão conectados aos pixels relativos às pessoas. Ou seja, a sombra projetada por uma pessoa pode não estar próxima a ela, o que indicaria a existência de uma outra pessoa no local, o que seria um erro, pois na verdade se trata apenas de uma sombra. Portanto, um tratamento diferenciado para as sombras se faz necessário.

Considere que a matriz $\mathbf{I}_t(x,y)$ denota o valor do pixel nas coordenadas (x,y) do quadro da seqüência de imagens capturado no instante t . A notação aqui utilizada (negrito) enfatiza que o valor de cada pixel é dado por um **vetor** num espaço de cores, onde o brilho é representado por uma única componente ($I_t^b(x,y)$) e a cromaticidade pelas componentes restantes ($\mathbf{I}_t^c(x,y)$). Considere também a existência de uma estimativa de fundo. De forma análoga, $\mathbf{B}_t(x,y)$, $B_t^b(x,y)$ e $\mathbf{B}_t^c(x,y)$ denotam respectivamente esta estimativa de fundo e suas respectivas componentes de brilho e cromaticidade no instante t .

Utiliza-se o método convencional de subtração de imagem, onde uma matriz booleana $M_t(x,y)$ terá o valor “verdadeiro” nas posições (x,y) da imagem corrente correspondentes ao fundo. Deste modo serão fundo ($M_t(x,y) = \text{verdadeiro}$) os pixels para os quais o módulo da diferença entre a imagem e a estimativa de fundo é inferior a um dado limiar T_L , conforme mostrado em pseudocódigo na Figura 3:

<p>se $\mathbf{I}_t(x,y) - \mathbf{B}_t(x,y) < T_L$ então $M_t(x,y) = \text{verdadeiro}$ senão $M_t(x,y) = \text{falso}$</p>

Figura 3: Algoritmo convencional de separação do fundo / primeiro plano.

O algoritmo anterior apresenta desempenho insatisfatório em áreas de sombra. Nestas áreas a diferença entre a imagem e a estimativa de fundo pode ser significativa (superior ao limiar T_L), embora esta área da imagem seja do fundo.

Por isto, o método convencional de subtração foi modificado incorporando a observação de Prati [PRATI 01], segundo a qual, “nas áreas de sombra a maior parte da diferença entre a imagem e a estimativa do fundo recai sobre a componente de brilho”. Kumar [KUMAR 03] também propõe algumas modificações na subtração convencional baseado nesta observação, entretanto torna o algoritmo mais complexo desnecessariamente.

Seguindo o pensamento de Prati [PRATI 01], o algoritmo proposto considera ainda como fundo os pixels em que a razão entre o módulo da diferença de cromaticidade e o módulo da diferença total entre imagem e estimativa de fundo é inferior a um dado limiar. Para evitar o inconveniente de uma possível divisão por zero nesta razão, considera-se apenas o módulo da diferença de cromaticidade menor que um determinado parâmetro T_C . Com isto, o algoritmo da Figura 3 se transforma no algoritmo apresentado na Figura 4 adiante:

<p>se $I_t(x, y) - B_t(x, y) < T_L$ então $M_t(x, y) = \text{verdadeiro}$ senão se $I_t(x, y) - B_t(x, y) < T_H$ E $I_t^c(x, y) - B_t^c(x, y) < T_C$ então $M_t(x, y) = \text{verdadeiro}$ senão $M_t(x, y) = \text{falso}$</p>
--

Figura 4: Algoritmo de separação do fundo / primeiro plano com tratamento de sombras.

A primeira condição “se” captura os pixels mais similares ao fundo. Já a segunda condição, que incorpora a observação de Prati [PRATI 01], tolera uma disparidade maior entre a imagem analisada e a estimativa de fundo, caso a diferença nas componentes de cromaticidade continuem moderadas. Com isto, os pixels relativos a sombras são identificados. Cabe destacar que, para que isto seja possível, é necessário um sistema de coordenadas de cores onde o brilho se expresse em apenas uma componente de cor, como é o caso do sistema escolhido.

O algoritmo supõe a existência de uma boa estimativa do fundo a cada instante t . O procedimentos relacionados com a obtenção e com a atualização desta estimativa serão abordados na seção 3.3. Os valores para os parâmetros T_L , T_H ($T_L < T_H$) e T_C são obtidos experimentalmente.

Na Figura 5 pode-se observar os resultados desta etapa de separação do fundo / primeiro plano passo a passo. Na Figura 5a tem-se a imagem de entrada,

já na Figura 5b está apresentada a matriz lógica obtida após a aplicação do primeiro “se” do algoritmo da Figura 4. Nesta imagem as áreas em branco representam o fundo encontrado. Na seqüência, a Figura 5c mostra a mesma matriz lógica, porém obtida depois da aplicação do segundo “se” do algoritmo mostrado na Figura 4, ou seja, após a aplicação do algoritmo completo. Nesta imagem, o fundo apresentado em branco já considera os pixels relativos a sombras.

Comparando as Figura 5b e Figura 5c, observa-se, na parte destacada abaixo dos pés da pessoa localizada à esquerda da imagem, um exemplo de remoção de pixels relativos a sombras. Por fim, a Figura 5d mostra o resultado de toda a etapa de separação de fundo / primeiro plano, incluindo o tratamento de ruídos. Um detalhamento sobre os procedimentos realizados para a obtenção desta última imagem é apresentado na próxima seção.

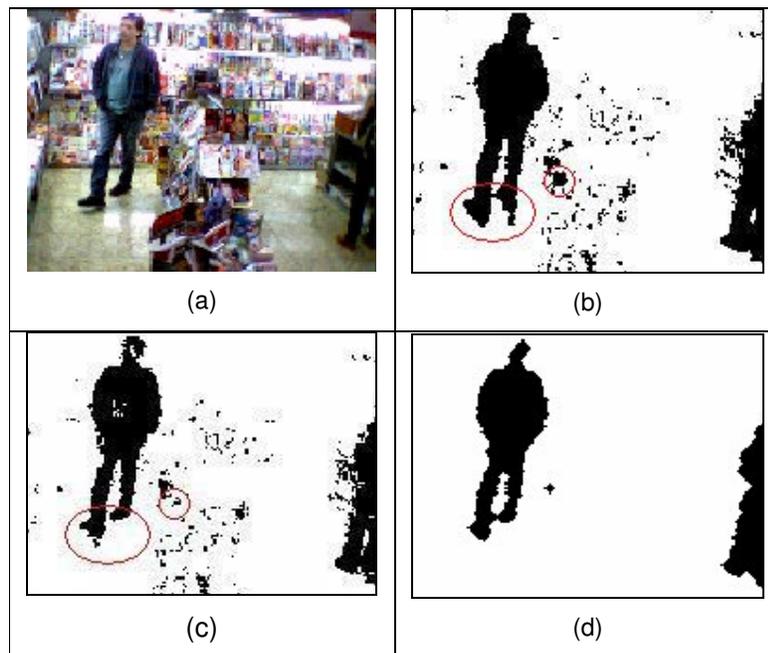


Figura 5: Resultados do algoritmo de separação do fundo / primeiro plano: (a) imagem de entrada; (b) pixels relativos ao fundo; (c) pixels relativos ao fundo e a sombras; (d) pixels relativos ao fundo e a sombras após operações morfológicas.

3.2.2. Tratamento de Ruído

Apesar da filtragem aplicada na fase de pré-processamento, que visa reduzir o efeito de ruído sobre a imagem de entrada, nota-se, ainda, que após o

tratamento de sombras algumas regiões pequenas permanecem. Tais regiões são ainda consequência de ruído na imagem de entrada.

Para eliminar estas ocorrências a matriz $M_t(x,y)$ é corrigida através da aplicação do algoritmo “abre-fecha” [PATTERN 00] descrito na Equação 1 abaixo.

$$M_t = (M_t \circ SE) \bullet SE \quad (1)$$

A operação de abertura (denotada pelo símbolo \circ) é definida como a erosão seguida da dilatação, e a operação de fechamento (denotada pelo símbolo \bullet) é definida como a dilatação seguida da erosão. SE denota o elemento estrutural usado nas operações. Nos experimentos descritos na próxima seção utilizou-se como elemento estrutural um disco de raio 2.

Por fim, uma operação para remoção de pixels isolados é aplicada. Isto completa o procedimento de tratamento de ruído nesta etapa. O efeito final pode ser notado comparando a Figura 5c com a Figura 5d.

3.3. Atualização da Estimativa do Fundo

O procedimento de separação de fundo / primeiro plano pressupõe a existência de uma boa estimativa de fundo. Esta seção descreve como estas estimativas de fundo são geradas e atualizadas a cada novo quadro analisado da seqüência de imagens. O processo aqui apresentado procura atender os seguintes objetivos.

- i) Permitir que o fundo seja modelado mesmo na presença de pessoas.
- ii) Tratar variações de iluminação no ambiente.
- iii) Incorporar alterações permanentes no fundo.

A primeira estimativa de fundo, utilizada quando o sistema é iniciado na separação de fundo / primeiro plano, deve ser providenciada de alguma forma. Pode ser uma imagem armazenada na memória do sistema, ou uma imagem capturada como parte dos procedimentos iniciais da operação do sistema.

Partindo desta primeira estimativa, um algoritmo de atualização dinâmica do fundo irá corrigindo-a, quadro a quadro, inclusive na presença de pessoas em movimento (i), de forma a tratar flutuações de iluminação (ii), e a incorporar modificações permanentes no fundo (iii).

O objetivo (ii) consiste em refletir na estimativa de fundo as variações de iluminação ambiente ao longo do tempo. No método aqui proposto admite-se que a variação de iluminação é lenta, ou seja, que a diferença no padrão de iluminação entre dois quadros consecutivos é pequena. Desta forma, toda vez que a diferença entre o quadro e a estimativa de fundo correntes, observada em um pixel, for pequena, a estimativa de fundo é corrigida ao longo do tempo, através da mistura entre o valor do pixel na imagem e o valor do pixel na estimativa de fundo.

Para isso, pode-se usar a matriz $M_t(x,y)$ obtida no passo anterior, que já contém a informação a cada quadro, dos pixels em que a estimativa de fundo deve ser corrigida. Esta correção só ocorre quando o pixel é considerado como fundo pelo procedimento de separação de fundo, descrito na seção 3.2. Desta forma, a atualização do fundo tolera a passagem de pessoas na imagem, como requer o objetivo (i). Quando uma pessoa se coloca diante do fundo, a correção do fundo é suspensa nos pixels relativos à pessoa, e é retomada assim que a pessoa se retira e o fundo fica novamente aparente.

Quando um pixel não for classificado como fundo por $K_{m\acute{a}x}$ quadros consecutivos, este pixel será considerado parte de uma alteração permanente (objetivo iii). Neste caso a nova estimativa de fundo passa a ser o valor do pixel no quadro corrente.

O algoritmo completo é apresentado em pseudocódigo na Figura 6.

```

se  $M_t(x, y) = \text{verdadeiro}$ 
  então  $B_{t+1}(x, y) = (1 - p) \cdot B_t(x, y) + p \cdot I_t(x, y)$ 
senão se  $|I_t(x, y) - I_{t-1}(x, y)| < e_b$ 
  então  $K(x, y) = K(x, y) + 1$ 
  se  $K(x, y) = K_{m\acute{a}x}$ 
    então  $B_{t+1}(x, y) = I_{t+1}(x, y)$ 
     $K(x, y) = 0$ 
  senão  $K(x, y) = 0$ ;

```

Figura 6: Algoritmo de atualização da estimativa do fundo.

Para os pixels da imagem corrente ($I_t(x,y)$) pertencentes ao fundo ($M_t(x,y) = \text{verdadeiro}$) a estimativa de fundo do próximo instante será dada por uma mistura entre a estimativa corrente ($B_t(x,y)$) e o valor do pixel da própria imagem ($I_t(x,y)$). O parâmetro p , que pode assumir valores entre 0 e 1, define o peso do novo valor do pixel no cálculo da nova estimativa de fundo. Visto de outra forma, p

estabelece a velocidade com que mudanças no fundo são assimiladas na estimativa.

Já para os pixels do primeiro plano ($M_i(x,y) = \text{falso}$), a estimativa de fundo será corrigida com o valor no quadro corrente, caso permaneçam estáveis por vários quadros consecutivos. Em outras palavras, se as mudanças de um pixel do primeiro plano entre quadros consecutivos for pequena, como dado pelo parâmetro e_b , durante $K_{máx}$ quadros consecutivos, a estimativa do fundo para este pixel receberá o valor do pixel no último quadro.

O algoritmo apresentado na Figura 6 contém a matriz $K(x,y)$ onde o número de quadros consecutivos sem mudanças significativas é armazenado para cada pixel. No começo do processo, $K(x,y)$ inicia com o valor zero e sempre que um pixel do primeiro plano sofrer mudanças grandes em relação ao quadro imediatamente anterior, $K(x,y)$ será zerado novamente.

O número máximo de quadros consecutivos $K_{máx}$, assim como os parâmetros p e e_b , devem ser definidos empiricamente para a aplicação.

3.4. Localização / Contagem de Pessoas

Uma vez que fundo e primeiro plano foram separados, o sistema deve encontrar e contabilizar as pessoas que se encontram no primeiro plano. Nesta etapa é importante que os seguintes objetivos sejam levados em consideração.

- i) Tratamento de grupos.
- ii) Tratamento de oclusões parciais.

A localização / contagem das pessoas é implementada através das seguintes etapas seqüenciais.

- Segmentação do primeiro plano.
- Agrupamento das regiões de interesse.

A seguir, cada uma destas etapas é detalhada.

3.4.1. Segmentação do Primeiro Plano

Esta seção descreve o procedimento de segmentação do primeiro plano adotado neste trabalho. O algoritmo proposto resulta em segmentos homogêneos e com dimensões reduzidas.

Diversas soluções propostas na literatura poderiam produzir resultados igualmente satisfatórios. Em particular o algoritmo do tipo crescimento de região

proposto em [BAATZ 00] parece bastante promissor em virtude de considerar conjuntamente cor e dimensão no critério para agregar sub-regiões. Entretanto, a implementação disponível no momento da realização deste trabalho é computacionalmente muito ineficiente, de modo que se optou pela alternativa descrita a seguir.

O método baseia-se no algoritmo de watershed, cuja descrição pode ser encontrada em vários livros textos como, por exemplo, em [GONZALE 02]. Um número especial de *Pattern Recognition* foi dedicado ao tema e [PATTERN 00] é uma excelente referência.

Para produzir resultados satisfatórios, a imagem a ser segmentada deve ser adequadamente processada antes de ser submetida ao algoritmo básico de watershed. Esta seção descreve o processamento implementado no protótipo construído para validação do método proposto (vide capítulo 4). O processamento se dá através de 5 passos seqüenciais, conforme descrito a seguir e ilustrado pela Figura 7. Embora o procedimento adiante seja aplicado apenas ao primeiro plano, a Figura 7 mostra os resultados intermediários do procedimento aplicado a toda a imagem de entrada, para facilitar a visualização.

- a) Inicialmente, deve-se buscar a imagem do primeiro plano (Figura 7a) após a aplicação do filtro de suavização feita na etapa de pré-processamento (Figura 7b).
- b) Em seguida, a magnitude do gradiente é calculada, para cada componente, utilizando a máscara de convolução Sobel [GONZALE 02].
- c) Calcula-se a média das magnitudes dos gradientes nas três componentes de cada pixel (Figura 7c e Figura 7d).
- d) A matriz com a média das magnitudes dos gradientes é modificada, de modo a criar linhas de cristas artificiais que formam uma grade com malhas de lados iguais a $\sqrt[2]{A_{máx}}$. Atribui-se aos pixels localizados sobre a grade o maior valor da matriz média (Figura 7e e Figura 7f). Isso garante que todos segmentos produzidos mais adiante terão área não superior a $A_{máx}$. A razão para limitar o tamanho dos segmentos é apresentada na seção 3.4.2.3.
- e) Aplica-se à matriz resultante a transformação conhecida como supressão de mínimos [SOELLI 03]. Esta transformação suprime os mínimos locais que não são mais profundos do que um parâmetro H , definido experimentalmente pelo usuário (Figura 7g e Figura 7h).

O objetivo disto é evitar o efeito de super segmentação que ocorreria, caso o algoritmo de segmentação *watershed* fosse aplicado diretamente sobre a matriz da média.

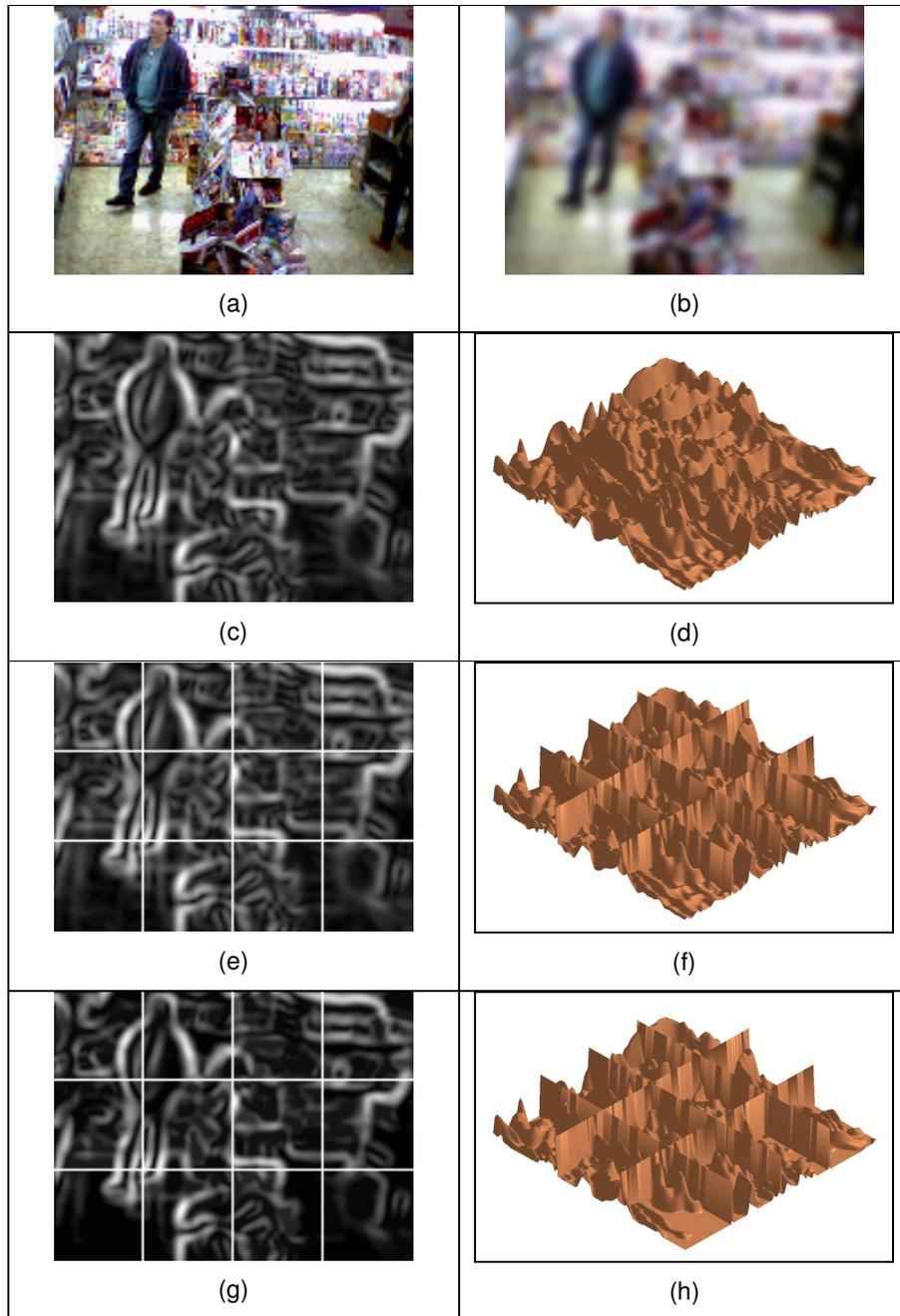


Figura 7: Processamento da imagem antes de ser submetida ao algoritmo básico de *watershed*: (a) imagem de entrada; (b) após suavização; (c) média do gradiente nas três componentes de cores em 2D (d) e em 3D; (e) após superposição da grade em 2D (e) e em 3D (f); após supressão de mínimos em 2D (g) e em 3D (h).

O algoritmo básico de watershed é então aplicado à matriz bi-dimensional produzida pela supressão de mínimos. A Figura 8 mostra um exemplo do resultado final da segmentação do primeiro plano.

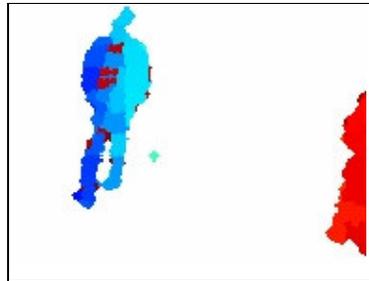


Figura 8: Resultado produzido pela segmentação do primeiro plano.

3.4.2. Agrupamento das Regiões de Interesse

O procedimento aplicado neste trabalho para localização / contagem de pessoas no primeiro plano se baseia numa técnica chamada coerência de movimento descrita em Shapiro [SHAPIRO 01]. Esta técnica foi escolhida, pois apresenta vantagens sobre os outros métodos apresentados no capítulo 2. Uma das vantagens que pode ser citada diz respeito à suposição de formas e/ou tamanhos fixos que é necessária em alguns dos trabalhos anteriores. Através desta técnica, isto não é preciso, o que faz com que o sistema seja mais tolerante a movimentos que envolvem inclinações ou agachamentos e a pessoas de alturas diferentes, mais gordas ou mais magras.

O método parte de três hipóteses principais.

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

Antes de partir para a descrição do método, convém definir o que se chamará daqui para diante neste texto de **região de interesse**. A região de interesse relativa a um segmento é a área da imagem de entrada delimitada pelo retângulo envolvente mínimo (bounding box) que contém o segmento. Associa-se ainda a cada região de interesse uma máscara lógica com número de linhas e colunas igual ao da região de interesse e que tem o valor “verdadeiro” nas

posições do retângulo envolvente mínimo ocupadas pelo segmento, e, “falso” nas demais posições.

Cabe destacar que neste aspecto o método proposto difere do método usado por Shapiro [SHAPIRO 01], uma vez que este trabalha com pontos de interesse ao invés de regiões de interesse. Esta seria uma vantagem sobre o método de Shapiro, pois com regiões de interesse a determinação de áreas correspondentes é mais eficiente, já que leva em conta uma vizinhança maior.

O método consiste de três etapas seqüenciais.

- a) Localizar para cada segmento do primeiro plano de um quadro, o segmento correspondente do quadro anterior com forma igual e que lhe seja mais similar (*best match*).
- b) Calcular o vetor de movimento entre quadros consecutivos para cada segmento do primeiro plano.
- c) Determinar os grupos de segmentos adjacentes com vetores de movimento semelhantes; cada grupo corresponderá em princípio a uma pessoa na cena.

Cada uma destas etapas é detalhada nas seções seguintes.

3.4.2.1. Localização de Segmentos Correspondentes

Para uma melhor eficiência computacional, a procura pelo segmento correspondente é restrita a uma área em volta da região de interesse em questão, como ilustrado na Figura 9. O centro da área de procura coincide com o centro da área de interesse. Suas dimensões são determinadas pelo deslocamento máximo ($R_{máx}$) que um segmento pode sofrer em quadros consecutivos, conforme a segunda hipótese (ii) em que se baseia o método.

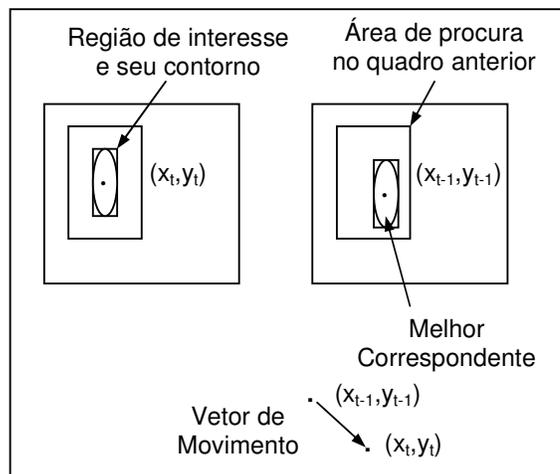


Figura 9: Correspondência das regiões de interesse.

Várias medidas de similaridade podem ser adotadas nesta busca. Neste trabalho a similaridade entre uma região de interesse e uma possível região correspondente na área de busca do quadro anterior é medida da maneira que se segue.

- Primeiramente, computa-se o módulo da diferença de cores entre todos os pares de pixels correspondentes.
- Calcula-se, em seguida, o valor do percentil π_i dos valores produzidos. Somente as posições indicadas na máscara associada à região de interesse entram neste cálculo.
- Associa-se, assim, a cada posição da área de busca um valor de similaridade.
- Se o menor valor de percentil encontrado estiver abaixo de um certo valor π_{\max} , pré-definido pelo usuário, encontrou-se o segmento correspondente.

Foram feitos testes utilizando correlação cruzada, como no método proposto por Shapiro [SHAPIRO 01], porém os resultados obtidos foram inferiores aos do método do percentil.

Para algumas regiões de interesse não são encontradas regiões correspondentes. Esta situação corresponde aos casos em que o menor valor de percentil encontrado supera π_{\max} . Tais casos serão tratados no momento em que as regiões de interesse são agrupadas, conforme descrito mais adiante.

O procedimento aqui detalhado pode ser aplicado utilizando todas as três componentes ou apenas uma ou duas componentes de cores. De fato, os experimentos relatados no capítulo 4, indicaram que a componente de brilho apenas já produz resultados satisfatórios.

3.4.2.2. Determinação dos Vetores de Movimento

De posse dos segmentos e de seus correspondentes no quadro anterior, os vetores que definem o movimento de cada região de interesse de um quadro para o outro podem ser computados. Sendo (x_t^i, y_t^i) e (x_{t-1}^i, y_{t-1}^i) as coordenadas dos centros da região de interesse do segmento s_i e da sua região correspondente no quadro anterior, respectivamente, conforme ilustrado na Figura 9, o vetor de movimento relativo a esta região será calculado conforme a Equação 2 abaixo.

$$mv_i = (x_t^i - x_{t-1}^i, y_t^i - y_{t-1}^i) \quad (2)$$

3.4.2.3. Agrupamento dos Segmentos

Finalmente, os segmentos devem ser reunidos em grupos que representam cada um uma pessoa. Para serem agrupados dois segmentos devem atender a algumas condições.

- Devem ser adjacentes.
- Devem ter movimentos coerentes do quadro anterior para o atual.
- Não podem ter sido atribuídos a grupos distintos no quadro anterior.

Os detalhes do algoritmo para formação dos grupos são apresentados na Figura 10 .

Note como o algoritmo trata uma região de interesse que não possui região correspondente no quadro anterior. Evidentemente não será possível calcular seu vetor de movimento e, portanto, a condição de movimento coerente não poderá ser verificada. Nestas situações, considera-se apenas a condição de adjacência. Em outras palavras, caso uma região de interesse sem correspondente no quadro anterior seja adjacente a um grupo que está sendo formado, este segmento é agregado ao grupo.

Desta maneira, ao final deste processo, cada grupo formado irá corresponder a um conjunto de regiões de interesse coerentes e adjacentes ou apenas adjacentes, que serão consideradas como pertencentes a uma única pessoa.

Apesar dos procedimentos descritos, permanecem, ainda, pequenos segmentos devido a sombras, a ruídos ou a mudanças na disposição de objetos do fundo. Para que tais segmentos não interfiram na correta contabilidade do número de pessoas, só são considerados como representando uma pessoa os grupos cuja área em número de pixels for superior a um parâmetro S_{min} definido pelo usuário.

Com isto, ao final desta etapa tem-se o número de pessoas presentes no ambiente em cada quadro e o posicionamento de cada uma delas. A Figura 11 mostra o resultado da aplicação do procedimento apresentado aos segmentos da Figura 8.

/ \mathbf{U} é o conjunto de todos os segmentos encontrados no passo anterior.
 / \mathbf{S} é o conjunto dos segmentos não agrupados.
 / \mathbf{G}_g é o g -ésimo grupo formado pelo algoritmo.
 / s_j, s_k são segmentos encontrados no passo anterior que ainda não foram agrupados.
 / L_j^A, L_k^A são os rótulos associados aos segmentos s_j, s_k no quadro anterior.
 / L_j^C, L_k^C são os rótulos associados aos segmentos s_j, s_k no quadro atual (corrente).
 / mv_j, mv_k são vetores de movimento associados aos segmentos s_j e s_k .
 / $D_{máx}$ é a distância máxima que evidencia a coerência do movimento de dois vetores.
 $\mathbf{S} = \mathbf{U}$;
 $g = 0$;
enquanto $\mathbf{S} \neq \emptyset$
 $g = g + 1$; / g recebe novo rótulo.
 $\mathbf{G}_g = s_j \in \mathbf{S}$; / s_j é um segmento qualquer ainda não agrupado.
 $\mathbf{S} = \mathbf{S} - s_j$; / s_j é retirado do conjunto de segmentos não agrupados.
 enquanto $\exists s_k \in \mathbf{S}$ tal que
 $(s_k \text{ é adjacente a } s_j \in \mathbf{G}_g) \text{ E}$
 $[(L_k^A = L_j^A \text{ E } |mv_k - mv_j| < D_{máx}) \text{ OU}$
 $s_k \text{ não tem correspondente no quadro anterior}]$
 $\mathbf{G}_g = \mathbf{G}_g \cup s_k$; / s_k é agregado ao grupo \mathbf{G}_g .
 $\mathbf{S} = \mathbf{S} - s_k$; / s_k sai do conjunto de segmentos não agrupados.
 $L_{k=g}^C$; / s_k recebe o rótulo g .

 $L_j^A = L_j^C, \forall j$ tal que $s_j \in \mathbf{U}$

Figura 10: Algoritmo para agrupamento de segmentos.

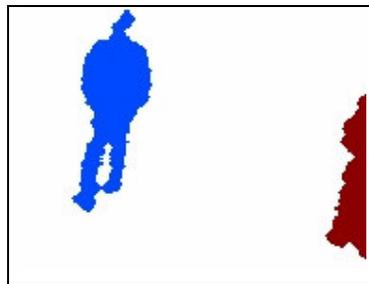


Figura 11: Resultado da agregação dos segmentos da Figura 8.

Neste ponto fica mais fácil entender a relevância de se impor um limite superior para a área dos segmentos do primeiro plano, como descrito na seção 3.4.1. Considere uma situação em que tronco e um dos braços formam dois grandes segmentos. Eventualmente o braço pode ter um movimento muito distinto do tronco, que pode até nem se mover em dois quadros consecutivos. Nesta situação os vetores de movimentos poderiam ser muito diferentes e, pelo critério da coerência de movimento, dariam origem a dois grupos e, portanto, a pessoas distintas.

Colocando um limite à área, os segmentos adjacentes de uma pessoa teriam de fato movimentos mais coerentes. Voltando ao exemplo citado, um segmento do ombro adjacente a um segmento do braço (ambos menores do que no caso anterior) se movem de modo mais coerente que um segmento relativo ao tronco e outro ao braço. Esta coerência tem mais chance de se manter ao longo de cada segmento adjacente do braço, de modo que ao final todo o tronco e todo o braço acabariam sendo agregados num único grupo. Note que, neste caso, o critério para sub-dividir um segmento homogêneo em segmentos menores é pouco relevante.

3.5. Detecção de Atitudes Suspeitas

As atitudes suspeitas que se pretende detectar são roubo e colocação de um objeto novo na cena. Ambas as situações se manifestam numa alteração duradoura na imagem. Esta condição já foi considerada no algoritmo para atualizar as estimativas de fundo, como visto na seção 3.3.

Seguindo o mesmo raciocínio, se uma área da imagem, distinta da estimativa de fundo permanece estável por $K'_{m\acute{a}x}$ ($K'_{m\acute{a}x} \leq K_{m\acute{a}x}$) quadros consecutivos, tem-se nesta área uma alteração suspeita. O valor de $K'_{m\acute{a}x}$ reflete o tempo que o sistema levará para detectar a situação suspeita e deve ser definido a partir das características da aplicação, em particular das exigências de segurança. Cabe lembrar que, $K'_{m\acute{a}x}$ muito baixo poderá resultar em um número elevado de alarmes falsos.

Num sistema de segurança real, pode-se impor condições adicionais para o acionamento de um alarme que levem em conta, por exemplo, o tamanho ou a forma do objeto em análise.