



INSTITUTO SUPERIOR TÉCNICO
Universidade Técnica de Lisboa

Detecção de Objectos Activos em Ambientes Dinâmicos

Sérgio Miguel dos Santos Copeto

Dissertação para a Obtenção do Grau de Mestre em
Engenharia Electrotécnica e de Computadores

Júri

Presidente: Prof Carlos Filipe Gomes Bispo
Orientador: Prof Jorge dos Santos Salvador Marques
Co-Orientador: Prof João Manuel Lage de Miranda Lemos
Vogal: Prof Alexandre José Malheiro Bernardino

Abril de 2012

Agradecimentos

Em primeiro lugar, gostaria de agradecer ao meu orientador, o professor Jorge Salvador Marques, por ter aceite a minha proposta de tema e sobretudo pelo apoio prestado ao longo desta tese. Ao meu co-orientador, o professor João Miranda Lemos, também pelo apoio prestado neste trabalho e também pela oportunidade que me deu em participar no projecto ARGUS, que me fez aprender imenso. Ainda nesse campo, gostaria de agradecer à Fundação para a Ciência e Tecnologia pela bolsa de investigação que me foi atribuída no âmbito do projecto ARGUS PTDC/EEA-CRO/098550/2008.

Tenho também que agradecer ao Doutor Miguel Lobo Antunes, ao Doutor Paulo Prata Ramos e a toda a equipa da Culturgest, que disponibilizou sem qualquer senão o grande auditório para a recolha dos vídeos de teste utilizados nesta tese, dando assim um grande contributo para a realização deste trabalho.

Um agradecimento especial a todos os meus amigos pelos incríveis momentos que passamos juntos, pelas inúmeras conversas e partilhas de ideias e sonhos, e por tudo aquilo que me traz memórias que ficarão para o resto da vida.

E é com o maior orgulho que dedico esta tese aos meus pais, à minha irmã e aos meus avós, que ao longo da vida me deram tudo o que precisava para construir o futuro que sempre desejei.

Resumo

Esta tese apresenta um sistema de detecção de objectos activos para sistemas de video-vigilância capaz de lidar com variações bruscas das condições de iluminação. O sistema proposto guarda múltiplas imagens de fundo, cada uma correspondendo a uma condição de iluminação diferente, sendo que o sistema identifica a condição de iluminação activa em cada instante de tempo. O sistema consegue ainda identificar condições de iluminação desconhecidas, adicionando-as à colecção de modelos de fundo, bem como actualizar os modelos existentes. No âmbito desta tese foi recolhido um conjunto de vídeos de teste, captados no palco do grande auditório da Culturgest. O sistema desenvolvido foi avaliado com base nestas sequências e usando diferentes espaços de cor, tendo-se escolhido os espaços RGB, HSV e $L^*a^*b^*$.

O sistema foi testado tanto na identificação das condições de iluminação activas em cada instante, como na segmentação de regiões activas com iluminação constante e variável. Foi ainda feita uma comparação com um modelo clássico, da qual se conclui que o sistema desenvolvido apresenta um desempenho superior, sobretudo no caso de variações bruscas das condições de iluminação.

Palavras chave: Sistemas de Video Vigilância, Ambientes Dinâmicos, Iluminação Variável

Abstract

This thesis presents an object detection system for video surveillance that can handle sudden illumination changes. The proposed system keeps multiple background models, each one representing an illumination condition, and the system performs the identification of the active illumination for each incoming frame. The system can also identify unknown illumination conditions and add them to the background model collection, as well as perform updates to the existing models. A set of test videos were recorded at the Culturgest auditorium in the scope of this thesis. The proposed system was evaluated on this data set, using different color representations, RGB, HSV and L*a*b*

The system was tested in terms of the active background model identification, as well as in the active region segmentation for both constant and varying illumination. A comparison with a classical model was also made. It was concluded that the proposed system performs better than the classical model, specially in the case of sudden light changes.

Keywords: Video Surveillance Systems, Dynamic Ambient, Sudden Illumination Changes.

Conteúdo

Agradecimentos	iii
Resumo	v
Abstract	vii
Lista de Figuras	xi
Lista de Tabelas	xiii
1 Introdução	1
1.1 Motivação	1
1.2 Formulação do Problema	1
1.3 Organização da Tese e Contribuições	2
2 Estado da Arte	3
2.1 Modelação de Imagens de Fundo	3
2.1.1 Modelação Estatística	4
2.2 Actualização do modelo da Imagem de Fundo	5
2.3 Segmentação de Objectos Activos por Subtracção de Fundo	6
2.4 Modelo de Mistura de Gaussianas	6
3 Sistema de Seguimento	9
3.1 Colecção de Imagens de Fundo	10
3.2 Gestão de Imagens de Fundo	11
3.2.1 Identificação de Modelos de Fundo	11
3.2.2 Actualização dos modelos de Imagens de Fundo	11
3.2.3 Adição de um Modelo de Imagem de Fundo	12
3.3 Segmentação de Regiões Activas	14

4	Conjunto de dados	15
4.1	Descrição Geral do Conjunto de Dados	15
4.2	Estabilidade da Cor no Tempo e no Espaço	17
4.3	Variância da imagem de fundo em Cada Intervalo Estacionário	23
4.4	Reacção à Presença de Objectos	29
5	Resultados Experimentais	31
5.1	Metodologia de Avaliação	31
5.2	Identificação de Imagens de Fundo	32
5.3	Detecção de Objectos Activos	36
5.3.1	Segmentação Com Iluminação Constante	36
5.3.2	Segmentação com Variações de Iluminação	38
5.4	Comparação com Modelo Clássico	40
6	Conclusão	45
6.1	Conclusões	45
6.2	Trabalho Futuro	46
	Bibliografia	47

Lista de Figuras

3.1	Diagrama de Blocos do Sistema a Implementar	10
4.1	Condições de iluminação usadas nos testes	16
4.2	Aparência das pessoas usadas nos testes	16
4.3	Descrição gráfica do conjunto de acontecimentos das sequências de vídeo adquiridas. Os valores entre 1 e 4 representam a condição de iluminação activa em cada instante, e os valores entre 5 e 6 indicam a ausência/presença de pessoas em cena.	17
4.4	Linhas escolhidas para o estudo da estabilidade da cor, representadas a branco.	18
4.5	Estabilidade da iluminação âmbar para os espaços de cor RGB (1ª linha), HSV (2ª linha) e L*a*b* (3ª linha). Os valores a vermelho, verde e azul representam o primeiro, segundo e terceiro canais de cor, respectivamente.	19
4.6	Estabilidade da iluminação azul para os espaços de cor RGB (1ª linha), HSV (2ª linha) e L*a*b* (3ª linha). Os valores a vermelho, verde e azul representam o primeiro, segundo e terceiro canais de cor, respectivamente.	20
4.7	Estabilidade da iluminação vermelha para os espaços de cor RGB (1ª linha), HSV (2ª linha) e L*a*b* (3ª linha). Os valores a vermelho, verde e azul representam o primeiro, segundo e terceiro canais de cor, respectivamente.	21
4.8	Estabilidade da iluminação verde para os espaços de cor RGB (1ª linha), HSV (2ª linha) e L*a*b* (3ª linha). Os valores a vermelho, verde e azul representam o primeiro, segundo e terceiro canais de cor, respectivamente.	22
4.9	Desvio padrão da imagem de fundo de iluminação âmbar. Escala de cor: azul escuro - valores pequenos, vermelho - valores elevados	25
4.10	Desvio padrão da imagem de fundo de iluminação azul. Escala de cor: azul escuro - valores pequenos, vermelho - valores elevados	26
4.11	Desvio padrão da imagem de fundo de iluminação vermelha. Escala de cor: azul escuro - valores pequenos, vermelho - valores elevados	27
4.12	Desvio padrão da imagem de fundo de iluminação verde. Escala de cor: azul escuro - valores pequenos, vermelho - valores elevados	28

4.13	Reacção à passagem de objectos na iluminação âmbar. Os valores a vermelho, verde e azul representam o primeiro, segundo e terceiro canais de cor, respectivamente.	29
4.14	Reacção à passagem de objectos na iluminação azul. Os valores a vermelho, verde e azul representam o primeiro, segundo e terceiro canais de cor, respectivamente.	30
4.15	Reacção à passagem de objectos na iluminação vermelha. Os valores a vermelho, verde e azul representam o primeiro, segundo e terceiro canais de cor, respectivamente.	30
4.16	Reacção à passagem de objectos na iluminação verde. Os valores a vermelho, verde e azul representam o primeiro, segundo e terceiro canais de cor, respectivamente.	30
5.1	Ground truth para os exemplos usados no teste de iluminação constante	32
5.2	Ground truth para os exemplos usados no teste de iluminação variável	32
5.3	Imagem de fundo identificada ao longo do tempo no primeiro teste. Os valores entre 1 e 4 representam o modelo de fundo seleccionado pelo algoritmo. O valor 0 representa a fase de aprendizagem de um novo modelo de fundo	34
5.4	Imagem de fundo indenticada ao longo do tempo no segundo teste. Os valores entre 1 e 4 representam o modelo de fundo seleccionado pelo algoritmo. O valor 0 representa a fase de aprendizagem de um novo modelo de fundo	35
5.5	Resultados da segmentação utilizando o sistema desenvolvido em condições de iluminação constante	37
5.6	Resultados da segmentação utilizando o sistema desenvolvido em condições de iluminação variável	39
5.7	Resultados da segmentação utilizando a Mistura de Gaussianas em condições de iluminação constante	41
5.8	Resultados da segmentação utilizando uma Mistura de Gaussianas em condições de iluminação variável	42

Lista de Tabelas

4.1	Média do Valor do Ruído nos Diferentes Espaços de Cor.	24
5.1	Matrizes de Confusão da segmentação usando o sistema implementado face a iluminação constante	43
5.2	Matrizes de Confusão da segmentação usando a Mistura de Gaussianas face a iluminação constante	43
5.3	Matrizes de Confusão da segmentação usando o sistema implementado face a variações de iluminação	43
5.4	Matrizes de Confusão da segmentação usando a Mistura de Gaussianas face a variações de iluminação	44

Capítulo 1

Introdução

1.1 Motivação

As sociedades modernas baseiam-se numa grande concentração de pessoas e meios em centros urbanos e grandes infra-estruturas, o que coloca grandes problemas de segurança. Este fenómeno levou ao aparecimento de sistemas de video-vigilância e à sua instalação em variados locais como bancos, casinos, aeroportos, centros comerciais e instalações militares. As aplicações são vastas, como a prevenção de crimes, controlo de processos industriais perigosos e segurança em transportes públicos. Os sistemas de vigilância podem ter uma maior ou menor dimensão, podendo ser compostos por uma única câmara, ou múltiplas câmaras ligadas em rede. As câmaras podem ser estáticas ou móveis, o que permite um campo de visão mais alargado. Com o avanço da tecnologia, estes sistemas tornaram-se mais inteligentes, conseguindo detectar objectos de interesse (por exemplo, malas abandonadas), detectar e seguir pessoas, fazer reconhecimento facial e detectar situações anómalas como circulação de veículos em contra-mão em auto-estradas.

Contudo, os sistemas de video-vigilância estão em constante evolução. É necessário que estes sejam mais robustos no que toca a falsas detecções, muitas vezes provocadas por alterações das condições de iluminação dos locais sob observação. Estas alterações podem manifestar-se de forma lenta, em resultado do movimento do sol ou das nuvens em cenas exteriores, como de forma rápida, como acontece quando se acendem ou apagam luzes numa cena interior.

1.2 Formulação do Problema

O objectivo desta tese é desenvolver um sistema de detecção de objectos robusto a variações bruscas de iluminação. Este problema é de grande importância por se tratar de um ponto de partida para a construção de sistemas mais fiáveis e com menos falhas. Esta tese propõe que sejam guar-

dadas múltiplas imagens de fundo, representando diferentes condições de iluminação possíveis, sendo possível adicionar novas imagens de fundo caso não as existentes não representem com suficiente rigor as condições de iluminação captadas num dado instante de tempo, eliminar imagens de fundo que não sejam utilizadas, e actualizar as imagens de fundo existentes. Desta forma, o sistema consegue adaptar-se não só a variações lentas de iluminação, como também adaptar-se a situações novas de forma rápida.

1.3 Organização da Tese e Contribuições

Seguidamente apresenta-se a estrutura desta dissertação. O Capítulo 2 apresenta uma introdução aos sistemas de video-vigilância, fazendo um resumo do estado da arte das ferramentas necessárias para este trabalho. O Capítulo 3 descreve de forma detalhada o trabalho desenvolvido, mostrando como são criados e geridos os modelos de fundo, a forma como a imagem de fundo é seleccionada em cada instante de tempo e o processo de segmentação das regiões activas encontradas.

Para esta tese foi adquirido um conjunto de 5 sequências de vídeo, captadas num palco de espectáculos gentilmente cedido pela Culturgest. O capítulo 4 apresenta o conjunto de dados recolhidos que são utilizados na avaliação do sistema desenvolvido no âmbito desta tese. Além disso, este capítulo faz uma análise detalhada das sequências de vídeo captadas, analisando a estabilidade da informação de cor ao longo do tempo, a forma como o ruído se distribui ao longo do espaço, bem como as alterações provocadas pela presença de objectos activos. Estes testes são realizados para 3 espaços de cor: RGB, HSV e L^*a^*b .

O Capítulo 5 apresenta os testes efectuados e os resultados experimentais obtidos. Neste capítulo é avaliada a forma como o sistema identifica as diferentes condições de iluminação presentes nas sequências de vídeo e os resultados obtidos na fase de segmentação de regiões activas, tanto no caso em que as condições de iluminação são invariantes ao longo do tempo, como também quando o sistema tem de lidar com alterações às condições de iluminação. Por último, é feita uma comparação do resultado da segmentação obtida com a segmentação obtida através da utilização de um modelo clássico, baseado em Mistura de Gaussianas [1].

As contribuições desta tese para o problema proposto são a aquisição de um conjunto de sequências de vídeo que permitam avaliar o desempenho de um sistema de video-vigilância face a variações bruscas das condições de iluminação, apresentando um estudo detalhado das mesmas sequências no que toca à estabilidade da cor e à presença de ruído para os diferentes espaços de cor em estudo. A contribuição principal é o desenvolvimento de um sistema robusto face a variações bruscas das condições de iluminação, sendo recolhidos e analisados os dados experimentais obtidos, e comparados com um modelo clássico e bem conhecido.

Capítulo 2

Estado da Arte

Este capítulo resume as várias técnicas de detecção de objectos activos em sistemas de video-vigilância baseados em câmeras estáticas. Seguidamente apresenta-se uma descrição geral do tema, após a qual se apresentam os métodos mais utilizados na resolução do problema proposto.

A detecção de objectos activos em sequências de vídeo é uma operação cada vez mais importante em sistemas de video-vigilância. Há diferentes métodos de segmentação que podem ser classificados em dois grandes grupos. O primeiro grupo utiliza um critério de homogeneidade espacial, nos quais são usados filtros morfológicos para detectar elementos de interesse na imagem, como o fluxo óptico [2] ou contornos activos [3]. Estas técnicas são capazes de detectar de forma robusta os limites dos objectos, mas a sua complexidade computacional torna-os pouco utilizados.

A segunda abordagem explora a detecção de variações em sequências de vídeo. Alguns algoritmos utilizam a diferença entre duas imagens consecutivas [4], mas o método mais comum é o método de subtracção de fundo (Background Subtraction). Este método consiste na comparação da imagem recebida em cada instante com uma imagem representativa do fundo da cena. Por serem robustos e rápidos a nível computacional, os métodos subtracção de fundo são amplamente utilizados neste tipo de sistemas, sendo também o tipo de métodos escolhido no âmbito desta tese.

As secções seguintes descrevem a forma como a imagem de fundo é estimada e a forma como os objectos activos são segmentados.

2.1 Modelação de Imagens de Fundo

Nas aplicações de video-vigilância que utilizam o método de subtracção de fundo, o primeiro passo é obter um modelo do cenário (chamada imagem de fundo), que será posteriormente utilizada para detecção de elementos activos. Vários métodos de modelação da imagem de fundo estão descritos na literatura. Seguidamente destacam-se algumas categorias [5]:

- Modelação Básica - Constitui o método mais simples, no qual a imagem de fundo é obtida através do cálculo da média ou a mediana de um conjunto de imagens obtidas em diferentes instantes de tempo. Este conjunto de treino pode ser actualizado periodicamente ao longo do tempo. [6]
- Modelação Estatística - Em que cada pixel da imagem de fundo é considerado como uma variável aleatória independente cuja distribuição de probabilidade pode ser modelada através de uma ou mais distribuições gaussianas ou por funções não-paramétricas (densidade de Kernel). [7]
- Modelação por Clustering - Assume-se que o valor de cada pixel da imagem é próximo de 1 de K valores, obtidos por clustering. Utilizando o algoritmo de K-médias é possível estimar as cores associadas a cada posição da imagem. [8]

Estes algoritmos têm que lidar com várias dificuldades. Por um lado, têm de ser robustos no que toca a compensar características como o ruído dos sensores. Contudo, as dificuldades principais estão relacionadas com os ambientes dinâmicos, ou seja, situações em que as características do ambiente sofrem grandes alterações. Em ambientes exteriores, estas alterações manifestam-se através da posição do sol (que afecta significativamente as condições de iluminação), o movimento das nuvens no céu (que pode ser mais ou menos rápido), a presença de sombras (cuja forma e intensidade varia consoante a posição do sol), a existência de vento (que pode fazer mover a câmara) e objectos móveis (p. ex., árvores ou bandeiras). Nos ambientes interiores, pode existir também a presença de luz natural (que afecta as condições de iluminação e a presença de sombras), mas há sobretudo o problema das mudanças bruscas das condições de iluminação (o acender ou apagar de luzes).

Tendo em conta as dificuldades descritas, escolheu-se a modelação estatística para a implementação deste trabalho. Este tipo de modelação permite não só compensar a presença de ruído, mas também permite lidar com o problema dos ambientes dinâmicos, através da descrição de múltiplas aparências possíveis para um dado pixel. Outro factor importante tem que ver com a forma como o modelo da imagem de fundo é actualizado, existindo várias formas de realizar essa tarefa, adaptando-se cada melhor consoante o tipo de ambiente em questão.

2.1.1 Modelação Estatística

O sinal de vídeo apresenta variações aleatórias ao longo do tempo, produzidas durante o processo de aquisição, mesmo quando a câmara e os objectos presentes na cena são estáticos. Esta propriedade sugere o uso de modelos probabilísticos para descrever a imagem de fundo. A modelação estatística engloba várias técnicas. A situação mais simples consiste em representar cada pixel da imagem como uma variável aleatória com distribuição normal (Gaussiana) com média e variância dependente do pixel [7]. A utilização de distribuições probabilísticas permite que o ruído da imagem seja representado através de uma variável aleatória cuja variância pode ser estimada. No entanto, este modelo apenas obtém

uma boa performance em cenas interiores, ou em cenas exteriores com condições estáveis de iluminação.

O modelo gaussiano tem limitações e foi estendido. Em [1] cada pixel é modelado através de uma mistura de Gaussianas com diferentes médias e variâncias. Esta técnica permite representar múltiplos valores comuns de um pixel, sendo utilizada em ambientes exteriores em que existem variações lentas e periódicas. Verifica-se também que este método causa falsos positivos na presença de ambientes dinâmicos, como por exemplo, quando existem folhas de árvores ou movimentos de água. Mais tarde, foi também estudada a possibilidade de utilizar o modelo de Mistura de Gaussianas em diversos espaços de cor [9].

Outra alternativa consiste abandonar o modelo gaussiano e estimar uma função de Kernel baseado num conjunto de imagens de treino, actualizadas ao longo do tempo [10]. Este modelo adapta-se melhor a ambientes exteriores, mas não apresenta uma boa performance face a variações bruscas de iluminação.

2.2 Actualização do modelo da Imagem de Fundo

As imagens captadas pelo sistema de vigilância alteram-se ao longo do tempo devido a variações de iluminação, nuvens ou alterações de cena (p.ex. alterações dos carros parados numa rua ou parque de estacionamento). Para diferentes técnicas de modelação de imagens de fundo, existem também diferentes processos de actualização temporal da imagem obtida. Em [7] é apresentado um algoritmo denominado PFinder, no qual cada pixel é actualizado utilizando uma média ponderada entre o valor da imagem de fundo e o valor da imagem adquirida, sendo a variância da distribuição actualizada de forma idêntica. Este método é adequado para variações lentas das condições de iluminação.

Em [1], o autor propõe um método de actualização para uma mistura de gaussianas constituído por dois passos: o primeiro consiste em associar o valor de um dado pixel da imagem adquirida à gaussiana mais provável, sendo que o segundo passo actualiza a os parâmetros dessa gaussiana.

Outra técnica foi utilizada por [11], na qual a imagem de fundo é actualizada utilizando um filtro de mediana sobre um conjunto de imagens armazenadas ao longo do tempo. Estas imagens podem ser adquiridas em instantes de amostragem consecutivos ou de forma mais esparsa no tempo, o que aumenta a estabilidade do modelo calculado.

Utilizando o conceito de filtro de mediana temporal, o algoritmo de actualização proposto em [10] utiliza duas janelas temporais para sintetizar a imagem de fundo. Uma das janelas guarda as últimas imagens adquiridas, detectando variações rápidas. A outra janela guarda imagens espaçadas no tempo, o que permite capturar uma representação mais estável da imagem de fundo.

2.3 Segmentação de Objectos Activos por Subtracção de Fundo

A detecção de regiões activas pode ser realizada recorrendo à imagem de fundo, avaliando desvios entre esta e cada nova frame adquirida, na medida em que cada diferença significativa é classificada como um objecto. Este processo é realizado pixel a pixel, e as regiões activas encontradas são marcadas para posterior processamento. Este processo é frequentemente designado por Subtracção de Fundo (Background Subtraction).

A subtracção de fundo é popular na literatura. Em [7] é calculada a verosimilhança entre um pixel da imagem adquirida e o mesmo pixel da imagem de fundo, definido através de uma distribuição normal. Caso essa probabilidade seja demasiado baixa, esse pixel será classificado como pertencente a uma região activa. Este modelo sofreu uma evolução para melhorar a sua performance em cenas de exterior. No modelo de mistura de gaussianas proposto em [1], um pixel é classificado como activo se seu valor não for suficientemente bem descrito por nenhuma gaussiana presente na mistura. Uma outra possibilidade consiste em incorporar informação espacial da cena para além da informação de cor. Em [10] é utilizada a função não-paramétrica para modelar a imagem de fundo, sendo feita uma correspondência não só entre um pixel da imagem adquirida e o pixel correspondente da imagem de fundo, mas é também contemplada uma vizinhança desse respectivo pixel. Além da informação de cor, em [12] é também usada informação de textura para realizar a segmentação, tornando o processo mais robusto, já que a informação de textura sofre pouca alteração na presença de mudanças de iluminação.

Uma abordagem alternativa consiste em representar as alterações de iluminação de um pixel sob a forma de estados discretos correspondentes a diferentes eventos. A abordagem por espaço de estados permite que certos casos difíceis de modelar através de métodos não supervisionados possam ser aprendidos através de sequências de treino [13] [14].

Uma das limitações dos métodos descritos acima é o facto de precisarem de uma imagem de fundo estática. Esta limitação foi contornada em [15] e [16]. Os métodos apresentados são capazes de lidar com ambientes variáveis no tempo, através da aprendizagem e previsão de padrões de movimento numa sequência de vídeo.

2.4 Modelo de Mistura de Gaussianas

Nesta tese, o sistema desenvolvido é comparado com um modelo clássico de detecção de objectos activos baseado em mistura de gaussianas [1], que é amplamente utilizado em aplicações de video-vigilância. Esta secção faz uma descrição mais detalhada do modelo proposto para comparação.

O modelo de mistura de gaussianas parte do pressuposto de que o valor de um dado pixel da imagem de fundo pode variar devido não só a variações lentas da iluminação (por exemplo, devido à movimentação do sol), mas também devido a variações dos próprios elementos constituintes da imagem de fundo (por exemplo, o movimento das folhas das árvores). Posto isto, assume-se que o valor de um pixel X é uma variável aleatória independente cuja distribuição de probabilidade é modelada através de uma mistura de distribuições gaussianas, ou seja, para um dado pixel X , a probabilidade do seu valor ter sido gerado pela mistura de gaussianas é dado por

$$P(X) = \sum_{i=1}^K \omega_i * \mathcal{N}\{X, \mu_i, \Sigma_i\} \quad (2.1)$$

em que K é o número total de modos da mistura (que assume geralmente valores entre 3 e 5), ω_i é uma estimativa do peso da i -ésima distribuição gaussiana da mistura, μ_i e Σ_i representam a média e a matriz de covariância do i -ésimo modo, que toma a forma

$$\Sigma_k = \sigma_k^2 I \quad (2.2)$$

ou seja, admite-se que todas as componentes de cor são independentes e têm a mesma variância.

Assim, no caso de um dado pixel pertencer à imagem de fundo, este será representado por uma das componentes da mistura e o seu valor será utilizado para a actualização da mesma. Se uma das K misturas corresponder ao valor de um pixel, o valor de ω_k é actualizado de acordo com a seguinte regra:

$$\omega_{k,t} = (1 - \alpha)\omega_{k,t-1} + \alpha(M_{k,t}) \quad (2.3)$$

onde α é uma constante de actualização com valores compreendidos entre 0 e 1, e $M_{k,t}$ é uma máscara binária que toma o valor 1 para a gaussiana correspondente ao valor do pixel, e 0 para as restantes. Após este processo, os novos valores de ω_k são normalizados. Para a distribuição gaussiana correspondente, os valores de μ e σ^2 (definido em (2.2)) são actualizados através de:

$$\mu_t = (1 - \rho)\mu_{t-1} + \rho X_t \quad (2.4)$$

$$\sigma_t^2 = (1 - \rho)\sigma_{t-1}^2 + \rho(X_t - \mu_t)^T(X_t - \mu_t) \quad (2.5)$$

onde

$$\rho = \alpha * \mathcal{N}\{X_t | \mu_k, \sigma_k^2\} \quad (2.6)$$

A escolha da imagem de fundo em cada instante passa por avaliar para cada pixel qual a gaussiana que melhor representa a imagem de fundo. Primeiro, as diversas distribuições são ordenadas de acordo com ω_k/σ_k . Este valor aumenta de acordo com a importância de cada distribuição de probabilidade. Em seguida, as primeiras B gaussianas são escolhidas para o modelo de fundo, onde

$$B = \arg \min_b \left(\sum_{k=1}^b \omega_k > T \right) \quad (2.7)$$

onde T é uma medida da porção mínima de dados que devem ser tidos em conta para o modelo de fundo, ou seja, se T for pequeno, apenas é escolhida uma distribuição, sendo que o número de distribuições incluídas no modelo de fundo aumenta proporcionalmente com T .

Finalmente, A segmentação de regiões activas é feita tendo em conta que um dado pixel da imagem é classificado como foreground se a probabilidade do seu valor ter sido gerado por uma das distribuições escolhidas para o modelo de fundo não for suficientemente elevada.

Capítulo 3

Sistema de Seguimento

O objectivo desta tese é desenvolver um sistema de seguimento de objectos activos para câmeras fixas que seja robusto a variações das condições de iluminação. As variações lentas, como a posição do sol ao longo do dia, são bem compensadas através dos diferentes métodos apresentados no capítulo anterior. Contudo, a principal dificuldade deste tipo de sistemas tem que ver com as variações bruscas das condições de iluminação, que acontecem em cenas interiores, quando é ligada ou desligada uma das fontes de iluminação. Os métodos existentes não são suficientemente robustos quando se torna necessário lidar com este tipo de situações. A solução proposta nesta tese consiste em recolher vários modelos para a imagem de fundo, cada um representando uma condição de iluminação distinta. O sistema deverá ser capaz de identificar em cada instante qual o tipo de iluminação da cena, e seleccionar qual a melhor imagem de fundo a utilizar na fase de segmentação de objectos activos. Além disso, o sistema deverá ser capaz de identificar alterações das condições de iluminação, e no caso de se tratar de um tipo de iluminação desconhecido, iniciar um processo de aprendizagem das novas condições, gerindo de forma automática o conjunto de modelos existentes. Este tipo de abordagem difere dos modelos existentes, já que a imagem de fundo é interpretada de forma global, havendo uma contribuição de todos os pixels para a identificação do tipo de iluminação presente na cena, e também porque existem duas formas de actualizar as imagens existentes, servindo uma para actualizar as imagens de fundo existentes, e a segunda para a fase de aprendizagem de novas imagens de fundo.

A figura 3.1 ilustra o diagrama de blocos do sistema desenvolvido nesta tese. A imagem adquirida em cada instante de tempo, $I(t)$, é analisada pelo bloco de gestão de imagens de fundo, que guarda as múltiplas aparências da cena com diferentes tipos de iluminação. Este bloco identifica a condição de iluminação, $B(t)$, activa nesse instante, e actualiza o respectivo modelo antes da fase de segmentação, onde se obtém uma máscara de foreground, $F(t)$, para esse instante de tempo.

As secções seguintes apresentam detalhadamente o processo identificação da imagem de fundo activa

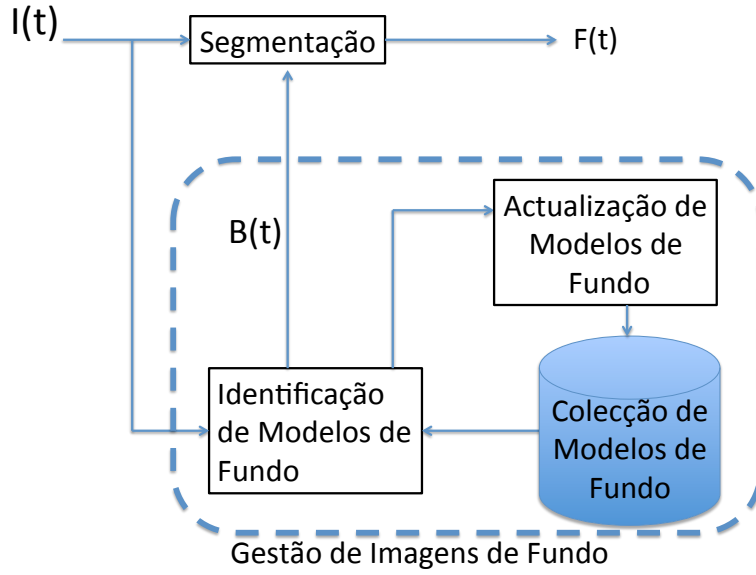


Figura 3.1: Diagrama de Blocos do Sistema a Implementar

em cada instante, bem como a forma como os diferentes modelos são construídos e actualizados. Por fim, é feita uma descrição do processo de segmentação para obtenção das regiões activas na imagem adquirida.

3.1 Colecção de Imagens de Fundo

Os algoritmos existentes guardam unicamente uma imagem de fundo ou as suas propriedades estatísticas. Mesmo utilizando um modelo de mistura de Gaussianas, estas técnicas não são suficientemente robustas quando existem alterações bruscas das condições de iluminação.

Esta tese usa uma abordagem de múltiplos modelos para a imagem de fundo, assumindo que se dispõe de uma colecção de J modelos. Em cada um é guardada uma aparência possível da imagem de fundo, sendo construída usando um modelo probabilístico para cada pixel. Cada pixel do modelo M^j é representado através de uma distribuição gaussiana

$$M^j(x) \sim \mathcal{N}(B^j(x), (\sigma^2)^j(x)) \quad (3.1)$$

em que $B^j(x) \in \mathbb{R}^3$ é o valor médio de $M^j(x)$, designado por imagem de fundo j , e $(\sigma^2)^j(x) \in \mathbb{R}^3$ é a variância de $M^j(x)$, sendo que tanto a média como a variância são estimadas a partir de imagens anteriores.

Em cada instante de tempo, é seleccionado o modelo M^j que melhor se adapta à imagem captada. Os passos para a identificação do modelo activo são descritos na secção seguinte.

3.2 Gestão de Imagens de Fundo

3.2.1 Identificação de Modelos de Fundo

Admite-se que se dispõe de uma colecção de J modelos de imagem de fundo, calculados previamente. A probabilidade do valor da imagem I no ponto x , canal de cor c e instante de tempo t ser gerado pelo modelo M^j é dada por:

$$P_c^j(x) = P\{I_c(t, x) | M^j(x)\} = \mathcal{N}\{I_c(t, x) | B_c^j(x), (\sigma_c^2)^j(x)\} \quad (3.2)$$

A equação (3.2) calcula a probabilidade para cada pixel de cada canal de cor. O passo seguinte é obter uma medida da semelhança entre a imagem adquirida e o modelo j para cada pixel, designada por $d^j(x)$. Este valor é obtido através de uma combinação probabilística de P_c^j , na qual se assume que as probabilidades de cada canal de cor são independentes, ou seja, $d^j(x)$ é calculado através de

$$d^j(x) = P_{c_1}^j(x) P_{c_2}^j(x) P_{c_3}^j(x) \quad (3.3)$$

onde c_1 , c_2 e c_3 denotam os três canais de cor. A combinação probabilística é a forma mais simples de se obter um valor para $d^j(x)$, permitindo também perceber em qual dos canais de cor se manifestam as maiores diferenças entre a o modelo de fundo e a imagem adquirida.

Uma vez calculada a combinação das probabilidades, é obtida uma medida global, D^j , da semelhança entre $I(t)$ e o modelo M^j , definida por

$$D^j = \prod_X d^j(x) \quad (3.4)$$

Posto isto, a escolha do modelo mais adequado passa pela avaliação do valor de D^j de todos os modelos existentes na colecção, sendo que a imagem de fundo activa no instante de tempo t , $B(t)$, é dada por:

$$B(t) = B^j : j = \arg \max_j D^j \quad (3.5)$$

3.2.2 Actualização dos modelos de Imagens de Fundo

O facto de haver múltiplos modelos para representar a imagem de fundo permite que o sistema se adapte a variações bruscas de iluminação. Para lidar com variações lentas, é necessário que os modelos existentes sejam actualizados. Para cada modelo M^j , é necessário actualizar não só o valor de B^j como também a variância $(\sigma^2)^j$ utilizada no cálculo das probabilidades descritas na secção 3.2.1. O sistema actualiza os modelos existentes através de um filtro de mediana temporal, uma vez que a imagem de fundo não pode conter objectos activos.

Considere-se um conjunto constituído por N imagens associadas a um dado modelo, W_N^j , recolhidas

em instantes de tempo separados de um período de amostragem T_a , durante o período activo de cada modelo. A actualização de B^j é realizada através do calculo da mediana pixel a pixel das imagens contidas em W_N^j , ou seja:

$$B_c^j(x) = \text{median}_{n=1, \dots, N}(W_n)_c^j(x) \quad (3.6)$$

É de notar que o valor de T_a é bastante importante para uma actualização correcta do modelo de fundo, na medida em que o intervalo entre amostragens tem que ser suficientemente grande para que a imagem B^j contenha apenas os elementos estáticos das imagens contidas na janela W_N , garantindo que nenhum objecto móvel é identificado como pertencente à imagem de fundo.

No que toca ao valor de $(\sigma^2)^j$, uma vez que as imagens contidas na janela W_N podem conter objectos activos, o seu valor não pode ser actualizado utilizando o estimador de máxima verosimilhança mais comum para o cálculo de variâncias de distribuições de probabilidade.

$$(\sigma^2)_c^j(x) = \frac{1}{N} \sum_{n=0}^N ((W_n)_c^j(x) - \overline{(W_N)_c^j(x)})^2 \quad (3.7)$$

Como tal, utiliza-se um estimador de desvio absoluto de mediana, MAD , obtido para cada pixel através de

$$MAD_c^j(x) = \text{median}_{n=1, \dots, N} |(W_n)_c^j(x) - B_c^j(x)| \quad (3.8)$$

Posto isto, o valor final para a variância de um dado pixel x para um dado espaço de cor c é dado por:

$$(\sigma^2)_c^j(x) = (1.4826[MAD_c^j(x)])^2 \quad (3.9)$$

expressão que é válida para dados gaussianos.

3.2.3 Adição de um Modelo de Imagem de Fundo

A forma de identificação do modelo activo apresentada em 3.2.1 é válida quando o modelo M^j escolhido descreve com suficiente rigor as condições de iluminação presentes. No entanto, pode ocorrer que surja uma condição de iluminação nova. Aquando da selecção do modelo M^j , esta situação é identificada através de

$$\forall_{j \in J} : D^j < (\eta_I)^{XC} \quad (3.10)$$

em que $\eta_I \in [0, 1]$ representa o grau mínimo de semelhança entre o melhor modelo M^j e a imagem adquirida, X e C representam o número total de pixels e o número de canais de cor da imagem, respectivamente. Esta equação indica que nenhum modelo descreve suficientemente bem a imagem adquirida. Neste caso, é necessário criar um modelo de fundo que se adapte às novas condições de

iluminação. Este novo modelo tem de ser obtido da forma mais rápida possível, mas ao mesmo tempo, garantindo que nenhum objecto activo é identificado como pertencente à imagem de fundo.

Tendo estes factores em conta, quando uma condição de iluminação nova é detectada, o sistema entra numa fase de adaptação, criando um modelo temporário M^* , onde será colocada a informação relevante para a construção do novo modelo de fundo, constituído pela imagem de fundo temporária B^* e um conjunto de imagens temporário W_N^* . O conjunto W_N^* é preenchido com imagens recolhidas a partir do momento em que uma nova condição de iluminação é detectada. Uma vez que a fase de adaptação tem de decorrer no menor tempo possível, estas imagens são recolhidas em instantes de tempo separados de um período de amostragem T_b menor que o utilizado na actualização de modelos de fundo conhecidos, ou seja

$$T_b < T_a \quad (3.11)$$

Até W_N^* ser totalmente preenchido para que o modelo M^* possa ser totalmente construído, o valor de B^* do modelo temporário é actualizado utilizando um mecanismo de actualização da nova imagem de fundo através do cálculo de uma média ponderada frame a frame que utiliza a seguinte fórmula:

$$B_c^*(x) = \alpha I_c(t, x) + (1 - \alpha) B_c^*(x) \quad (3.12)$$

em que α , $\alpha \in [0, 1]$, é uma constante de actualização.

Uma vez preenchida a janela temporal, os valores de B^* e σ^{2*} são obtidos através do processo de actualização de modelos de fundo, apresentado na secção 3.2.2. Após este passo, o modelo temporário final M^* é comparado com todos os modelos existentes na colecção, por forma a conferir que se trata de um modelo completamente novo. Este processo é idêntico ao apresentado na secção 3.2.1, com a diferença de que a probabilidade P^j é calculada comparando o modelo temporário M^* com os modelos existentes na colecção, ou seja:

$$P^j = P\{B^*|M^j\} = \mathcal{N}\{B^*|B^j, (\sigma^2)^j\} \quad (3.13)$$

Obtidos os valores de D^j , o novo modelo é adicionado à colecção se

$$\forall j \in J : D^j < (\eta_M)^{XC} \quad (3.14)$$

em que η_M , $\eta_M \in [0, 1]$ representa o grau mínimo de semelhança entre o melhor modelo M^j e o modelo temporário, M^* , estimado.

Uma vez adicionado um novo modelo à colecção, este é tratado como um modelo conhecido, sendo actualizado usando o período de amostragem T_a original.

Assume-se que a colecção guarda um número fixo de modelos, sendo por isso necessário que os modelos antigos, ou menos utilizados, possam ser substituídos caso um novo modelo surja e precise de ser adicionado à colecção de imagens de fundo. Este processo é realizado através de uma análise de ocorrências dos modelos ao longo do tempo, sendo contabilizado o número de ocorrências de um dado modelo ao longo do tempo. No caso de ser necessário substituir um dos modelos, será substituído aquele cujo número de ocorrências for menor.

3.3 Segmentação de Regiões Activas

Esta secção aborda a forma como os objectos activos são segmentados. Esta tarefa é realizada através da subtracção do fundo activo $B(t)$, calculado de acordo com (3.5), à imagem $I(t)$ adquirida em cada instante de tempo. Por forma a melhorar a qualidade da segmentação, este processo é realizado separadamente para todos os canais de cor quer da imagem adquirida, quer da imagem de fundo activa. Assim, considere-se uma variável S que resulta do cálculo da diferença absoluta entre a imagem $I(t)$ adquirida e o fundo activo $B(t)$, isto é, para cada pixel da imagem, e para canal de cor c :

$$S_c(x) = |B_c(t, x) - I_c(t, x)| \quad (3.15)$$

Uma vez calculada a diferença entre as duas imagens para cada canal de cor, é necessário condensar a informação dos três canais de cor por forma a obter a segmentação final. A informação é condensada através de uma combinação lógica dos canais de cor. Primeiramente, é calculada uma binarização da matriz S , obtida da seguinte forma:

$$F_c(x) = \begin{cases} 1, & \text{se } S_c(x) \geq \delta_c \\ 0, & \text{se } S_c(x) < \delta_c \end{cases} \quad (3.16)$$

onde δ_c é denominado threshold de segmentação para o canal de cor c . A máscara de segmentação final para um dado instante de tempo, $F(t)$, é então calculada com recurso à informação obtida em (3.16), assumindo que um pixel x é classificado como pertencente a um objecto (pixel activo) se esse pixel for classificado como activo em pelo menos um canal de cor, ou seja:

$$F(x) = F_{c_1}(x) \vee F_{c_2}(x) \vee F_{c_3}(x) \quad (3.17)$$

Esta forma de calcular a matriz de segmentação permite ter um maior controlo sobre a informação resultante da subtracção das duas imagens, na medida em que o threshold de segmentação é independente entre os três canais de cor.

Capítulo 4

Conjunto de dados

Este capítulo descreve as sequências de vídeo adquiridas no âmbito desta tese, para teste e avaliação do sistema proposto no Capítulo 3.

4.1 Descrição Geral do Conjunto de Dados

Os dados a analisar são constituídos por sequências de vídeo, captadas num palco de espectáculos, gentilmente disponibilizado pela Culturgest, cujo cenário é constituído pela base do palco e uma tela branca. No espaço são projectados diferentes tipos de iluminação, que variam instantaneamente ao longo do tempo, havendo no total quatro condições de iluminação diferentes, de luz de cor âmbar, azul, vermelha e verde. O conteúdo dessas sequências no que diz respeito à iluminação e presença de pessoas no palco é caracterizado nas figuras 4.1 e 4.2. É de referir que as pessoas que colaboraram neste estudo se apresentaram vestidas de forma habitual, não lhes tendo sido recomendado nenhum tipo específico de vestuário ou de cor.

Foram adquiridas, no total 5 sequências de vídeo utilizando a câmara FaceTime HD em posição fixa e com controlo automático de iluminação e focagem, cuja descrição gráfica está ilustrada na figura 4.3. A primeira sequência de vídeo contém no total 6060 frames e tem como objectivo avaliar a forma como as diferentes condições de iluminação são identificadas. Na figura 4.3a é possível observar a evolução das condições de iluminação ao longo do tempo. A segunda sequência, com um total de 6109 frames, estuda a identificação e segmentação de um objecto, cuja presença ocorre em condições de iluminação constantes no tempo. A figura 4.3b mostra não só a evolução da iluminação, como também o conjunto de frames em que existe presença de objectos activos para serem identificados. A terceira sequência contém 2762 frames e tem uma estrutura semelhante à sequência anterior (representada na figura 4.3c), com a diferença de que são apresentados dois objectos em cena, existindo um cruzamento entre eles. As últimas duas sequências, representadas nas figuras 4.3d e 4.3e, estudam a forma como o sistema mantém a identificação dos objectos activos no caso em que as condições de iluminação se

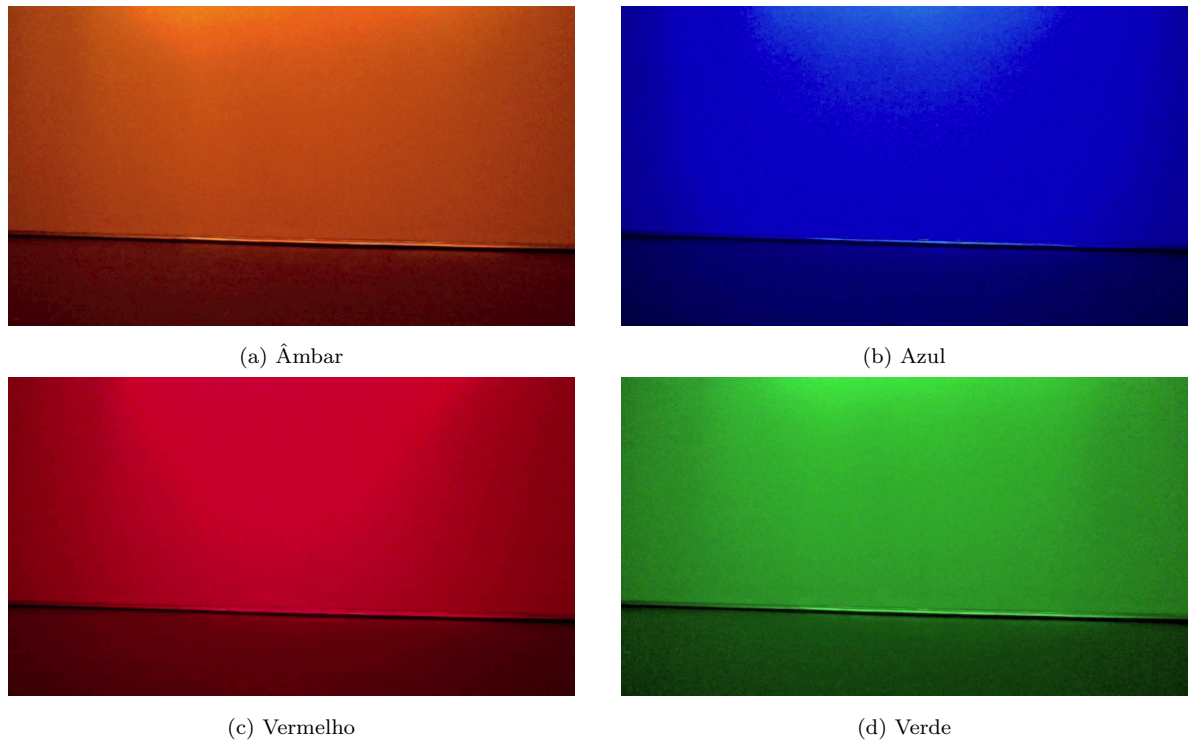


Figura 4.1: Condições de iluminação usadas nos testes

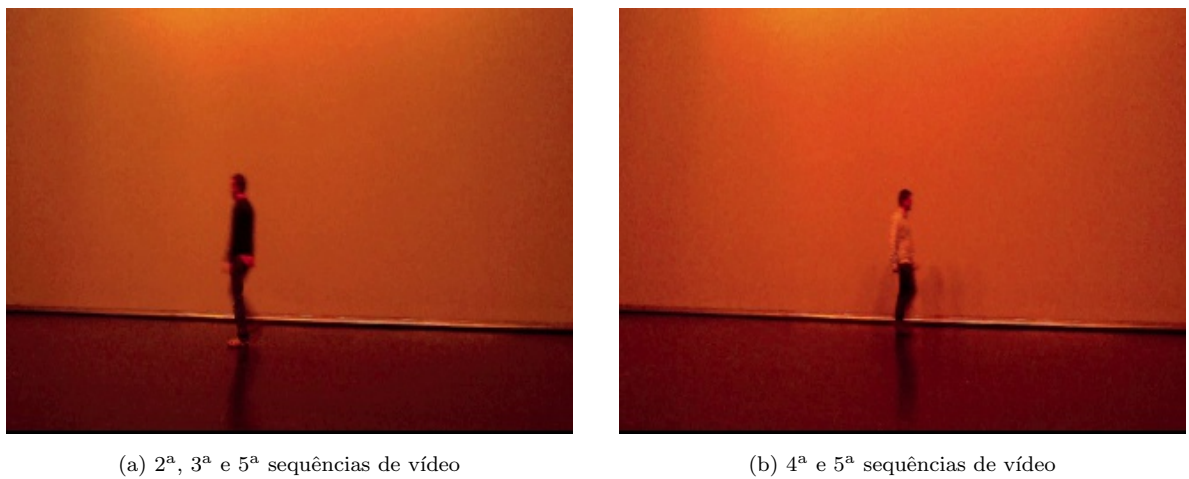


Figura 4.2: Aparência das pessoas usadas nos testes

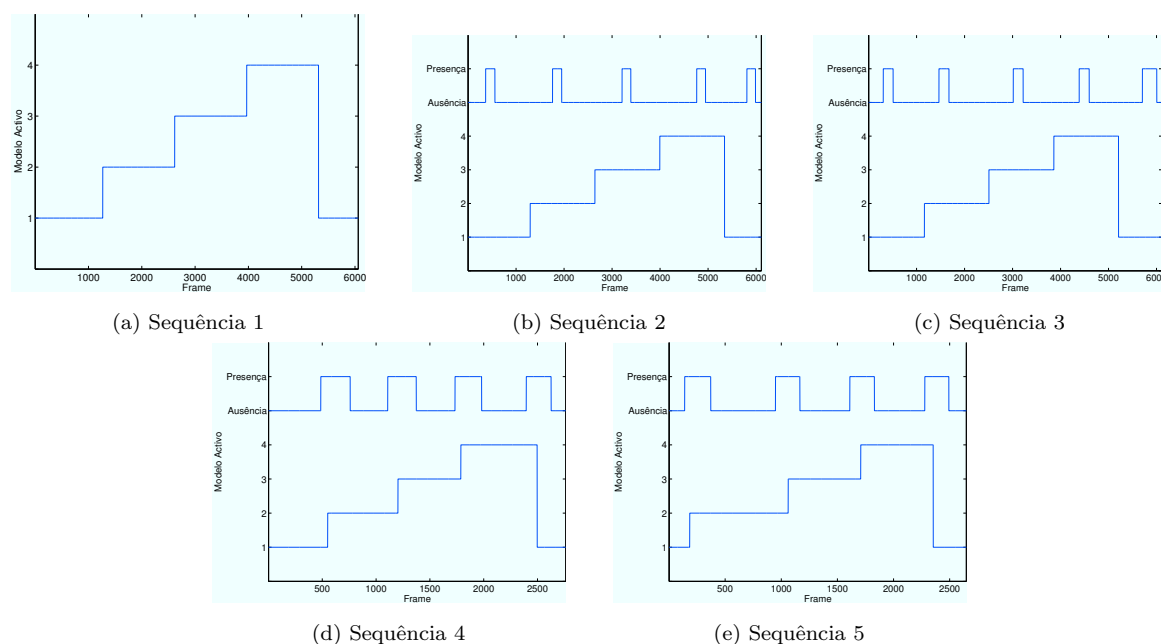


Figura 4.3: Descrição gráfica do conjunto de acontecimentos das sequências de vídeo adquiridas. Os valores entre 1 e 4 representam a condição de iluminação activa em cada instante, e os valores entre 5 e 6 indicam a ausência/presença de pessoas em cena.

alteram durante a presença de elementos activos em cena, pelo que na quarta sequência existe apenas um objecto activo em cena, e na quinta existem dois. As últimas duas sequências têm 6112 e 2648 frames, respectivamente.

Em seguida apresentam-se uma caracterização das sequências de vídeo no que toca à estabilidade de cor ao longo do tempo e do espaço, à forma como o ruído é distribuído, e a forma como a passagem de um objecto afecta o valor de um pixel.

4.2 Estabilidade da Cor no Tempo e no Espaço

O estudo da estabilidade da cor permite caracterizar qual o espaço de cor que poderá oferecer maior robustez tanto na selecção do fundo activo, como na segmentação de regiões activas, sendo testados os espaços de cor RGB, HSV e $L^*a^*b^*$. Para cada tipo de iluminação, é recolhido um conjunto de 10 imagens da primeira sequência de vídeo, com um intervalo de amostragem de 1 segundo. Posto isto, é representada sob a forma de um gráfico a informação relativa a 3 linhas específicas das imagens adquiridas, em cada espaço de cor em análise. Escolheram-se as linhas 50, 270 e 480, correspondentes ao topo da tela, meio da imagem e base do palco, respectivamente. A figura 4.4 ilustra uma percepção visual das linhas escolhidas. Para cada espaço de cor, os valores para o primeiro, segundo e terceiro canais de cor são apresentados a vermelho, verde e azul, respectivamente.

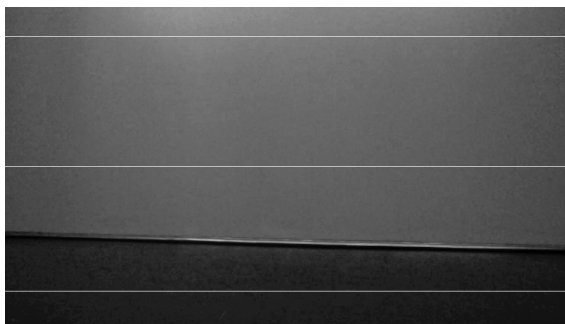


Figura 4.4: Linhas escolhidas para o estudo da estabilidade da cor, representadas a branco.

Os resultados obtidos estão ilustrados nas figuras 4.5 a 4.8, nas quais uma maior dispersão vertical da informação de cada canal de cor indica uma menor estabilidade da cor, e conseqüentemente mais ruído. Desde já se pode observar que o espaço de cor que apresenta maior estabilidade é o $L^*a^*b^*$, dado que os seus valores apresentam pouca dispersão na maioria dos casos.

De facto, no espaço de cor $L^*a^*b^*$, as duas últimas componentes são normalizadas por forma a mapear uniformemente a informação de cor ao longo de dois eixos. Cada um dos eixos representa uma propriedade facilmente identificável, como a relação vermelho/verde no caso do canal a^* , ou a relação azul/amarelo no caso do canal b^* . Estas características tornam este espaço de cor mais robusto no que toca a descrição da tonalidade de uma dada cor.

É de notar que, na figura 4.5f, o valor de H possui uma linha de valores próximo de 0 e outra de valores próximos de 255. Tal facto deve-se à forma como o espaço de cor HSV é definido, sendo que os valores de H representam um ângulo no cone HSV. O valor de H é definido a menos de um múltiplo de 360° e os valores próximos de 0° ou 360° representam a mesma tonalidade. Uma vez que existe ruído, os valores de H podem oscilar entre valores próximos de 0° ou 360° , mesmo que tal facto não represente uma diferença significativa a nível de cor.

Por inspecção visual dos gráficos obtidos é possível afirmar que a condição de iluminação que produz maior ruído na imagem é a verde, ao passo que a que possui menor ruído é a vermelha. Uma situação a reparar tem que ver com a forma como o ruído se distribui na linha 50 da condição de iluminação azul.

O facto do nível de saturação da cor ser baixo faz com que haja um aumento da incerteza do valor da cor nessa zona. Ainda assim, o espaço de cor $L^*a^*b^*$ consegue mais uma vez representar a informação em questão nessa zona, oferecendo uma maior estabilidade.

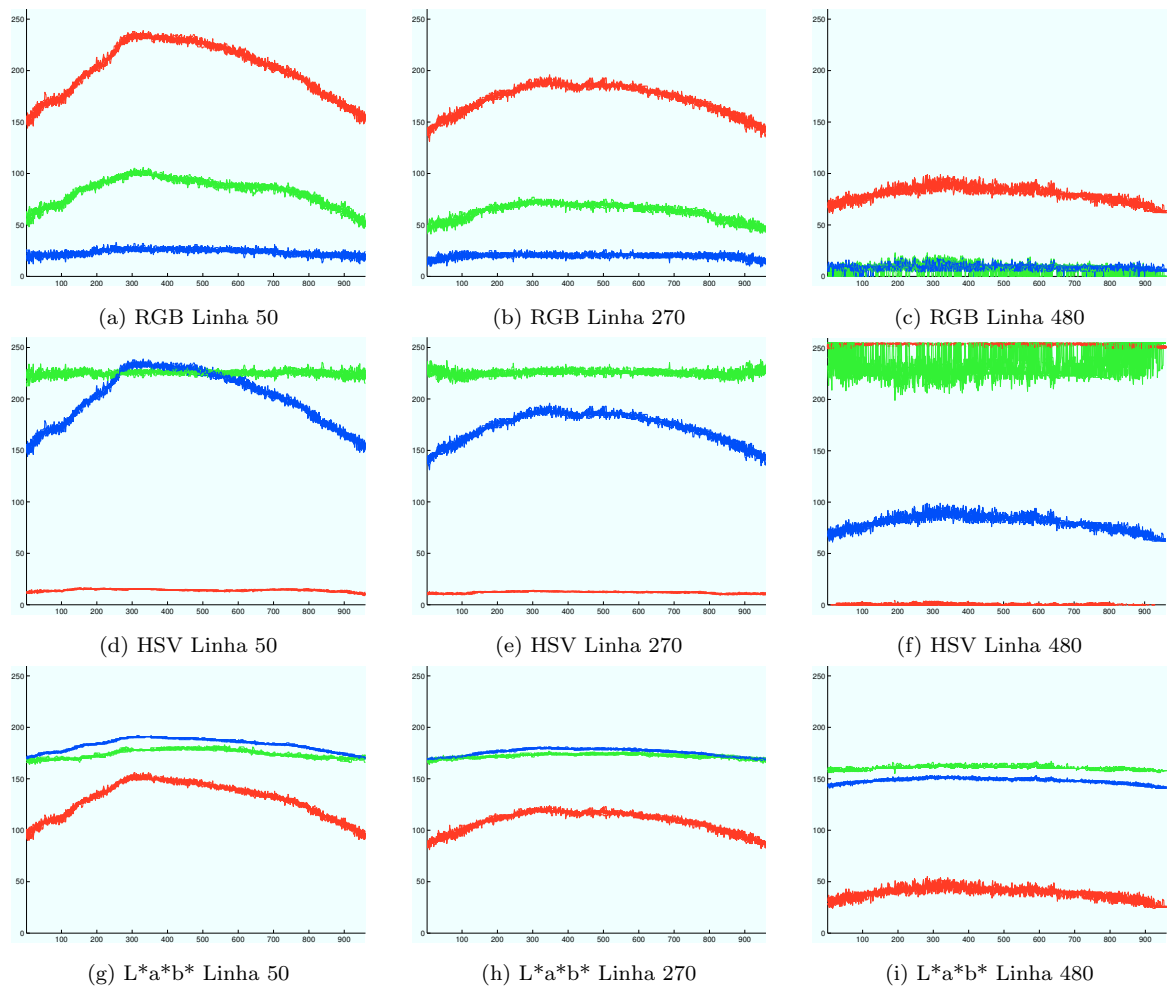


Figura 4.5: Estabilidade da iluminação âmbar para os espaços de cor RGB (1^a linha), HSV (2^a linha) e L*a*b* (3^a linha). Os valores a vermelho, verde e azul representam o primeiro, segundo e terceiro canais de cor, respectivamente.

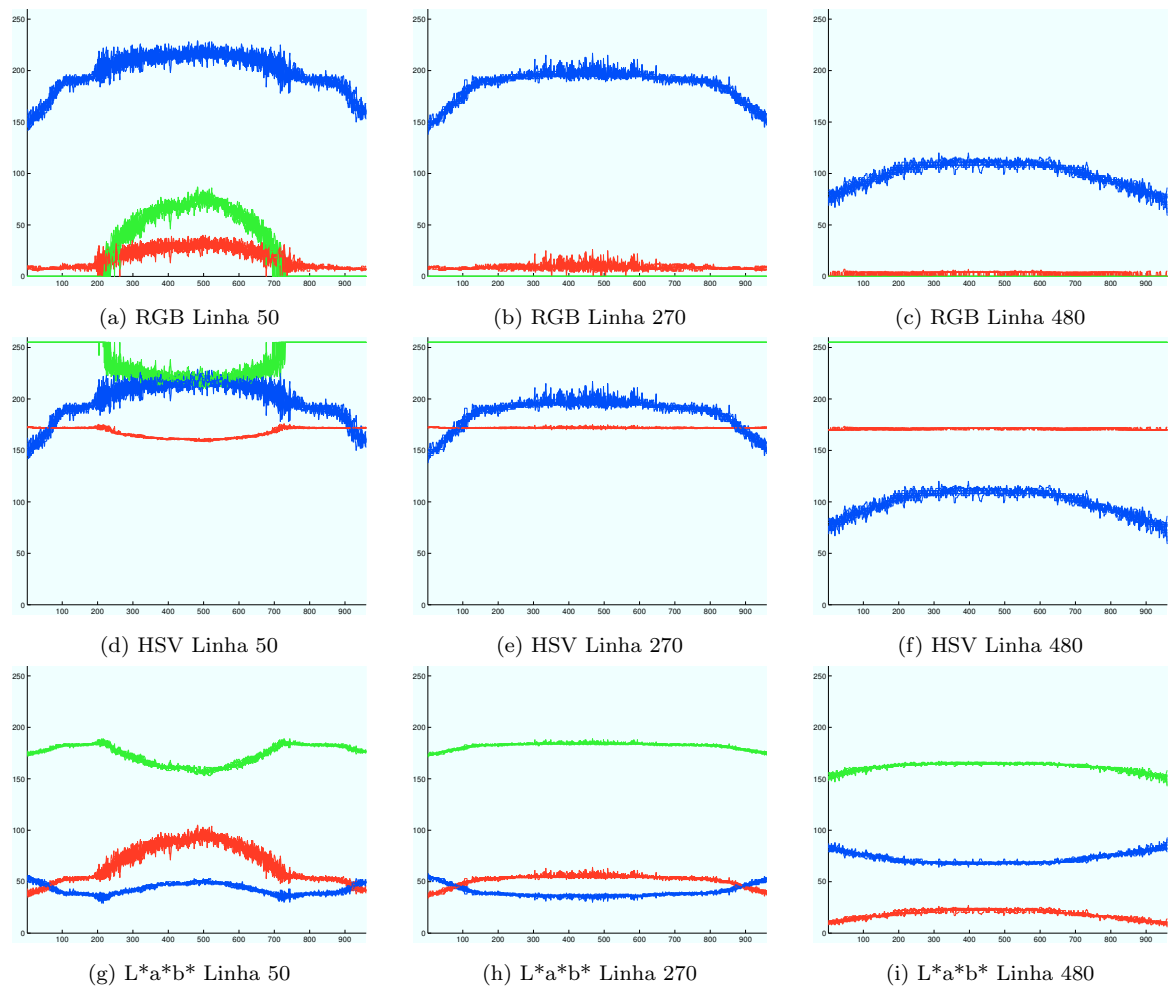


Figura 4.6: Estabilidade da iluminação azul para os espaços de cor RGB (1^a linha), HSV (2^a linha) e L*a*b* (3^a linha). Os valores a vermelho, verde e azul representam o primeiro, segundo e terceiro canais de cor, respectivamente.

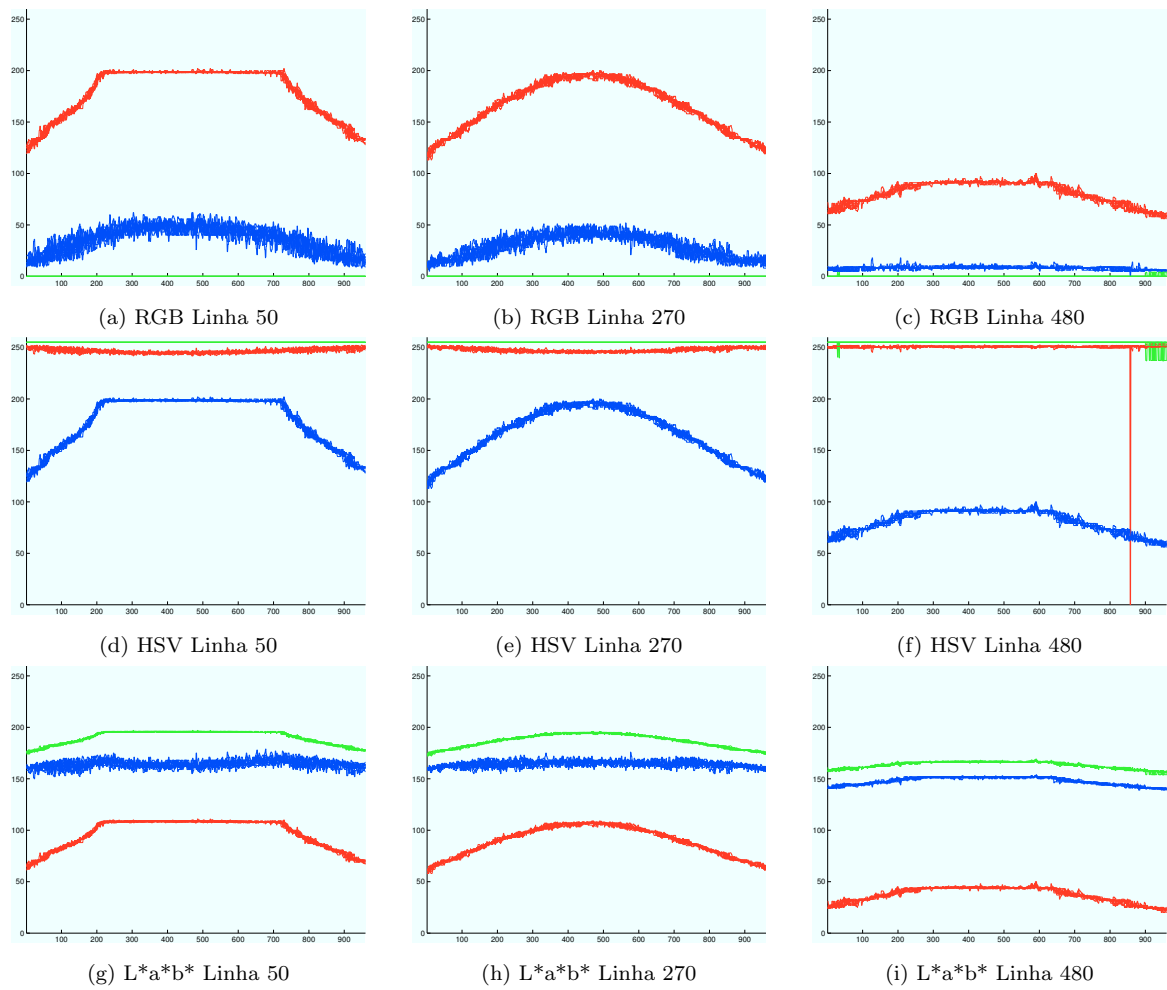


Figura 4.7: Estabilidade da iluminação vermelha para os espaços de cor RGB (1^a linha), HSV (2^a linha) e L*a*b* (3^a linha). Os valores a vermelho, verde e azul representam o primeiro, segundo e terceiro canais de cor, respectivamente.

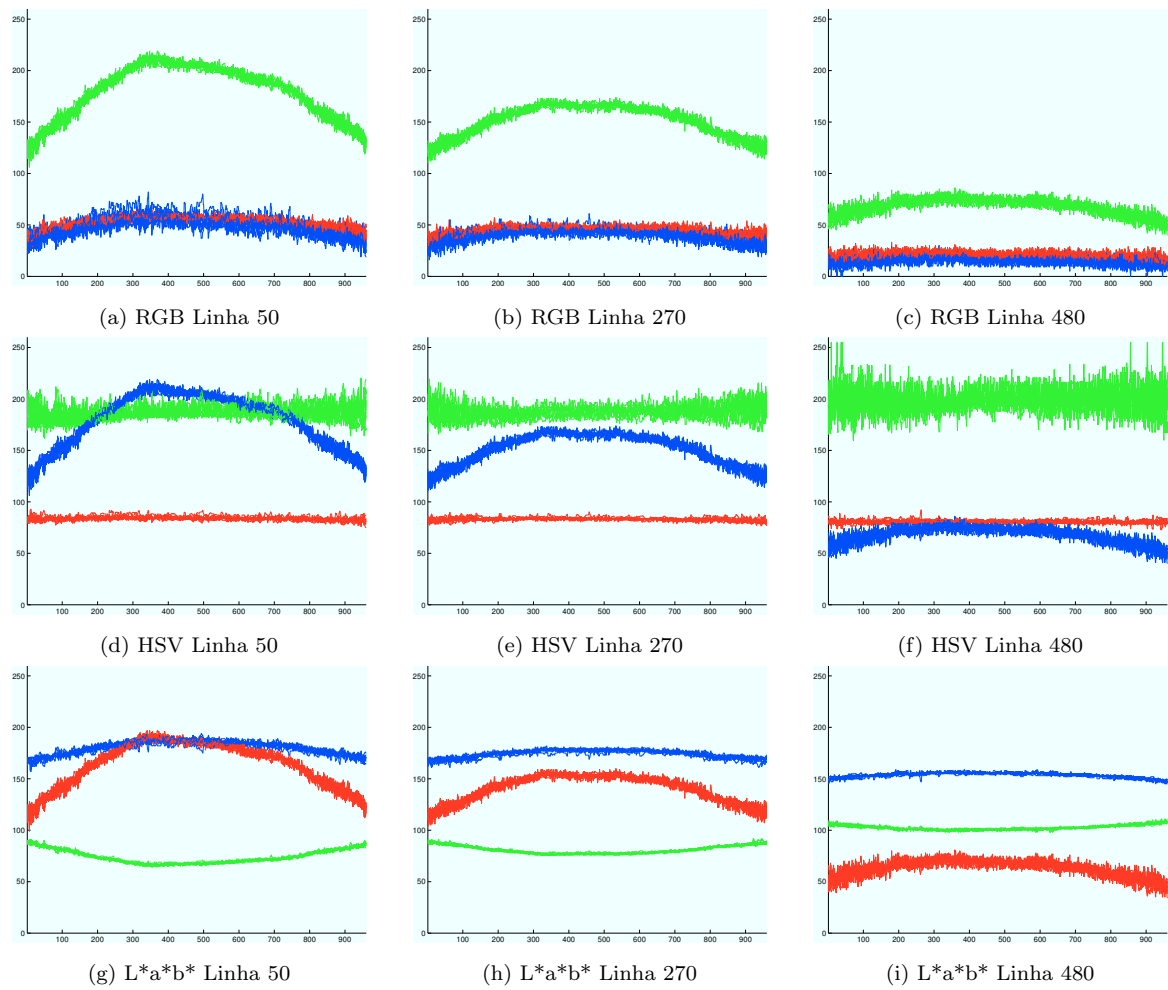


Figura 4.8: Estabilidade da iluminação verde para os espaços de cor RGB (1^a linha), HSV (2^a linha) e L*a*b* (3^a linha). Os valores a vermelho, verde e azul representam o primeiro, segundo e terceiro canais de cor, respectivamente.

4.3 Variância da imagem de fundo em Cada Intervalo Estacionário

O objectivo desta secção é verificar a forma como o ruído varia ao longo do espaço, e se esta variação está relacionada com certas zonas ou espaços de cor.

A incerteza de ruído é obtida calculando o desvio padrão de cada pixel nas imagens adquiridas, tendo-se utilizado o conjunto de 10 imagens por tipo de iluminação adquiridas nas condições descritas na secção anterior. Para cada pixel, é calculada a média das imagens de teste, ou seja:

$$\overline{I_c(x)} = \frac{1}{N} \sum_{n=0}^N I_c(n, x) \quad (4.1)$$

em que $I_c(n, x)$ designa o valor da c -ésima componente de cor da imagem I , no n -ésimo instante de aquisição e na posição x . A estimativa da variância para cada pixel é então obtido através de

$$(\sigma_C)^2(x) = \frac{1}{N} \sum_{n=0}^N (I_c(n, x) - \overline{I_c(x)})^2 \quad (4.2)$$

sendo o desvio padrão dado por

$$\sigma_c(x) = \sqrt{(\sigma_C)^2(x)}. \quad (4.3)$$

Este processo é realizado para todos os espaços de cor em estudo. Uma vez que os valores de $\sigma^c(x)$ podem ser bastante diferentes entre as diferentes zonas e os diferentes espaços de cor, é feita uma normalização da gama de variação de todos os coeficientes de cor em relação ao maior valor de σ^c entre todos os valores calculados. Este passo permite obter uma boa medida de comparação entre os diferentes espaços de cor a comparar, cujos resultados estão ilustrados nas figuras 4.9 a 4.12.

Analisando as imagens obtidas, pode-se concluir que o ruído não se distribui uniformemente ao longo do espaço, e que este depende não só do espaço de cor utilizado, mas também das condições de iluminação presentes. Tomando como exemplo a condição de iluminação azul, é possível notar que existe menos ruído nas zonas onde as cores são mais saturadas. Além disso, as zonas mais escuras da imagem são também menos bem caracterizadas, ou seja, quando a intensidade de um pixel é demasiado baixa, torna-se mais difícil obter uma representação estável da informação relativa à tonalidade da cor, facto que se nota especialmente no espaço de cor HSV.

Uma vez que por inspecção visual das figuras 4.9 a 4.12 não é possível obter uma medida quantitativa do ruído, é calculada uma média dos valores obtidos para cada espaço de cor e cada condição de iluminação, de acordo com

$$\bar{\sigma}_c = \frac{1}{X} \sum_{x=1}^X \sigma_c(x). \quad (4.4)$$

Os valores obtidos são normalizados, isto é, o valor de cada canal de cor é dividido pelo seu valor máximo admissível (255 da maioria dos casos) de forma a que a gama de variação seja $[0, 1]$. Os resultados podem ser observados na tabela 4.1.

Espaço de Cor	RGB	HSV	L*a*b*
Ambar	0.093	0.138	0.047
Azul	0.094	0.071	0.054
Vermelho	0.076	0.044	0.051
Verde	0.136	0.167	0.074

Tabela 4.1: Média do Valor do Ruído nos Diferentes Espaços de Cor.

Os resultados obtidos confirmam que o espaço de cor L*a*b* é o que obtém melhores resultados tanto a nível de ruído, como a nível da estabilidade da cor. Posto isto, é de esperar que na detecção de objectos, este espaço de cor seja também mais robusto que os outros em estudo. A partir da informação presente na tabela, é possível definir um valor para o limiar de segmentação apresentado na secção 3.3. Uma vez que a maioria dos valores obtidos para o desvio padrão são inferiores a 0.1, é possível afirmar que este valor pode ser utilizado para a segmentação das regiões activas encontradas. Uma vez que a gama de valores a nível digital se situa entre 0 e 255, o valor para o limiar de segmentação escolhido é de 25.

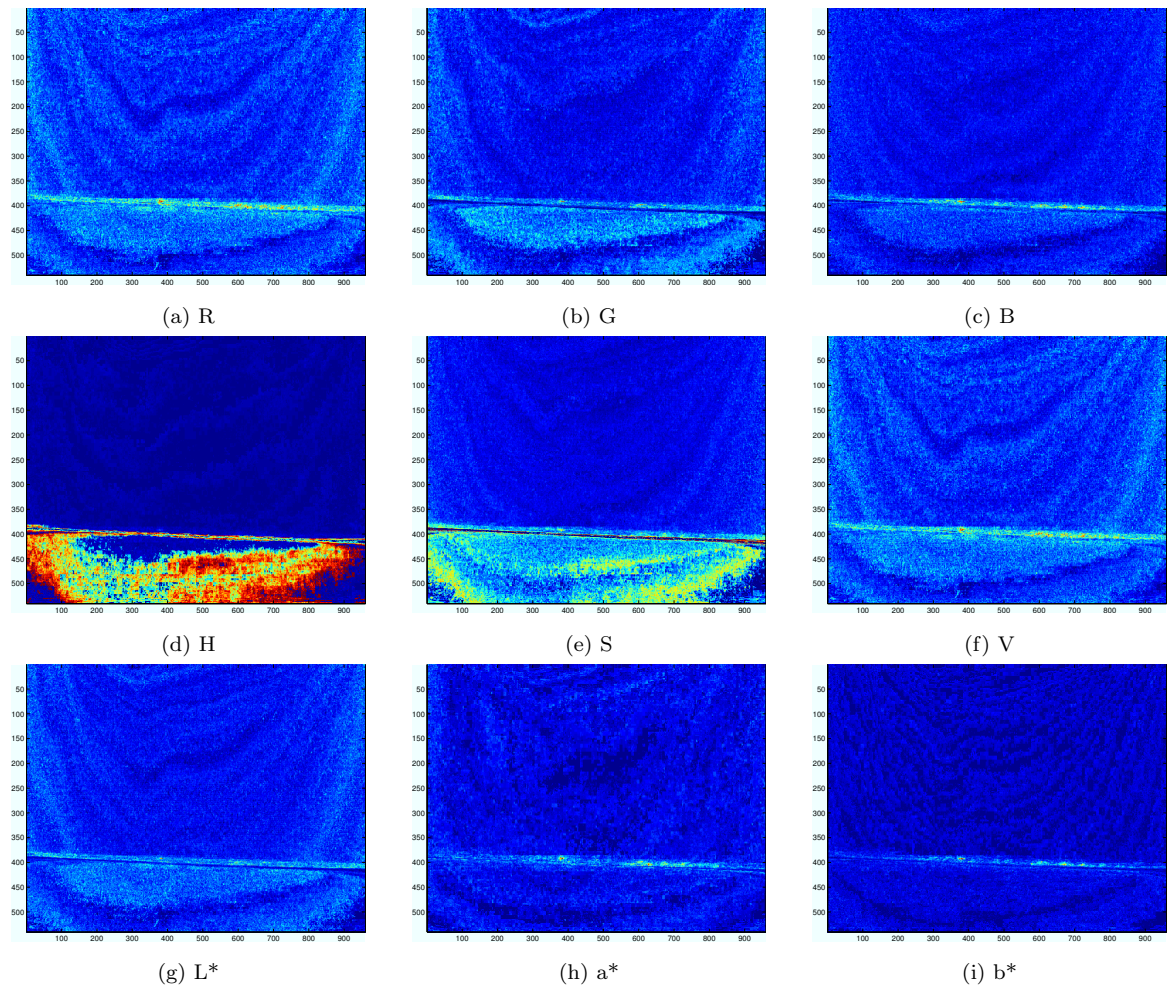


Figura 4.9: Desvio padrão da imagem de fundo de iluminação âmbar. Escala de cor: azul escuro - valores pequenos, vermelho - valores elevados

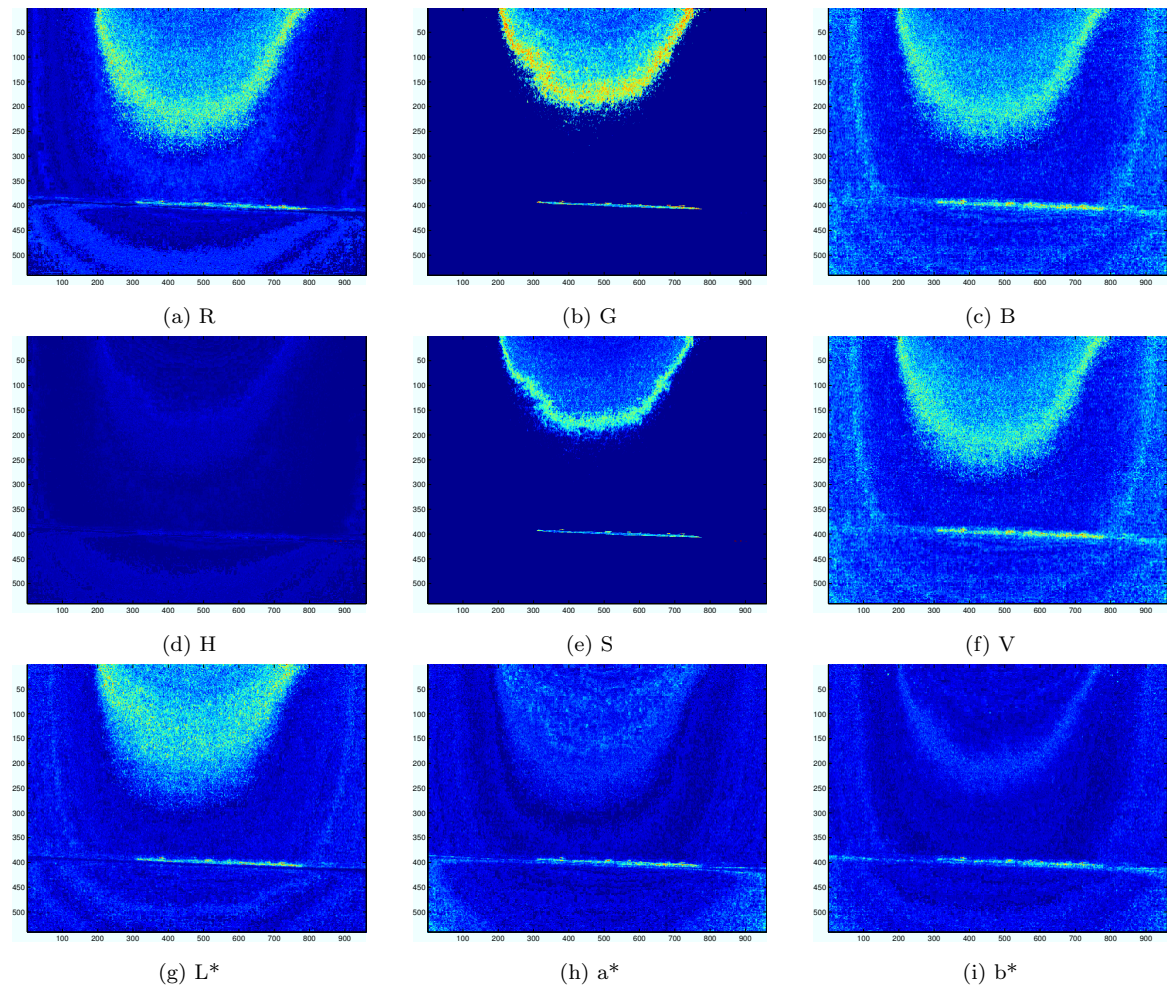


Figura 4.10: Desvio padrão da imagem de fundo de iluminação azul. Escala de cor: azul escuro - valores pequenos, vermelho - valores elevados

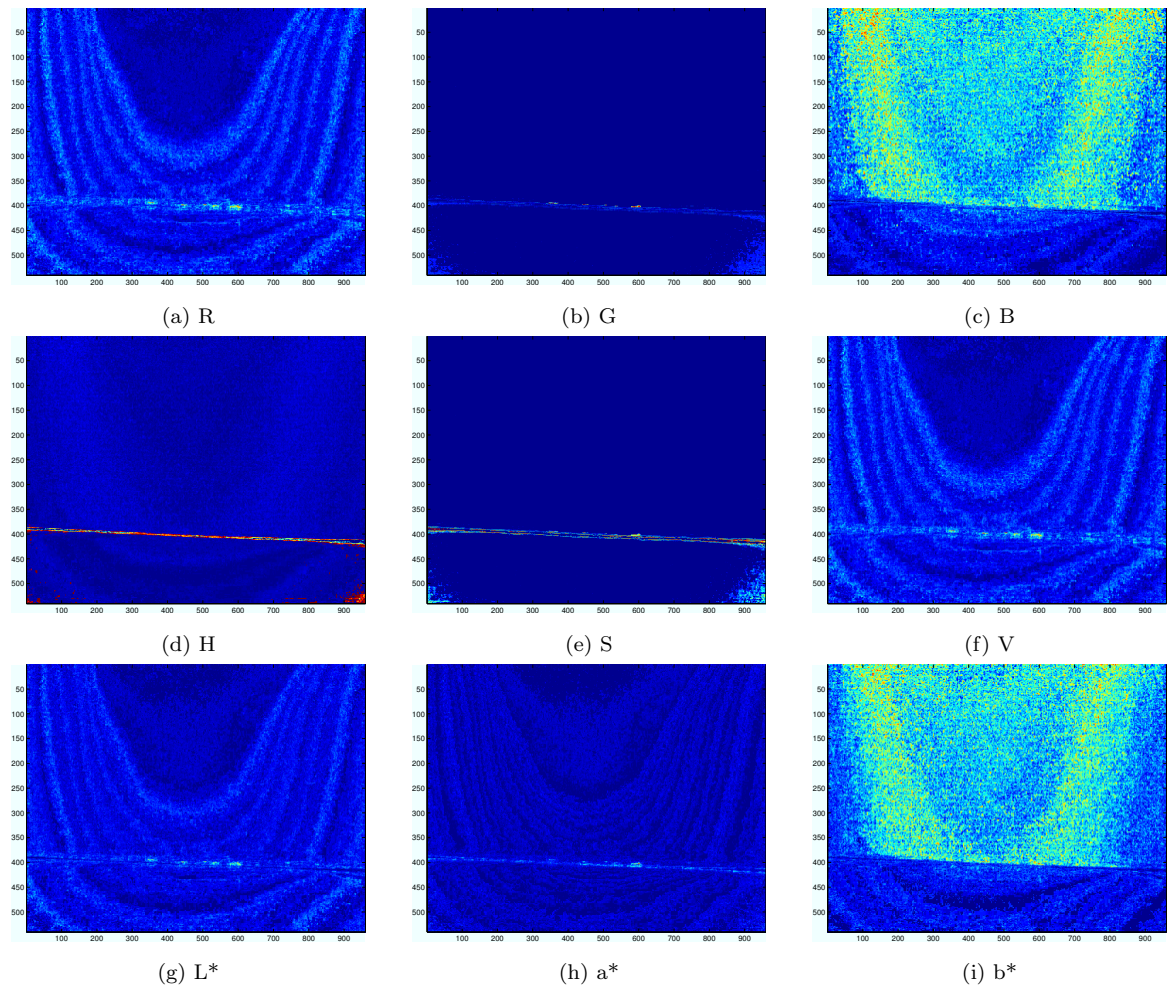


Figura 4.11: Desvio padrão da imagem de fundo de iluminação vermelha. Escala de cor: azul escuro - valores pequenos, vermelho - valores elevados

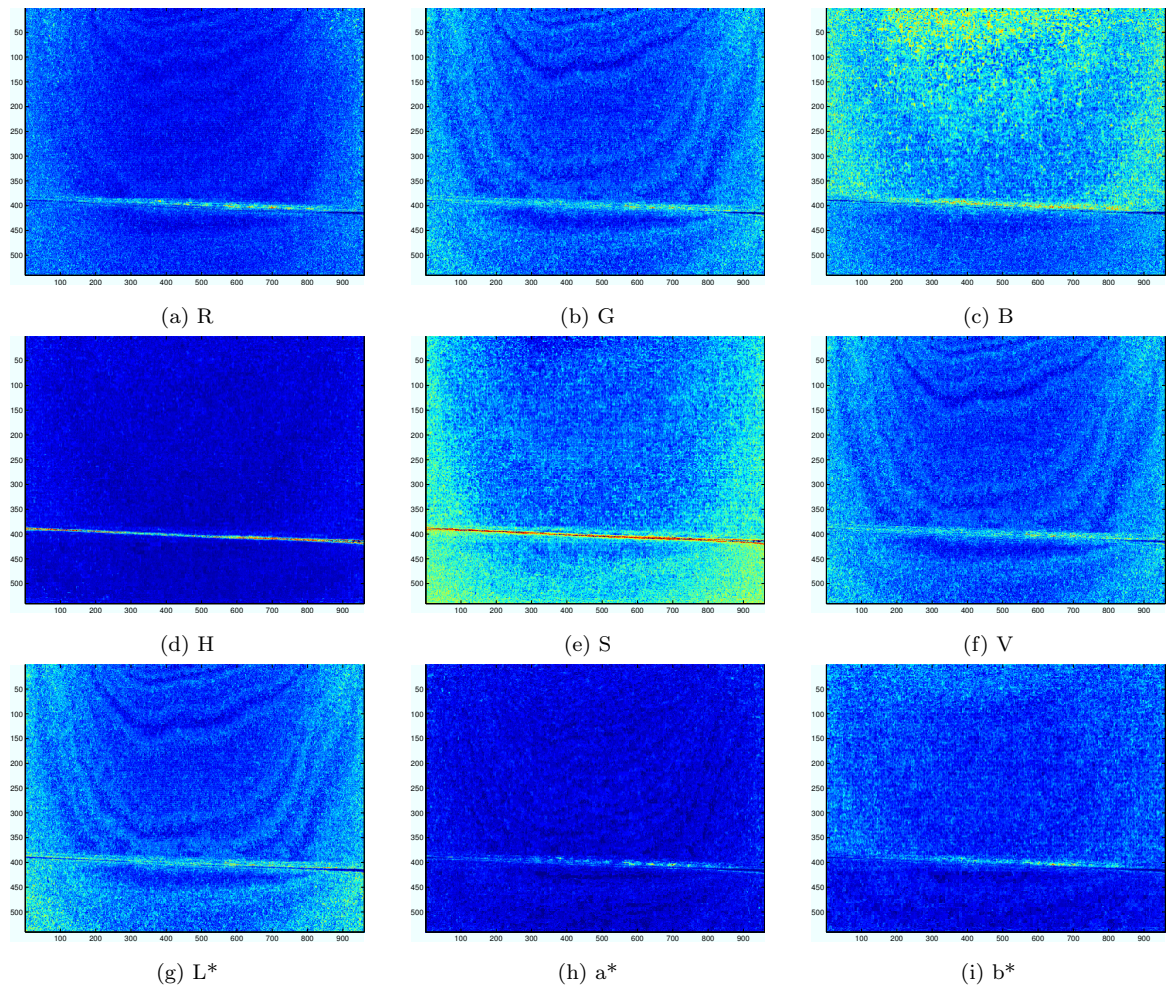


Figura 4.12: Desvio padrão da imagem de fundo de iluminação verde. Escala de cor: azul escuro - valores pequenos, vermelho - valores elevados

4.4 Reacção à Presença de Objectos

Esta secção tem como objectivo estudar a forma como a presença de um objecto (pessoa em palco) altera os valores de um dado pixel em relação aos elementos estáticos de cada condição de iluminação, sendo esta a alteração que terá de ser detectada automaticamente pelo sistema de vigilância. Para este passo, são analisadas quatro sequências vídeo nas quais existe a passagem de um objecto. As sequências são subsequências da segunda sequência de vídeo, sendo que cada uma representa uma condição de iluminação diferente, tendo todas as subsequências uma duração de 175 frames.

É então escolhido um pixel comum a todas as sequências de vídeo e feita uma representação temporal do seu valor para cada tipo de iluminação e para cada espaço de cor. Neste teste, foi escolhido o pixel de coordenadas (300,480), por se localizar sensivelmente a meio da imagem e num local onde existe presença de um objecto (pessoa).

Os resultados obtidos estão ilustrados nas figuras 4.13 a 4.16. Por inspecção dos gráficos obtidos, verifica-se que existem diferenças consideráveis no valor dos pixels, pelo que a detecção de objectos pode ser feita utilizando qualquer um dos espaços de cor em estudo. Ainda assim, é possível observar mais uma vez que o espaço de cor $L^*a^*b^*$ parece ser o mais adequado para o efeito, uma vez que os valores para quando existe passagem de um objecto são mais estáveis que nos restantes espaços de cor, mas também porque não existe saturação dos valores para o caso em que é detectado um objecto.

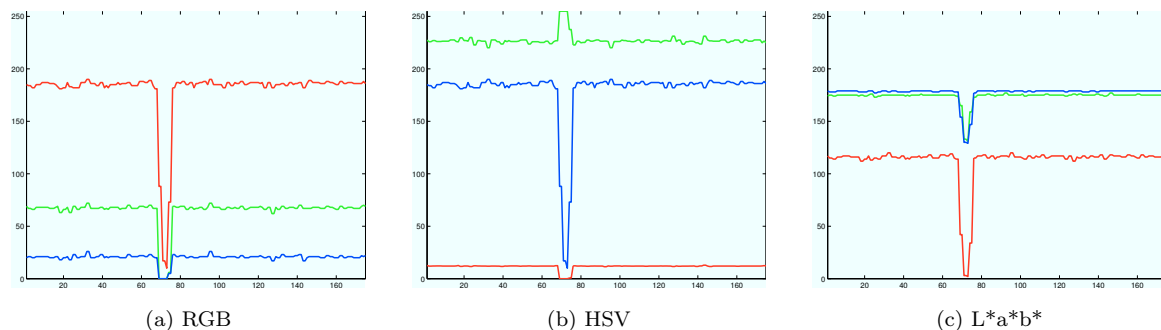


Figura 4.13: Reacção à passagem de objectos na iluminação âmbar. Os valores a vermelho, verde e azul representam o primeiro, segundo e terceiro canais de cor, respectivamente.

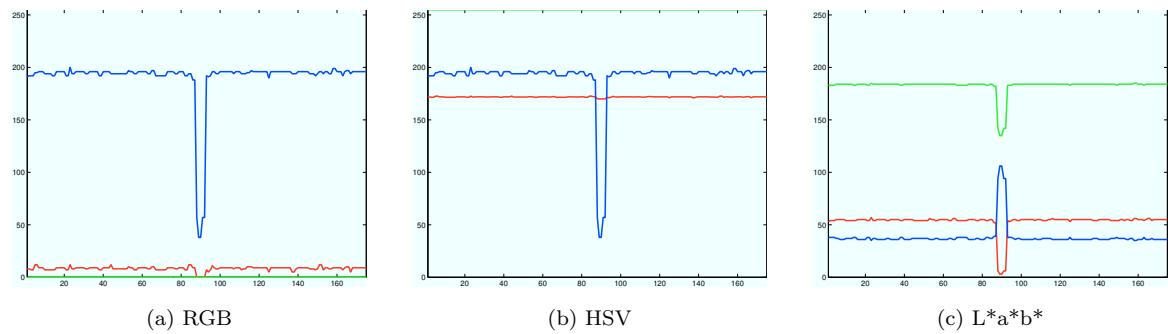


Figura 4.14: Reacção à passagem de objectos na iluminação azul. Os valores a vermelho, verde e azul representam o primeiro, segundo e terceiro canais de cor, respectivamente.

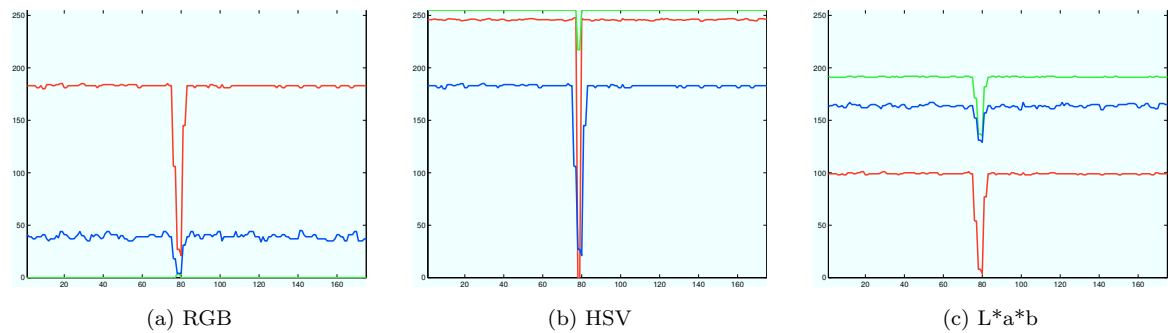


Figura 4.15: Reacção à passagem de objectos na iluminação vermelha. Os valores a vermelho, verde e azul representam o primeiro, segundo e terceiro canais de cor, respectivamente.

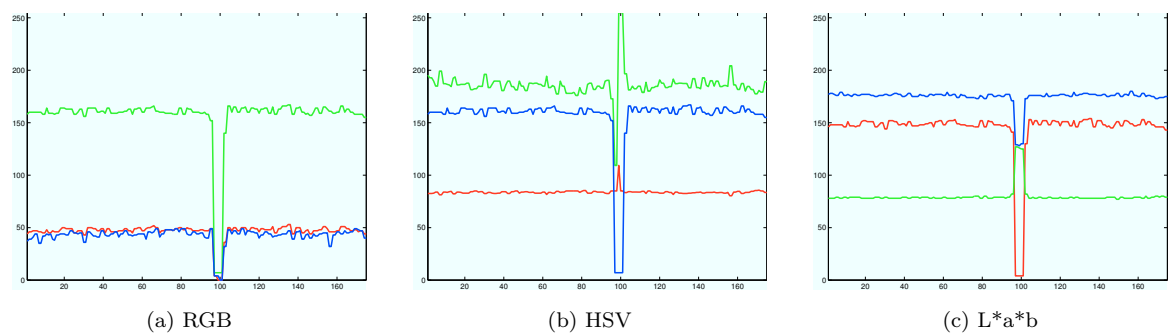


Figura 4.16: Reacção à passagem de objectos na iluminação verde. Os valores a vermelho, verde e azul representam o primeiro, segundo e terceiro canais de cor, respectivamente.

Capítulo 5

Resultados Experimentais

Este capítulo tem como objectivo avaliar o sistema de detecção de objectos para condições de iluminação não estacionárias, proposto nesta tese. Os testes realizados avaliam a forma como o sistema identifica as diferentes imagens de fundo ao longo do tempo, bem como a forma como este se comporta face à presença de objectos activos. É também avaliada a forma como os parâmetros de gestão da colecção de fundos afecta o desempenho da identificação do modelo activo em cada instante. Por fim, o sistema proposto é comparado com um modelo clássico baseado em Mistura de Gaussianas [1], por forma a avaliar as melhorias do sistema desenvolvido nesta tese face a um modelo bem conhecido.

5.1 Metodologia de Avaliação

Nesta secção é feita uma descrição pormenorizada do conjunto de testes efectuados ao sistema proposto. O primeiro teste consiste em registar a evolução temporal da condição de iluminação identificada pelo sistema, representando-a de forma gráfica, por forma a comparar com o valor esperado.

De seguida é feita uma avaliação da detecção de objectos activos para o caso de iluminação constante e para o caso em que existem alterações de iluminação. Para ambos os testes, é recolhido um conjunto de imagens correspondentes ao resultado da segmentação efectuada pelo sistema em diferentes instantes de tempo, para todas as condições de iluminação e para os diversos espaços de cor em estudo. O último passo é fazer uma comparação do resultado de segmentação do sistema com um modelo clássico, tendo-se escolhido o modelo de Mistura de Gaussianas, apresentado em [1]. Para cada caso, a saída do sistema é comparada com uma segmentação de ground truth, que se ilustra nas figuras 5.1 e 5.2. O desempenho do algoritmo de segmentação é caracterizado através da matriz de confusão que contém as probabilidades de verdadeiros negativos, falsos negativos, falsos positivos e verdadeiros positivos.

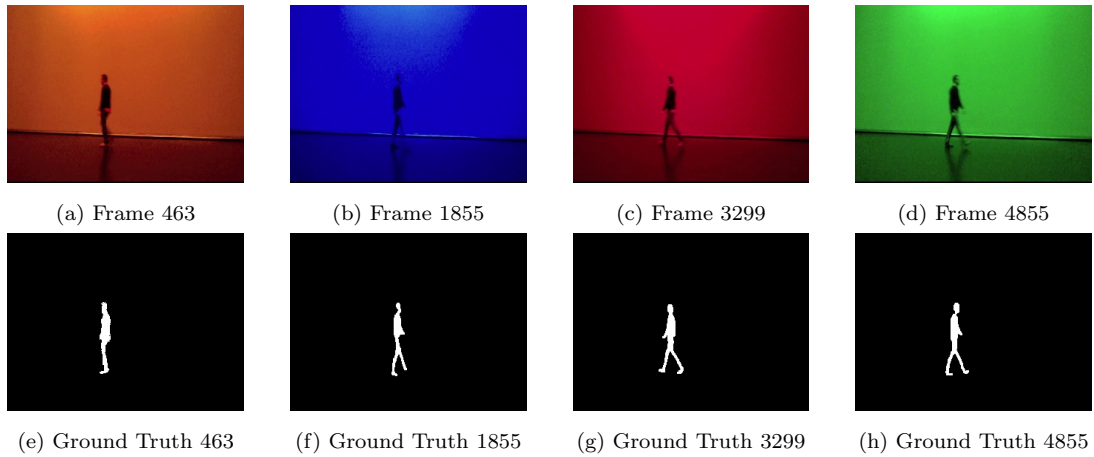


Figura 5.1: Ground truth para os exemplos usados no teste de iluminação constante

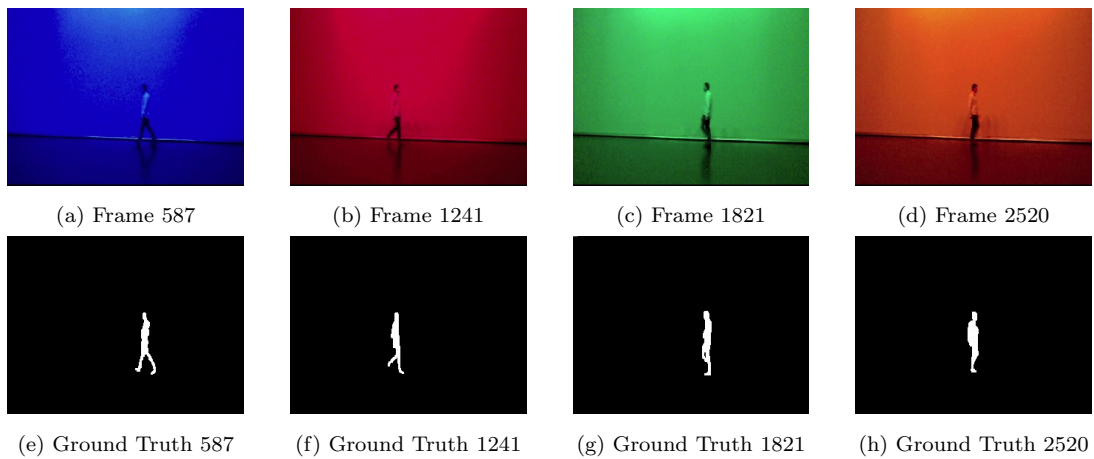


Figura 5.2: Ground truth para os exemplos usados no teste de iluminação variável

É importante referir que, nestes testes, o sistema não tem conhecimento prévio de nenhuma imagem de fundo, e admite-se que o número máximo de modelos de fundo que o sistema pode gerir é 4.

5.2 Identificação de Imagens de Fundo

Esta secção permite avaliar a forma como o sistema se comporta na identificação da imagem de fundo activa em cada instante, bem como na identificação e adição de novos modelos. Neste teste foi utilizada a primeira sequência de vídeo, visto não haver presença de elementos activos. Esta sequência é analisada utilizando todos os espaços de cor em estudo.

Existem quatro parâmetros de interesse no algoritmo de gestão de modelos de fundo, apresentados no capítulo 3:

- T_a - Intervalo entre recolha de imagens para o processo de actualização de um modelo conhecido.
- T_b - Intervalo entre recolha de imagens para o processo de construção de um novo modelo.
- η_I - Taxa de semelhança entre um imagem adquirida e um modelo da colecção.
- η_M - Taxa de semelhança entre dois modelos da colecção.

É importante que para um modelo de imagem de fundo existente na colecção, o intervalo temporal entre a adição de novas imagens à janela temporal seja grande o suficiente por forma a não englobar objectos móveis no modelo, garantido que é obtida a imagem de fundo mais estável possível. Da mesma forma, o intervalo entre a adição de novas frames para a construção de um novo modelo tem de contemplar as características descritas acima. Contudo, o valor escolhido tem de ser mais baixo que no caso anterior, por forma a que a construção do novo modelo de fundo seja a mais rápida possível, garantindo também a estabilidade do modelo construído. Como tal, escolheu-se um intervalo de amostragem de 15 frames para o caso de modelos pertencentes à colecção de fundos e um intervalo de amostragem de 8 frames no caso da construção de um novo modelo de fundo.

O terceiro parâmetro, η_I , definido na secção 3.2.3, regula a semelhança entre a imagem adquirida e as imagens de fundo existentes na colecção, na medida em que uma determinada imagem de fundo é identificada como activa quando o seu grau de semelhança com a imagem adquirida está acima de um certo valor. O último parâmetro, η_M , definido na secção 3.2.3, regula a gestão das imagens de fundo existentes na colecção, na medida em que uma imagem de fundo é considerada nova se não houver nenhuma imagem de fundo na colecção cujo grau de semelhança acima de determinado valor.

Por forma a verificar a influência dos dois últimos parâmetros descritos, são feitos dois testes, sendo o primeiro realizado utilizando os seguintes valores:

- η_I - 70%
- η_M - 90%

Os resultados obtidos estão ilustrados na figura 5.3, nos quais os valores entre 1 e 4 descritos nos gráficos indicam o modelo activo em cada instante, e o valor 0 indica que o sistema detecta uma alteração nas condições de iluminação, entrando na fase de aprendizagem de um novo modelo. Desde já se verifica que quando existe uma alteração da condição de iluminação de azul para vermelho e de vermelho para verde, a identificação destas alterações não é reconhecida nos espaços de cor HSV e $L^*a^*b^*$. Este facto significa que o valor escolhido para a semelhança entre a imagem adquirida e as imagens de fundo da colecção é baixo, tornando o sistema pouco sensível a variações significativas. Além disso, na transição da condição de iluminação verde para âmbar no espaço de cor RGB, o sistema interpreta a iluminação âmbar como nova, o que significa que o valor escolhido para a semelhança entre duas imagens de fundo

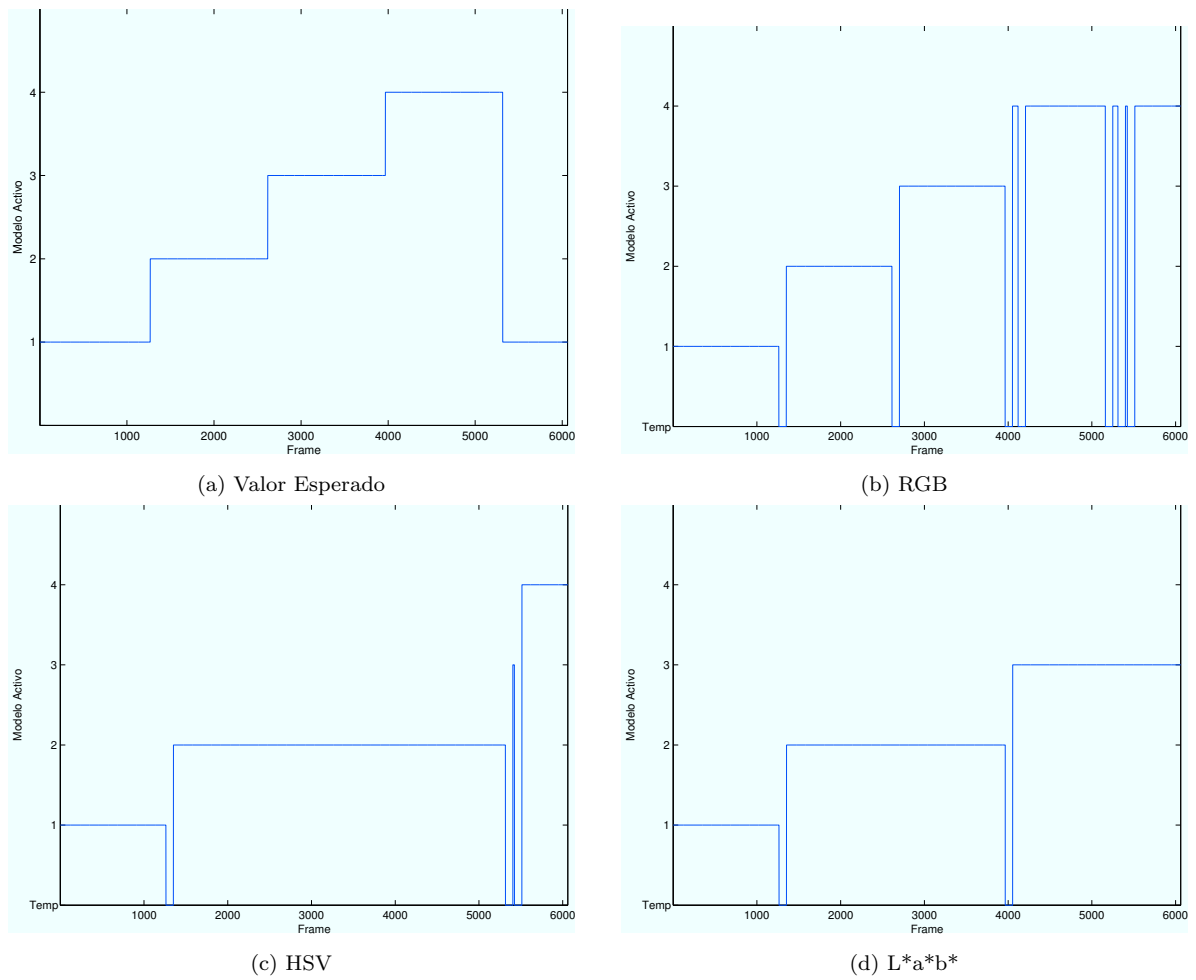


Figura 5.3: Imagem de fundo identificada ao longo do tempo no primeiro teste. Os valores entre 1 e 4 representam o modelo de fundo seleccionado pelo algoritmo. O valor 0 representa a fase de aprendizagem de um novo modelo de fundo

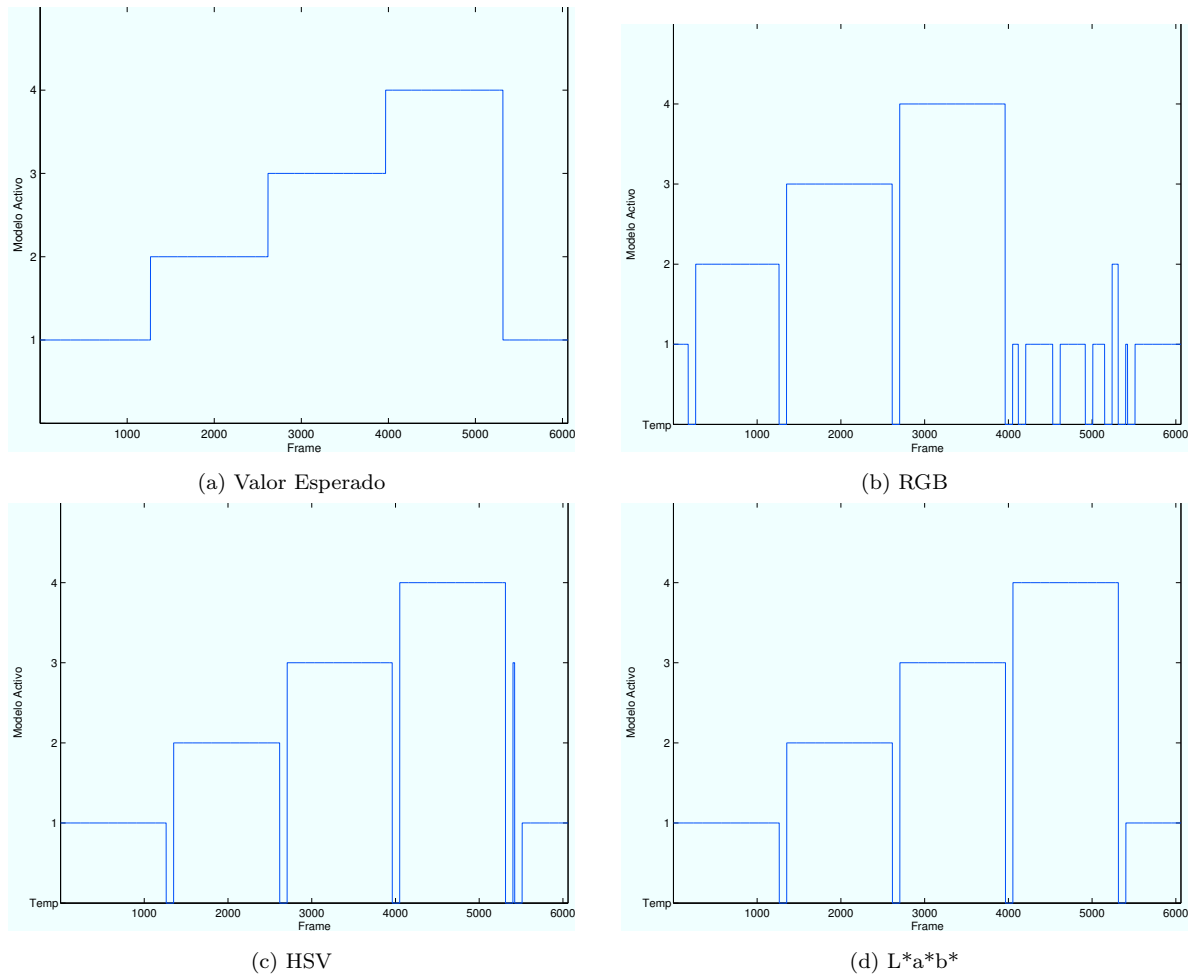


Figura 5.4: Imagem de fundo indentificada ao longo do tempo no segundo teste. Os valores entre 1 e 4 representam o modelo de fundo seleccionado pelo algoritmo. O valor 0 representa a fase de aprendizagem de um novo modelo de fundo

terá de ser reduzido.

Por forma a tentar melhorar os resultados obtidos no primeiro caso, é realizado um segundo teste que utiliza os seguintes valores:

- η_I - 80 %
- η_M - 80 %

A figura 5.4 apresenta os resultados para os valores testados no segundo teste. Conclui-se que o sistema detecta correctamente o modelo de fundo quando se usa o espaço de cor $L^*a^*b^*$ para os parâmetros utilizados. No que toca ao espaço de cor HSV, a identificação é feita também de forma correcta, excepto na identificação da ultima condição de iluminação (âmbar), que é posteriormente corrigida,

sendo identificada a condição de iluminação prevista.

É possível observar que o espaço de cor RGB apresenta bastante instabilidade, muito devido à presença de ruído, surgindo o primeiro problema de identificação entre as frames 1 e 1000, período no qual é identificada uma segunda imagem de fundo, apesar das condições de iluminação serem constantes, como se pode constatar na figura 5.4a. Este problema está relacionado não só com o parâmetro de identificação da imagem de fundo activa, mas também com o parâmetro de comparação das imagens de fundo existentes, cujo valor terá de ser menor para este espaço de cor.

Face aos resultados obtidos nos dois testes, é possível concluir que o segundo conjunto de valores produz melhores resultados que no primeiro caso, sendo que o espaço de cor $L^*a^*b^*$ identifica de forma correcta todas as variações de iluminação.

5.3 Detecção de Objectos Activos

Esta secção permite avaliar a forma como os objectos activos são segmentados. São efectuados dois testes, um deles para o caso em que não existe variação das condições de iluminação durante a passagem de elementos activos, tendo assim uma imagem de fundo constante e estabilizada. O segundo teste avalia o comportamento do sistema na situação em que existem variações das condições de iluminação durante a presença de objectos activos em cena.

5.3.1 Segmentação Com Iluminação Constante

Para avaliar o desempenho do sistema foram escolhidas quatro frames da segunda sequência de vídeo e extraídos os objectos activos recorrendo ao algoritmo descrito na secção 3.3 para todos os espaços de cor em estudo. As frames a analisar têm de ser escolhidas por forma a que as imagens de fundo estejam completamente estimadas, além de que o objecto a identificar deve ser completamente observável, de preferência a meio da frame. Posto isto, foram escolhidas as frames 463, 1855, 3299 e 4855.

Uma variável importante neste ponto é o limiar de segmentação δ_c a escolher (ver (3.17)). O limiar deve ser suficiente elevado para não classificar pixels do fundo, afectados por ruído aditivo, como pixels do objecto activo, bem como não pode ser demasiado elevado, podendo prejudicar a boa identificação dos elementos activos. Face à análise efectuada no capítulo 4, escolheu-se um valor para o limiar de segmentação de 25, sendo este valor usado para todos os espaços de cor, e para todos os canais de cor.

Os resultados da segmentação estão ilustrados na figura 5.5. Tal como previsto no capítulo anterior, o espaço de cor $L^*a^*b^*$ é aquele que obtém o melhor resultado, principalmente porque apresenta menos

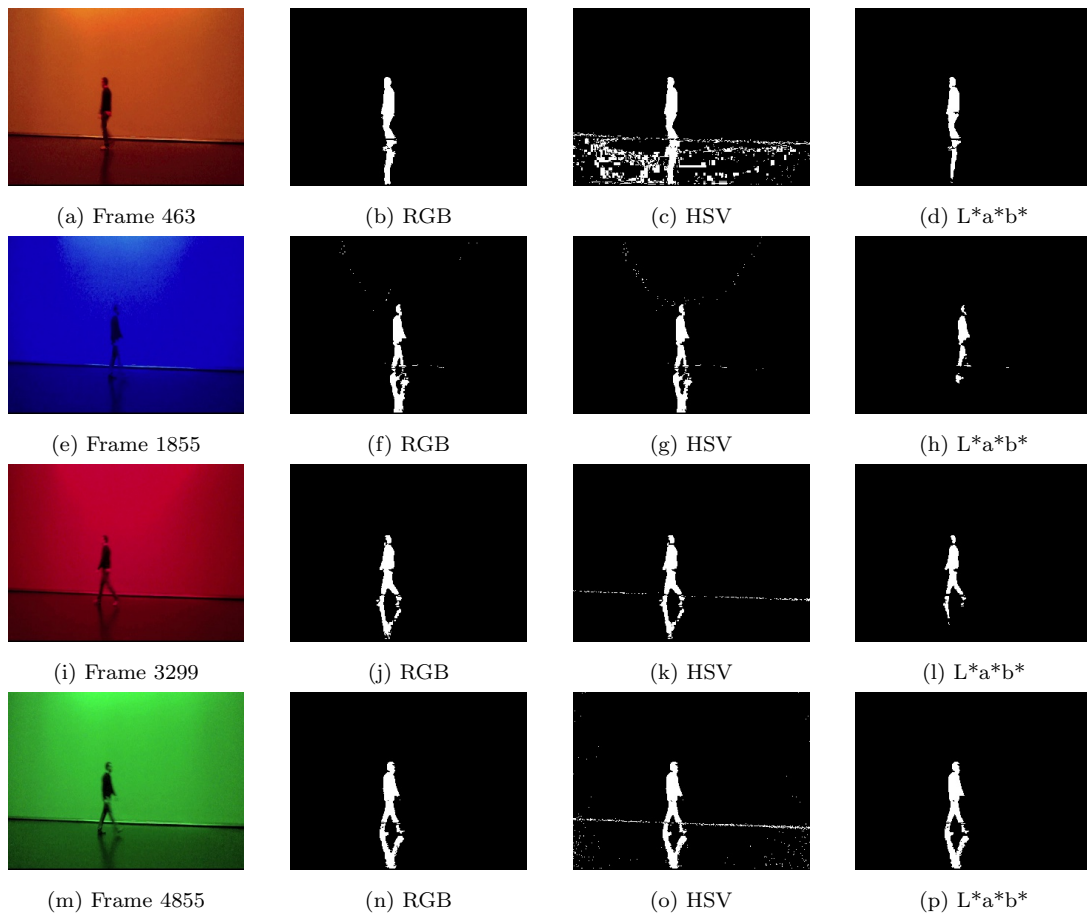


Figura 5.5: Resultados da segmentação utilizando o sistema desenvolvido em condições de iluminação constante

ruído que os outros espaços de cor em estudo, mas também porque este é mais estável a nível da interpretação da cor. O espaço de cor RGB produz igualmente bons resultados, com excepção da condição de iluminação azul, situação na qual algum ruído é reconhecido como sendo um objecto activo. No entanto, esta situação pode ser resolvida através de operações morfológicas à imagem binária obtida, removendo áreas pequenas que tenham sido mal classificadas.

Quanto ao espaço HSV, verifica-se uma maior presença de ruído, confirmando as previsões do capítulo anterior, sendo que este problema pode ser parcialmente resolvido aumentando o valor do limiar de segmentação para este espaço de cor. Contudo, o ruído presente na figura 5.7c está relacionado com a forma como este espaço de cor é definido.

É de notar que o valor escolhido para o limiar de segmentação permite que as sombras provocadas pelos objectos móveis sejam também considerados como elementos activos, podendo este valor ser alterado consoante a aplicação dada ao sistema implementado.

5.3.2 Segmentação com Variações de Iluminação

Para o segundo teste de segmentação foi utilizada a quarta sequência de vídeo. Na escolha das frames a recolher é necessário ter em conta que estas têm de pertencer à fase de aprendizagem de novas imagens de fundo, sendo recolhidas frames aproximadamente um segundo após haver uma alteração da iluminação. Assim sendo, foram escolhidas as frames 587, 1241, 1821 e 2520. Tal como no caso de iluminação constante, o valor escolhido para o limiar de segmentação é também de 25.

Uma variável importante para este teste é a taxa de actualização utilizada para a imagem de fundo temporária, definida em 3.2.3, uma vez que é necessário que as variações de iluminação sejam rapidamente incluídas na imagem de fundo, sugerindo a um valor de α elevado, mas que ao mesmo tempo seja capaz de excluir desta imagem temporária possíveis objectos activos. Como tal, escolheu-se um valor de α de 0.15, ou seja, a imagem adquirida num instante de tempo tem um peso de 15% no cálculo da imagem de fundo para o instante seguinte.

A figura 5.6 mostra os resultados da segmentação obtidos neste caso. Uma vez que estão presentes objectos móveis aquando da alteração de iluminação, é natural que o resultado da segmentação não apresente resultados tão bons como no caso de iluminação constante. Ainda assim, a maioria dos pixels identificados pertencem à zona onde o objecto se localiza, ainda que este não seja totalmente segmentado, devido ao facto de o objecto a identificar se encontrar em movimento. Outro facto positivo a ter em conta é que, na maioria dos casos, as regiões onde não existe presença de objectos não contêm pixels classificados como activos, tirando os casos da frame 2520 para os espaços de cor RGB e HSV, e as frames 1241 e 1821 para o espaço de cor HSV. Face aos resultados obtidos, pode considerar-se o

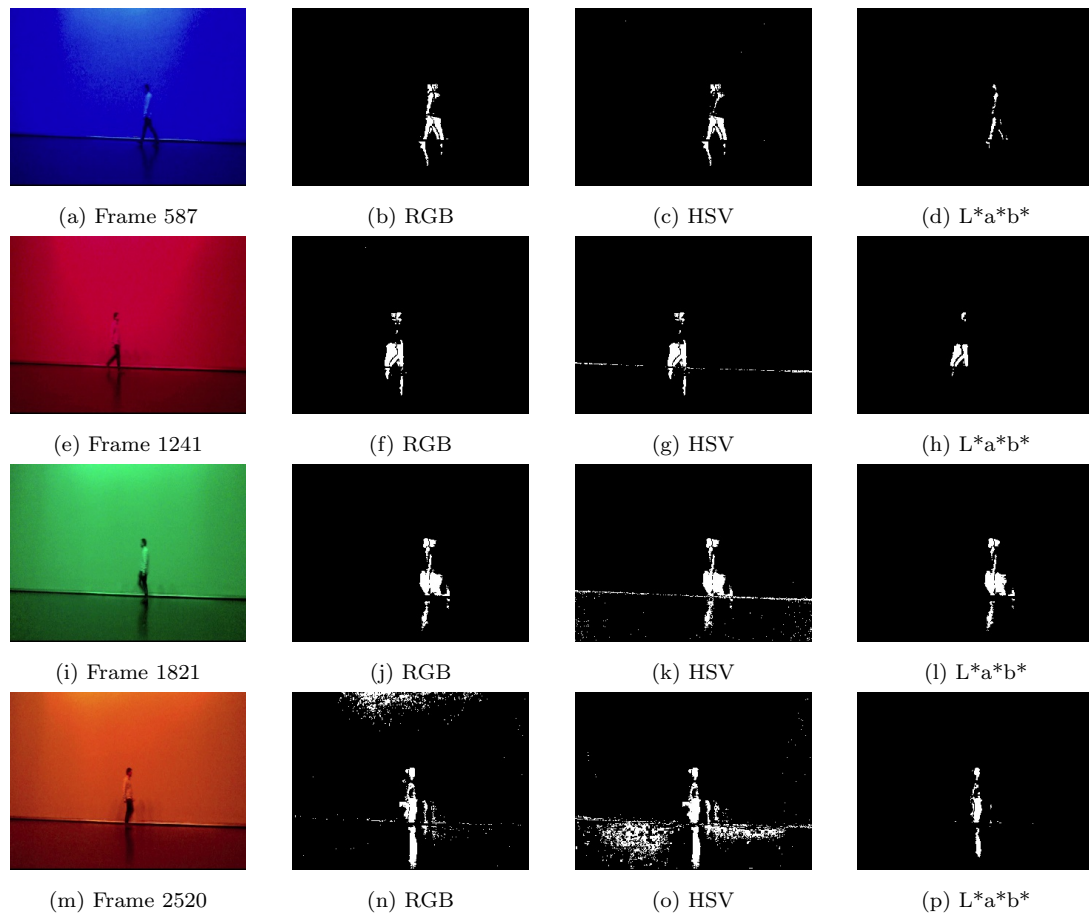


Figura 5.6: Resultados da segmentação utilizando o sistema desenvolvido em condições de iluminação variável

valor de α adequado para esta situação.

5.4 Comparação com Modelo Clássico

Por forma a avaliar a performance do sistema, é feita uma comparação entre o sistema desenvolvido e um modelo clássico. O modelo escolhido foi o Modelo de Mistura de Gaussianas [1], por ser amplamente utilizado e apresentar uma boa performance. Todos os parâmetros que têm que ver com a segmentação de objectos foram inicializados com os mesmos valores, ou seja, o limiar de segmentação utilizado é 25 quer para o sistema desenvolvido, quer para o algoritmo de Mistura de Gaussianas. Os exemplos de segmentação obtidos utilizando o modelo de Mistura de Gaussianas podem ser observados nas figuras 5.7 e 5.8.

Para efectuar a comparação entre o sistema desenvolvido e o modelo clássico, são realizados dois testes em que se adquire um conjunto de frames resultantes da segmentação. O primeiro analisa o caso de iluminação constante e utiliza a segunda e a terceira sequência de vídeo, recolhendo 8 frames em que há presença de elementos activos em cena, duas frames por cada condição de iluminação. No que toca ao caso em que existe variação das condições de iluminação, são utilizadas a quarta e a quinta sequências de vídeo, sendo guardadas duas frames correspondentes a 1 e 2 segundos após haver uma alteração de iluminação da cena. Estas frames são recolhidas para cada tipo de iluminação das sequências, perfazendo um total de 8 frames.

Para cada teste, a saída do sistema é comparada com uma segmentação de ground truth, que pode tomar o valor 0 ou 1 e é comparada com a saída (0 ou 1) do algoritmo automático. Posto isto, é construída uma matriz de confusão calculando as taxas de verdadeiros positivos (TP), verdadeiros negativos (TN), falsos positivos (FP) e falsos negativos (FN), obtidas através de:

$$TP = \frac{tp}{tp + fn} \quad (5.1)$$

$$TN = \frac{tn}{tn + fp} \quad (5.2)$$

$$FP = \frac{fp}{tn + fp} \quad (5.3)$$

$$FN = \frac{fn}{tp + fn} \quad (5.4)$$

em que tp , tn , fp e fn representam o número total de verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos, respectivamente.

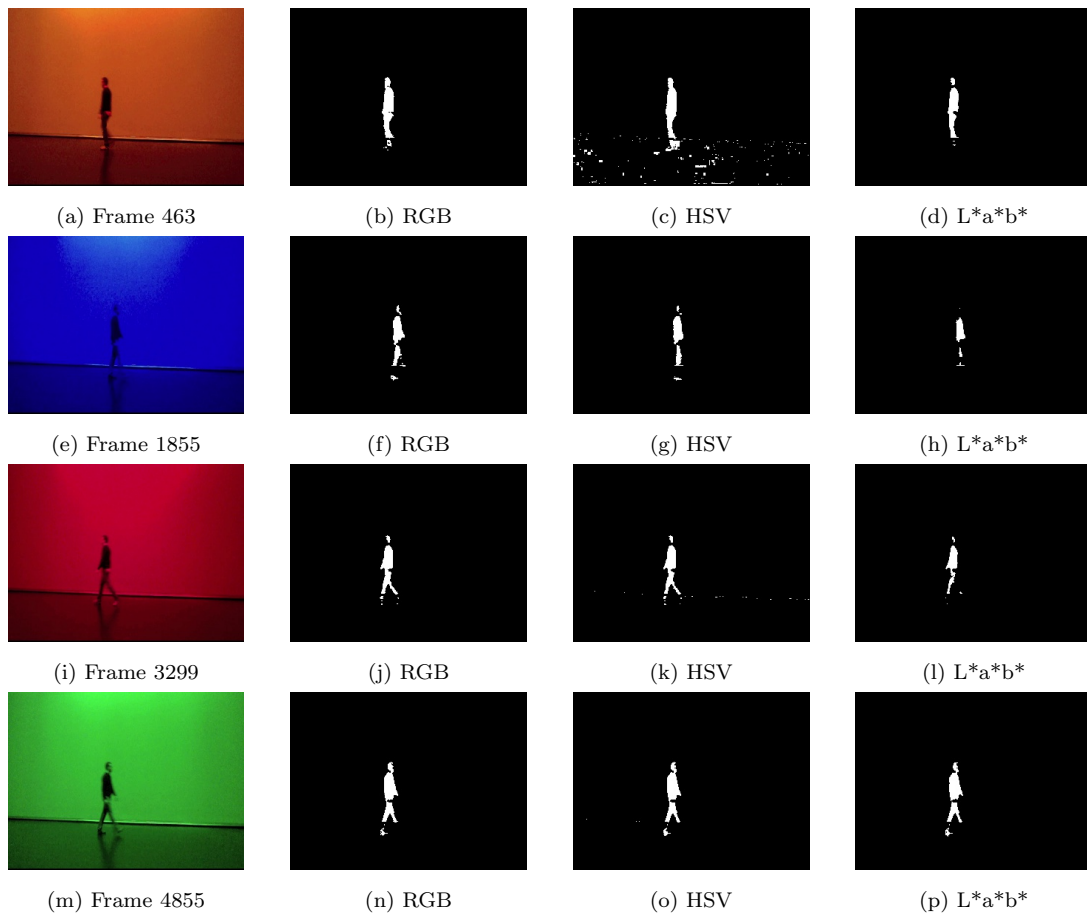


Figura 5.7: Resultados da segmentação utilizando a Mistura de Gaussianas em condições de iluminação constante

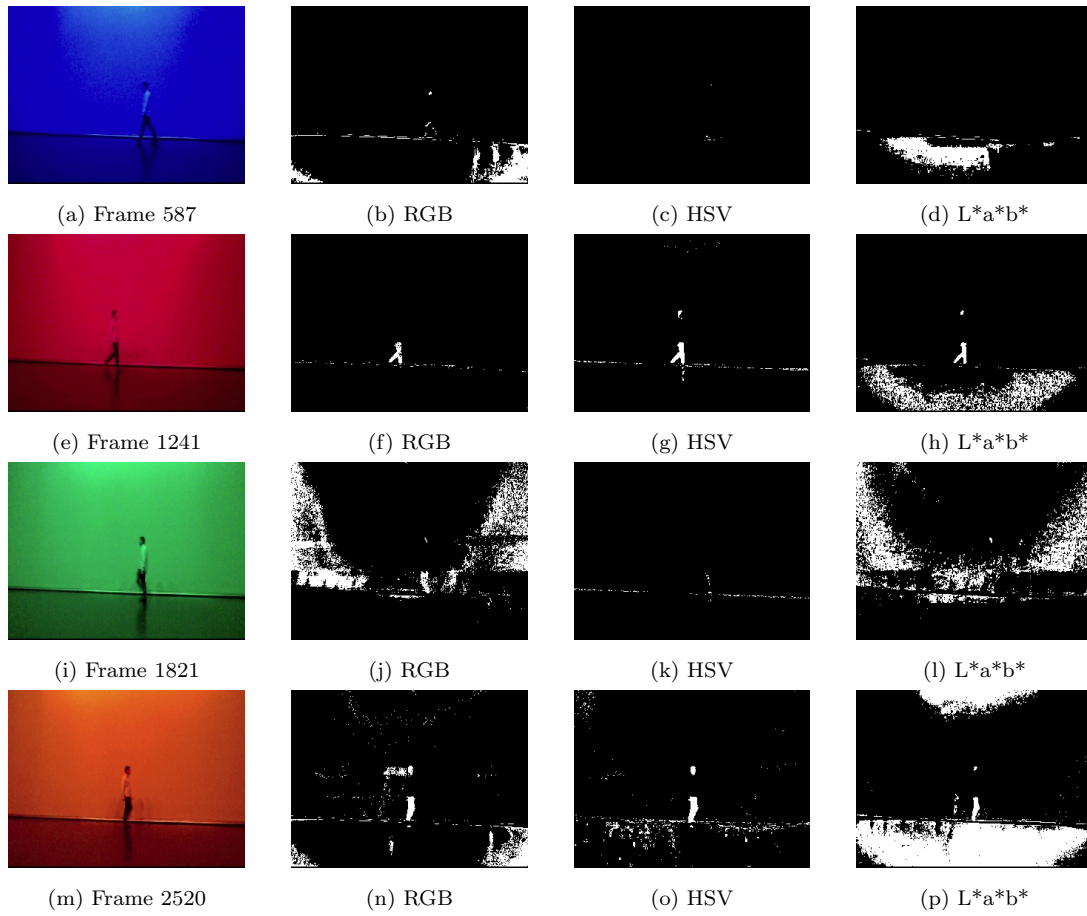


Figura 5.8: Resultados da segmentação utilizando uma Mistura de Gaussianas em condições de iluminação variável

		Classificação RGB		Classificação HSV		Classificação L*a*b*	
		Negativo	Positivo	Negativo	Positivo	Negativo	Positivo
Valor Real	Negativo	98.84 %	1.16 %	97.37 %	2.63 %	99.35 %	0.65 %
	Positivo	5.88 %	94.12 %	5.65 %	94.35 %	12.24 %	87.76 %

Tabela 5.1: Matrizes de Confusão da segmentação usando o sistema implementado face a iluminação constante

		Classificação RGB		Classificação HSV		Classificação L*a*b*	
		Negativo	Positivo	Negativo	Positivo	Negativo	Positivo
Valor Real	Negativo	99.83 %	0.17 %	99.60 %	0.40 %	99.88 %	0.12 %
	Positivo	20.08 %	79.92 %	27.93 %	72.07 %	41.26 %	58.74 %

Tabela 5.2: Matrizes de Confusão da segmentação usando a Mistura de Gaussianas face a iluminação constante

As tabelas 5.1 a 5.4 mostram as matrizes de confusão obtidas para os dois cenários e para os diferentes espaços de cor. Para o caso da segunda e terceira sequências de vídeo (tabelas 5.1 e 5.2), é possível verificar que a segmentação obtida através do sistema desenvolvido nesta tese consegue um melhor desempenho na identificação de elementos activos em cena, já que a sua taxa de verdadeiros positivos é maior que no caso do modelo de Mistura de Gaussianas. Por outro lado, a taxa de falsos positivos é menor no caso da Mistura de Gaussianas, devido ao facto de a sensibilidade do sistema desenvolvido face a variações de intensidade de um pixel ser elevada.

		Classificação RGB		Classificação HSV		Classificação L*a*b*	
		Negativo	Positivo	Negativo	Positivo	Negativo	Positivo
Valor Real	Negativo	98.64 %	1.36 %	97.76 %	2.24 %	99.55 %	0.45 %
	Positivo	42.36 %	57.64 %	44.98 %	55.02 %	55.69 %	44.31 %

Tabela 5.3: Matrizes de Confusão da segmentação usando o sistema implementado face a variações de iluminação

Quanto ao caso em que existem variações das condições de iluminação e elementos activos em cena (tabelas 5.3 e 5.4), os resultados obtidos são bastante satisfatórios, na medida em que a taxa de pixels bem identificados utilizando o sistema desenvolvido é claramente superior ao modelo clássico, além de que a taxa de falsos positivos é também mais baixa, significando que a maior parte das possíveis identificações devido a mudanças das condições de iluminação são absorvidas pela forma como a fase

		Classificação RGB		Classificação HSV		Classificação L*a*b*	
		Negativo	Positivo	Negativo	Positivo	Negativo	Positivo
Valor Real	Negativo	94.95 %	5.05 %	99.47 %	0.53 %	87.20 %	12.80 %
	Positivo	73.59 %	26.41 %	78.17 %	21.83 %	77.96 %	22.04 %

Tabela 5.4: Matrizes de Confusão da segmentação usando a Mistura de Gaussianas face a variações de iluminação

de aprendizagem de novas imagens de fundo é realizada.

Fazendo uma avaliação global dos dados recolhidos, os resultados obtidos podem ser considerados satisfatórios. Os valores escolhidos para os parâmetros do sistema implementado encontram-se adequados, produzindo resultados satisfatórios no que toca tanto à identificação do tipo de iluminação activo em cada instante, tanto no que toca à segmentação das regiões activas, principalmente no no caso da segmentação face a variações das condições de iluminação, onde se observam melhorias significativas em relação ao caso clássico.

Capítulo 6

Conclusão

6.1 Conclusões

Nesta tese foi proposto um sistema de detecção de objectos activos para sistemas de video-vigilância, robusto a variações bruscas de iluminação. A solução proposta consiste em guardar múltiplas aparências da imagem de fundo, cada uma representando um tipo de iluminação diferente, sendo o sistema capaz de identificar em cada instante de tempo qual o tipo de iluminação activo na cena captada, utilizando a informação do modelo de fundo activo na fase de segmentação das regiões activas na imagem. Este processo é realizado sem conhecimento prévio das condições de iluminação utilizadas na sequência de vídeo, ou seja, o sistema tem capacidade de aprender e construir novos modelos de fundo no caso de ser detectada uma condição de iluminação totalmente desconhecida.

Por forma a testar convenientemente o sistema desenvolvido, foi adquirido um conjunto de dados, composto por cinco sequências de vídeo captadas no palco do grande auditório da Culturgest, sendo feita uma análise detalhada das sequências adquiridas no que toca à estabilidade da cor, a presença e distribuição de ruído e a evolução do valor de um pixel à presença de elementos activos em cena. Estes testes foram efectuados para três espaços de cor diferentes: RGB, HSV e $L^*a^*b^*$, sendo que face aos dados obtidos, o espaço de cor $L^*a^*b^*$ foi aquele que obteve melhores resultados, sobretudo devido à sua maior estabilidade ao nível da representação da cor em relação aos espaços RGB e HSV.

As conclusões tiradas através do estudo das sequências de vídeo foram confirmadas nos testes efectuados ao sistema desenvolvido. O primeiro consistiu em verificar a forma como o sistema cria novos modelos de fundo e identificava as diferentes condições de iluminação ao longo do tempo, sendo que o espaço de cor $L^*a^*b^*$ obteve os melhores resultados, fazendo corresponder todos os valores obtidos com os valores esperados. O segundo teste avaliou o algoritmo de segmentação proposto nesta tese, quer para o caso em que as condições de iluminação são constantes durante o trajecto das pessoas em cena, quer no caso em que há alterações nas condições de iluminação. No primeiro caso, verificou-se

que a segmentação foi feita de forma aceitável, sendo que no caso do espaço de cor HSV os resultados contaram com um maior valor de falsos positivos, e mais uma vez se verificou que a menor taxa de falsos positivos foi obtida no espaço de cor $L^*a^*b^*$. No caso da segmentação com variações de iluminação, os resultados obtidos podem ser considerados bons, já que as taxas de falsos positivos obtidas continuaram baixas, ainda que a taxa de falsos negativos tenha aumentado.

Por fim, o sistema proposto nesta tese foi comparado com o modelo clássico de Mistura de Gaussianas [1], avaliando o resultado da segmentação para o caso de iluminação constante e para o caso de iluminação variável. Construídas as matrizes de confusão para todos os casos, os resultados obtidos mostram que o desempenho do sistema proposto é superior ao do modelo clássico, principalmente no caso em que existem variações bruscas das condições de iluminação.

Num balanço geral, a forma como o sistema é capaz de identificar, gerir e actualizar as diferentes condições de iluminação activas em cada instante permite que a detecção de objectos activos seja feita com uma boa performance, mesmo no caso em que o sistema tem de lidar com variações bruscas das condições de iluminação. Face aos resultados obtidos, é possível afirmar que os objectivos propostos nesta tese foram cumpridos.

6.2 Trabalho Futuro

Quanto ao futuro do sistema apresentado, seria desejável que os resultados da segmentação fossem melhorados no caso em que existem variações de iluminação, por forma a reduzir o número de falsos negativos. Outro ponto a acrescentar ao sistema é um algoritmo de seguimento das regiões activas identificadas, por forma a conseguir identificar univocamente cada objecto, resolver cruzamentos entre pessoas e manter a correcta identificação de cada objecto aquando da alteração das condições de iluminação.

Bibliografia

- [1] C. Stauffer and W. Grimson, “Adaptive background mixture models for real-time tracking,” vol. 2, pp. 637–663, 1999.
- [2] J. Duncan and T.-C. Chou, “On the detection of motion and the computation of optical flow,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 14, no. 3, pp. 346–352, mar 1992.
- [3] M. Kass, A. Witkin, and D. Terzopoulos, “Snakes: Active contour models,” *INTERNATIONAL JOURNAL OF COMPUTER VISION*, vol. 1, no. 4, pp. 321–331, 1988.
- [4] S.-Y. Chien, S.-Y. Ma, and L.-G. Chen, “Efficient moving object segmentation algorithm using background registration technique,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 12, no. 7, pp. 577–586, jul 2002.
- [5] T. Bouwmans, “Recent advanced statistical background modeling for foreground detection: A systematic survey,” *Recent Patents on Computer Science*, vol. 4, no. 3, September 2011.
- [6] J. Zheng, Y. Wang, N. L. Nihan, and M. E. Hallenbeck, “Extracting roadway background image: Mode-based approach,” *Transportation Research Record*, vol. 1944, no. 1944, pp. 82–88, 2006.
- [7] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, “Pfinder: real-time tracking of the human body,” in *Automatic Face and Gesture Recognition, 1996., Proceedings of the Second International Conference on*, oct 1996, pp. 51–56.
- [8] D. Butler and S. Sridharan, “Real-time adaptive background segmentation,” 2003.
- [9] O. Alata and L. Quintard, “Is there a Best Color Space for Color Image Characterization or Representation based on Multivariate Gaussian Mixture Model ?” *Computer Vision and Image Understanding*, vol. 113, no. 8, pp. 867–877, Aug. 2009.
- [10] A. Elgammal, D. Harwood, and L. Davis, “Non-parametric model for background subtraction,” vol. 1843, pp. 751–767, 2000.

- [11] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, “Detecting moving objects, ghosts, and shadows in video streams,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 10, pp. 1337 – 1342, oct. 2003.
- [12] L. Li and M. Leung, “Integrating intensity and texture differences for robust change detection,” *Image Processing, IEEE Transactions on*, vol. 11, no. 2, pp. 105 –112, feb 2002.
- [13] J. Rittscher, J. Kato, S. Joga, and A. Blake, “A probabilistic background model for tracking,” 2000, pp. 336–350.
- [14] B. Stenger, V. Ramesh, N. Paragios, F. Coetzee, and J. M. Buhmann, “Topology free hidden markov models: Application to background modeling.” in *ICCV’2001*, 2001, pp. 294–301.
- [15] A. Monnet, A. Mittal, N. Paragios, and V. Ramesh, “Background modeling and subtraction of dynamic scenes,” in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, vol. 2, oct. 2003, pp. 1305 –1312.
- [16] J. Zhong and S. Sclaroff, “Segmenting foreground objects from a dynamic textured background via a robust kalman filter,” in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, vol. 1, oct. 2003, pp. 44–50.