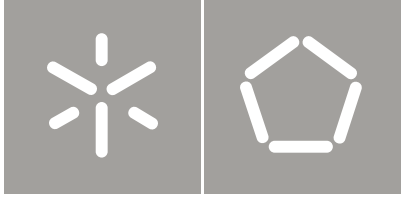
Universidade do Minho
Escola de Engenharia

Nuno Pedro Rodrigues Peixoto

Videovigilância Inteligente em Ambientes Aquáticos: Detecção Precoce de Afogamento em Piscinas Domésticas

Setembro de 2010



Universidade do Minho
Escola de Engenharia

Nuno Pedro Rodrigues Peixoto

Videovigilância Inteligente em Ambientes Aquáticos: Detecção Precoce de Afogamento em Piscinas Domésticas

Tese de Doutoramento
Área de Informática Industrial

Trabalho efectuado sob a orientação de
Professor Doutor José Mendes
Professor Doutor Adriano Tavares

Setembro de 2010

Agradecimentos

Em primeiro lugar gostaria de agradecer aos meus orientadores científicos, o Doutor José Mendes e o Doutor Adriano Tavares, que me guiaram ao longo dos trabalhos de doutoramento e se mostraram sempre disponíveis para esclarecer as minhas dúvidas. A sua ajuda no que se refere à tomada de decisões foram fundamentais para que, no fim, se chegasse a bom porto.

Gostaria também de deixar uma nota de agradecimento aos meus pais, Mário José Teixeira Peixoto e Maria de Lurdes Dias Rodrigues, que me incentivaram a atingir este nível, tendo sempre proporcionado todas as condições para tal, por vezes, com enorme sacrifício.

Agradeço ainda à minha namorada, Marta Abreu, que acompanhou esta minha escalada, fornecendo o conforto emocional e o apoio moral, tão necessários nos momentos mais difíceis. Ao meu irmão, Mário Peixoto, agradeço a disponibilidade de me ouvir e ajudar no desenho de alguns esquemas 3D da piscina utilizados nesta tese. Ao meu amigo Miguel Abreu agradeço a disponibilidade da piscina para os testes do sistema.

Por último fica um agradecimento especial àqueles que estiveram perto de mim durante todos os dias de trabalho e com os quais pude conversar acerca dos mais variados assuntos, Nuno Cardoso, Nuno Brito, André Mota, Paulo Santos, Bruno Garim, Nikhil Girdhar e Patrício Teixeira.

Não posso deixar de agradecer à Fundação para a Ciência e a Tecnologia pelo apoio concedido no âmbito do Programa de Bolsas de Formação Avançada, através da concessão e financiamento da bolsa de Doutoramento com a referência SFRH/BD/28227/2006 durante quatro anos.

Resumo

Videovigilância Inteligente em Ambientes Aquáticos: Detecção Precoce de Afogamento em Piscinas Domésticas

Neste documento, apresenta-se um conjunto de trabalhos relativos à utilização de videovigilância inteligente para a análise de comportamentos humanos em ambientes aquáticos, orientado à detecção precoce de afogamento em piscinas domésticas. Recorrendo a imagens de vídeo capturadas com uma câmara comum de videovigilância e um computador pessoal para efectuar todas as análises necessárias ao reconhecimento de pessoas e à inferência do seu comportamento no ambiente aquático referido de forma autónoma. Trata-se, portanto, de um trabalho baseado em videovigilância inteligente com a capacidade de autonomamente monitorizar as actividades dos utilizadores de uma piscina doméstica despoletando um alerta em caso de detecção de afogamento. A solução apresentada é composta por três módulos principais de processamento que operam de forma sequencial: segmentação de imagens, seguimento de objectos e reconhecimento de comportamento.

O primeiro módulo captura uma imagem proveniente de uma câmara e efectua a segmentação fornecendo uma imagem binária com todos os objectos que não pertencem à cena. Neste campo foram desenvolvidas novas técnicas de segmentação de movimento por estimação de *background* que conseguem eliminar os efeitos das reflexões especulares variáveis e a oscilação da superfície da água muito características de ambientes aquáticos. Neste módulo é ainda efectuado um segundo passo de segmentação dentro do rectângulo que engloba as regiões encontradas na primeira fase da segmentação proporcionando melhoramentos muito significativos nos resultados obtidos através do algoritmo *k-means*.

O segundo módulo, de seguimento de objectos, recebe a imagem binária gerada pelo módulo anterior e identifica cada objecto exclusiva e consecutivamente. Este módulo utiliza um algoritmo de correspondência não balanceado que permite a atribuição dos objectos detectados no fotograma actual aos objectos previstos para o fotograma seguinte. A previsão dos objectos é efectuada por intermédio de um filtro de *kalman*, sendo a correspondência baseada na similaridade de características dos objectos. Este módulo permite a identificação e seguimento de vários objectos em simultâneo tratando oclusões temporárias parciais e totais.

O último módulo recebe as características principais de cada objecto, fornecidas pelo módulo de seguimento, efectua o reconhecimento e inferência do seu comportamento. Este módulo mede várias características dos objectos ao longo do tempo e classifica os padrões encontrados por intermédio de uma árvore de decisão baseada em classificadores *naive Bayes*. Esta árvore permite assim distinguir pessoas de outros objectos e saber se estão localizados na piscina ou fora da piscina. Por último, para todos os objectos identificados como pessoas que se encontram na piscina, este módulo infere o seu comportamento com base nas decisões da árvore associadas a uma máquina de estados finita. Aquando da detecção de um comportamento de risco, que revele um potencial afogamento, são desencadeados alertas de vários níveis, de acordo com o risco associado.

Finalmente foi desenvolvido um protótipo para implementação dos algoritmos desenvolvidos que permite, através de vídeos gravados numa piscina doméstica, testar o completo funcionamento do sistema. Nos vídeos recolhidos foram simuladas situações de afogamento de vários tipos, tendo em atenção as várias fases associadas ao afogamento, de modo a tornar o teste o mais real possível. Os resultados obtidos neste trabalho demonstram a viabilidade da implementação de um sistema para a detecção precoce de afogamento em ambientes domésticos.

Abstract

Intelligent Video Surveillance Analytics in Aquatic Environments: Early Drowning Detection at Domestic Swimming Pools

This document describes all the work carried out in the development of a project of an intelligent video surveillance analytics solution for human behaviour analysis in aquatic environments, dedicated to the early drowning detection at domestic swimming pools. Using video images captured by a standard video surveillance camera and a personal computer to carry out all the needed analysis towards the recognition of people and the deduction of their behaviour in the referred aquatic environment in a completely autonomous way with the ability of generating alarms. The solution devised in this project is composed by three modules operating in a sequential way: image segmentation, object tracking and human behaviour recognition.

The first module is responsible for image capturing from a standard video surveillance camera and providing as output a binary image of all objects that do not belong to the background. In this module, new segmentation techniques for background estimation were developed capable of eliminating the problems caused by specular reflections on the water as well the normal oscillation of the surface of the water. In this module, it also shown that a two step segmentation can improve its results in a significant way by the application of the k-means algorithm in the objects obtained in the first step of this module.

The second module, object tracking, uses the output of the first module to identify each object in an unequivocal and exclusive way. This module uses an algorithm of non balanced correspondence which allows the establishment of a link of the same object in sequential images. The object location prediction from a frame to the next frame is achieved using a *kalman* filter, where the link between objects is based in the similarity of the characteristics of those objects. This module allows the unequivocal identification and tracking of several objects at the same time with a robust performance even in the case of total or partial occlusions.

The last module input are the main characteristics of each object, which is responsible for the behaviour recognition of each object. Several characteristics of each object are measured over time and the found patterns classified using a decision tree based in *naive Bayes* classifiers. This

decision tree allows the separation between people and other objects as well as knowing if they are located in the water or outside the water. Further, for all objects identified as people, this module analyses its behaviour based on finite state machine. Whenever, a potential early drowning event is detected several alerts are raised according to the calculated risk.

Finally, all the modules were integrated in a prototype that is capable of testing the full operation of the solution proposed, using recorded videos in a domestic swimming pool. In the videos several simulations of drowning situations were created in order to emulate the different phases of a drowning event. The results achieved in this project show the viability of the implementation of such a solution for early drowning detection in domestic swimming pools.

Índice

Siglas e Acrónimos

APSI	Associação para a Promoção da Segurança Infantil
CCD	*Charge-coupled Device*
CIE	*Commission Internationale d'Eclairag*
DEWS	*Drowning Early Detection System*
EKF	*Extended Kalman Filter*
EM	*Expectation Maximization*
FAR	*False Acceptance Rate*
FRR	*False Rejection Rate*
FSM	*Finite State Machine*
HMM	*Hidden Markov Model*
HSV	*Hue Saturation Value*
IIR	*Infinite Impulse Response*
JPDAF	*Joint Probability Data Association Filter*
KDE	*Kernel Density Estimation*
LED	*Light Emission Diode*
LSE	*Least-Squares Error*
MAP	*Maximum a Poteriori*
MHT	*Multiple Hypothesis Tracking*
MOG	*Mixture of Gaussians*
NB	*Naive Bayes*
NIR	*Neutral Interface Reflection*
PCA	*Principal Component Analysis*
PMHT	*Probabilistic Multiple Hypothesis Tracking*
RGB	*Red Green Blue*
RM	*Reduced Model*
SIFT	*Scale-Invariant Feature Transform*
SVM	*Support Vector Machine*
TVS	*Temporal Spatio-Velocity*
UKF	*Unscented Kalman Filter*

Índice de Figuras

Índice de Tabelas

Índice de Algoritmos

Capítulo 1

1.1 Introdução

A evolução tecnológica ao nível dos computadores e dos dispositivos de captura de imagem permitiram o rápido avanço dos sistemas de vigilância baseados em vídeo. Actualmente assiste-se a uma necessidade de segurança cada vez maior, principalmente depois da crescente vaga de terrorismo que se tem vindo a sofrer a uma escala global. Exemplos disso são os ataques de 11 de Setembro de 2001 nos Estados Unidos da América, 11 de Março de 2004 em Madrid e 7 de Julho de 2005 em Londres. Todas estas preocupações fazem voltar a comunidade científica para o estudo e desenvolvimento de soluções de vídeo vigilância inteligentes, capazes de detectar e inferir, automaticamente, comportamentos das pessoas em cena, de modo a alertar os operadores de segurança na eventualidade da ocorrência de uma situação anormal. Deste modo, a videovigilância engloba um vasto conjunto de aplicações. Entre essas aplicações destacam-se a monitorização de pessoas em aeroportos e estações de metro bem como de veículos em auto estradas, locais públicos tais como bancos, centros comerciais e parques de estacionamento ou locais privados como habitações.

Nesta tese aborda-se uma aplicação muito específica no âmbito da vídeo vigilância inteligente associada à monitorização de actividades humanas. Todos os trabalhos foram centrados no desenvolvimento de um sistema de detecção precoce de vítimas de afogamento para utilização em piscinas tipicamente domésticas. Os objectivos de um sistema deste tipo prendem-se com a capacidade de detectar pessoas num ambiente aquático complexo com fiabilidade e robustez, analisar o seu comportamento e tirar conclusões sobre a situação das mesmas gerando um alerta caso esse estado indique o início de um afogamento.

O ambiente aquático associado à piscina é caracterizado pela sua complexidade derivada das oscilações da água e das consequentes reflexões de luz causadas por essa mesma oscilação. Este dinamismo associado à cena exige algoritmos de processamento de imagem capazes de solucionar o problema de um *background* continuamente variável. Um sistema deste tipo é normalmente composto por três estágios essenciais, refira-se: a detecção de objectos, o seguimento e por último a análise do seu comportamento com inferência acerca do seu estado.

1.2 Motivação

Num relatório publicado pela Organização Mundial de Saúde (Krug, 1999), em 1998, quase meio milhão de mortes foram causadas devido a acidentes resultantes de afogamento e 57% desses casos aconteceram com crianças até aos 14 anos de idade. Um relatório mais recente, publicado pela Unicef (Adamson, Micklewright, & Wright, 2001), concluiu que nos 26 países mais ricos do mundo as causas de morte por acidente em crianças são as mais frequentes. No tipo de acidentes encontram-se em primeiro lugar os acidentes rodoviários e em segundo lugar os afogamentos. Na União Europeia, mais de 70% das vítimas de afogamento são do sexo masculino com idades compreendidas entre 1 e 4 anos (Boshuizen, Treurniet, & Marteloh, 1997). Mas mesmo nos casos de acidente que não resultam em morte podem resultar danos cerebrais irreversíveis que irão incapacitar a vítima para o resto da vida. Um estudo vocacionado para a análise dos custos resultantes da submersão de vítimas (CPSC, 2001) concluiu que, nos Estados Unidos da América, mais de 4000 dólares são gastos numa vítima que recupere totalmente, sendo que, nos casos em que as vítimas sofrem lesões neurológicas graves os custos podem ascender aos 160000 dólares. Portugal segue as tendências encontradas em estudos realizados sobre a União Europeia, tal como a Associação para Promoção da Segurança Infantil (APSI) tem vindo a esclarecer nos seus relatórios anuais. Estudos efectuados por esta entidade demonstram que em crianças mais novas o afogamento ocorre sobretudo em piscinas privadas, pertencentes aos próprios familiares das vítimas. Este é um caso muito grave de saúde pública que todos os anos se mostra cada vez mais intenso, sobretudo porque o número de piscinas também tem vindo a aumentar, gerando potenciais perdas de anos de vida.

Contudo, a vigilância autónoma, baseada na análise inteligente de vídeo, seria uma solução muito eficaz, capaz de gerar um alerta sempre que uma potencial situação de afogamento fosse detectada. As causas mais comuns de afogamento, tal como a queda de crianças em piscinas domésticas seriam facilmente detectadas por um sistema deste género, sendo deste modo possível salvá-las atempadamente. Ao ser gerado um alerta, das mais variadas formas, é possível que um adulto, que normalmente está sempre por perto, possa acorrer à zona da piscina para salvar a criança. A motivação que se prende com este trabalho está directamente relacionada com a possibilidade de combater um problema persistente há já muito tempo e com tendência a aumentar. Numa perspectiva tecnológica, a motivação advém da possibilidade do estudo e

expansão de conhecimentos em áreas como o processamento e análise de imagem, visão por computador, inteligência artificial e aprendizagem automática.

1.3 Desafios

Os maiores desafios deste trabalho encontram-se na dificuldade imposta pela cena exterior na detecção das zonas correspondentes aos objectos presentes na piscina. O movimento continuamente oscilatório da água faz com que seja extremamente difícil segmentar a imagem, uma vez que não é possível fazer uma estimação e como tal uma modelação efectiva do *background*, falhando deste modo as técnicas tradicionais de segmentação (Eng, Toh, Yau, & Wang, 2008). A água apresenta índices de reflexão elevados, o que, combinado com a oscilação provoca o aparecimento de múltiplas reflexões da fonte luminosa. Além da reflexão, a refracção provocada pela água constitui outro problema, pois os objectos mergulhados aparecem tanto mais distorcidos quanto maior for a oscilação da água à superfície. Em certos casos a oscilação é tanta que objectos mergulhados são completamente ofuscados pela água, não sendo possível, nem mesmo para um ser humano detectar a presença desses objectos. O seguimento de vários objectos em simultâneo com tratamento de oclusões, entradas e saídas da cena constituem um enorme desafio no trabalho proposto. A lista de desafios conta ainda com a necessidade de distinguir objectos de pessoas, inferindo o comportamento destas ao longo do tempo de modo a encontrar padrões de afogamento. Por último, o objectivo de efectuar toda esta análise recorrendo apenas a câmaras exteriores e ao menor número possível é mais um desafio.

1.4 Objectivos da investigação

Os objectivos deste trabalho de doutoramento centram-se na implementação de um sistema de videovigilância inteligente, de terceira geração, com a finalidade de detectar potenciais vítimas de afogamento, durante as actividades normais, envolvidas na utilização de uma piscina. A arquitectura de software do sistema é baseada num *pipeline* constituído por três módulos de processamento, tal como a Figura 1.4.1 evidencia.

O módulo de segmentação destaca os objectos de interesse da cena, fornecendo como saída uma imagem binária com a área ocupada por estes na imagem. De seguida, o módulo de seguimento irá identificar, exclusivamente ao longo do tempo, cada objecto encontrado na fase anterior. Numa última fase o módulo de análise de comportamento tem como objectivo a recolha e classificação de padrões presentes nas características dos objectos seguidos de modo a inferir as suas acções.

Caso o comportamento inferido coincida com uma acção correspondente a uma situação de afogamento é gerado um alerta.

Em termos de arquitectura o sistema é constituído por uma ou mais câmaras externas, dependendo do tamanho da área a vigiar, e um computador comum. Cabe a esta unidade central de processamento recolher as imagens provenientes das câmaras e efectuar os vários processos descritos anteriormente.

Figura 1.4.1: Arquitectura de software do sistema.

1.5 Contributos

Durante os trabalhos efectuados foi gerado novo conhecimento em várias áreas, nomeadamente, nas áreas relativas aos diferentes módulos de processamento que compõem o sistema. Deste modo, foi concebido um algoritmo híbrido de segmentação de movimento especializado para ambientes aquáticos complexos nos quais o *background* é altamente variável. O processo de segmentação é composto por duas fases. Na primeira fase é gerada uma máscara binária com os objectos detectados e na segunda fase é efectuado um segundo passo de segmentação a cada objecto individualmente de modo a aumentar a qualidade do processo. No que respeita ao módulo de seguimento foi concebido um algoritmo capaz de lidar com entradas, saídas e oclusões de objectos. Este algoritmo é baseado numa adaptação da solução do problema da correspondência não balanceada em conjunto com filtros de *Kalman* para inferir as posições dos objectos detectados. Por último foi desenvolvido um esquema de reconhecimento de objectos e inferência do seu comportamento descrito por uma máquina de estados finita. A comutação de estados nessa máquina é proporcionada pela detecção de padrões nas características extraídas dos objectos por intermédio de uma árvore de decisão baseada em classificadores probabilísticos

gerados através de aprendizagem supervisionada. De seguida resumem-se as contribuições resultantes do trabalho efectuado:

a) Novo algoritmo de segmentação de movimento especialmente adaptado para ambientes aquáticos complexos baseado na componente tonalidade do espaço de cores HSV (Peixoto, Cardoso, Cabral, Tavares, & Mendes, 2009).
b) Algoritmo de segmentação de movimento híbrido com re-segmentação por objectos intitulado de "*Hybrid Motion Segmentation for Object Detection on Complex Aquatic Scenes*". Submetido para *IEEE Transactions on Industrial Informatics*.

1.6 Organização da tese

Nos próximos capítulos são abordados cada um dos três módulos principais de processamento que compõem o sistema. Assim, o Capítulo 2 trata o problema da segmentação de movimento num ambiente aquático complexo, o Capítulo 3 incide sobre o seguimento de vários objectos em simultâneo e o Capítulo 4 aborda o reconhecimento de objectos e a inferência do seu comportamento. Os resultados do teste global do sistema em ambiente real são analisados no Capítulo 5. Finalmente, no Capítulo 6 reúnem-se as conclusões finais e perspectivam-se algumas linhas de orientação para trabalho futuro.

Capítulo 2

2 Segmentação de Objectos em Movimento num Ambiente Aquático Complexo

2.1 Introdução

O módulo de segmentação de movimento é o primeiro do sistema e tem como objectivo gerar uma imagem binária com as regiões que não façam parte da cena, isto é, com os objectos que nela se deslocam. A entrada deste módulo corresponde a uma sequência de imagens no formato RGB provenientes de uma câmara comum de videovigilância. O mapa binário gerado não deve conter regiões onde existam reflexões especulares e movimentos associados à água. A colocação ou remoção de novos objectos da cena, isto é, alterações à cena também não devem aparecer no mapa binário de *foreground*, pelo menos, durante muito tempo. Deve assim existir actualização do *background* de modo a comportar variações do mesmo, causadas, por exemplo, pela variação das condições atmosféricas.

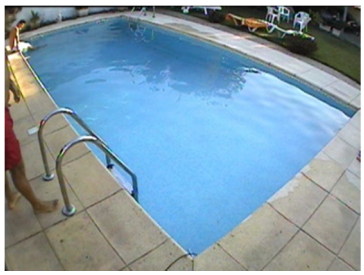
Neste segundo capítulo são abordados os problemas subjacentes à detecção de movimento num ambiente aquático complexo. As técnicas actuais na área da segmentação de movimento ainda demonstram várias lacunas no perfeito cumprimento desse objectivo. Em seguida destaca-se o estado da arte relativamente às técnicas de segmentação de movimento por estimação de background desenvolvidas até ao momento. Numa terceira fase sugere-se um modelo de representação da superfície da água que contenha as características essenciais da mesma com relevância para a segmentação. A escolha de um modelo de representação de cor que minimiza os problemas encontrados é sugerida. Essa sugestão é baseada no modelo de representação da superfície da água e no modelo de reflexão dicromático. Utilizando o espaço de representação de cor escolhido é sugerida uma técnica de segmentação híbrida que permite detectar os objectos presentes na piscina. Os resultados experimentais em diferentes situações que causam os problemas expostos são depois destacados de modo a avaliar a qualidade do algoritmo, sendo em última instância comparado com outro algoritmo com o mesmo objectivo descrito na secção correspondente ao estado da arte. Finalmente tem lugar a discussão dos resultados experimentais obtidos e as necessárias conclusões depois das comparações efectuadas.

2.2 Problemas Associados à Segmentação de Movimento num Ambiente Aquático

A segmentação de objectos em movimento faz normalmente parte do primeiro estágio de processamento num sistema de vídeo vigilância inteligente de terceira geração (Valera & Velastin, 2005). Contudo, esta operação é muito dependente da própria cena não existindo nenhum algoritmo que obtenha bons resultados em todos os casos, mas sim, existindo vários algoritmos, cada um deles mais adaptado para determinado tipo de cenário do que outros. De seguida apresenta-se um levantamento dos problemas mais importantes que afectam negativamente o processo de segmentação de movimento baseado nos métodos tradicionais de subtracção de *background* com modelação de *background* adaptativa.

No caso da segmentação de objectos em movimento numa piscina, um tipo de cena tipicamente exterior, existe uma forte influência da variação da intensidade luminosa da luz solar, tal como as imagens (a) e (b) da Figura 2.2.1 pretendem ilustrar. Nesta figura apresenta-se também o desaparecimento da sombra causada pela escada da piscina na imagem (b). Além disso, a fonte de luz também varia a sua posição relativamente à cena, ao longo do dia, acompanhando a variação da intensidade luminosa, alterando a forma e a posição das sombras bem como das reflexões especulares, como se pode verificar nas imagens da Figura 2.2.2.

a)

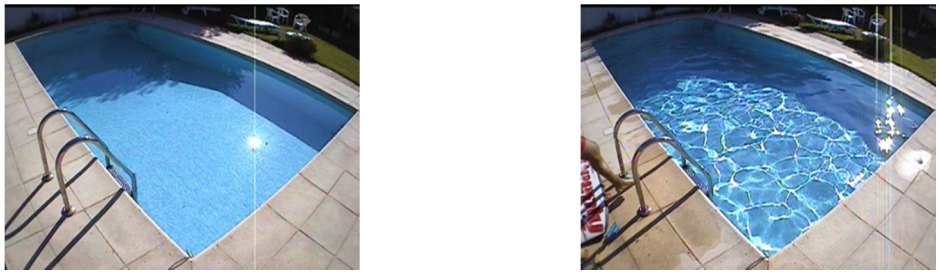


b)

Figura 2.2.1: Variação da intensidade da luz solar ao longo do dia.

a) Iluminação da cena pela luz do sol ao início da tarde;
b) Iluminação da cena pela luz do sol, quando este começa a desaparecer, ao final da tarde.

As condições atmosféricas também são um factor muito importante a considerar, tendo um forte impacto na quantidade de luminosidade que atinge a cena. Na Figura 2.2.3 podem ser comparadas duas situações em diferentes condições atmosféricas. Em (a) o céu está limpo e em (b) o sol está encoberto por nuvens.



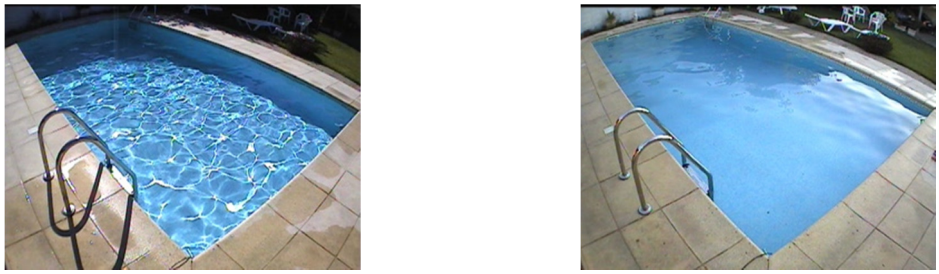
a) b)

Figura 2.2.2: Variação da posição da fonte de luz ao longo do dia, tendo influência também na variação da forma da projecção das sombras, tal como se pode verificar na escada e na posição das reflexões especulares.

a) Imagem capturada ao meio da tarde;
b) Imagem capturada no final da tarde. As sombras aparecem mais esticadas e as reflexões especulares estão mais perto do canto direito da piscina.

Repare-se nas diferenças entre as duas cenas, principalmente no que diz respeito à superfície da água e às sombras. A cena em (a) apresenta sombras e as projecções da luz solar no fundo da piscina, enquanto na cena em (b) não existem sombras, tal como se pode ver na zona envolvente da escada e as projecções da fonte de luz no fundo da piscina não existem. No entanto, em (b), existem as reflexões especulares dos objectos que rodeiam a piscina, tal como a reflexão da imagem das árvores no canto direito mais longínquo do observador. É ainda de considerar o facto da própria câmara implementar correcções na quantidade de luz permitida na excitação do sensor CCD (*Charge-coupled Device*), resultando numa diferença menor entre as imagens (a) e (b) da Figura 2.2.3. De qualquer modo, o problema aparece quando existem transições de um estado para outro, ocorrendo com elas a variação brusca da intensidade luminosa sobre toda a cena e causando problemas na adaptação do modelo de *background*.

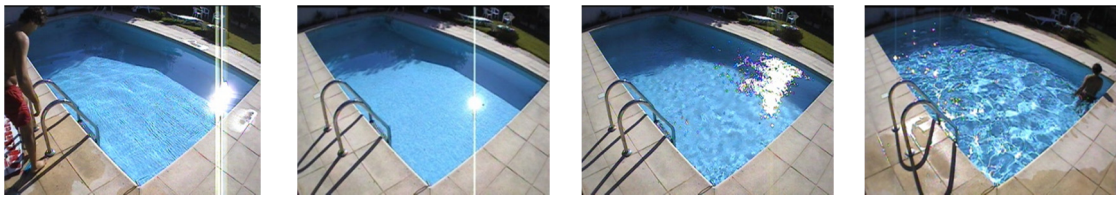
Além dos problemas característicos de uma cena exterior, descritos anteriormente, existem outros associados ao movimento oscilatório de árvores ou vegetação, em geral, causados pelo vento. A investigação nesta área permitiu a resolução parcial deste problema com recurso à soma da variação do fluxo óptico (Horn & Schunk, 1980) nesses objectos, anulando a detecção de movimentos deste tipo (Collins T. R., et al., 2000). Contudo, num ambiente aquático característico de uma piscina existe o problema da oscilação contínua da superfície da água. Esta oscilação não segue nenhum padrão, pois é causada pelo choque de ondas de diferentes frequências, direcções e amplitudes, conferindo características aleatórias ao movimento da água. Este movimento causa o aparecimento de reflexões especulares em diferentes posições e com diferentes formas ao longo do tempo, tal como as imagens da Figura 2.2.4 pretendem demonstrar.



a) b)

Figura 2.2.3: Comparação entre a mesma cena à mesma hora do dia com o céu limpo e com o céu nublado. A quantidade de luz que é irradiada para a cena é muito menor em b.

a) Iluminação da cena pela luz do sol em condições de céu limpo;
b) Iluminação da cena pela luz do sol em condições de céu nublado.



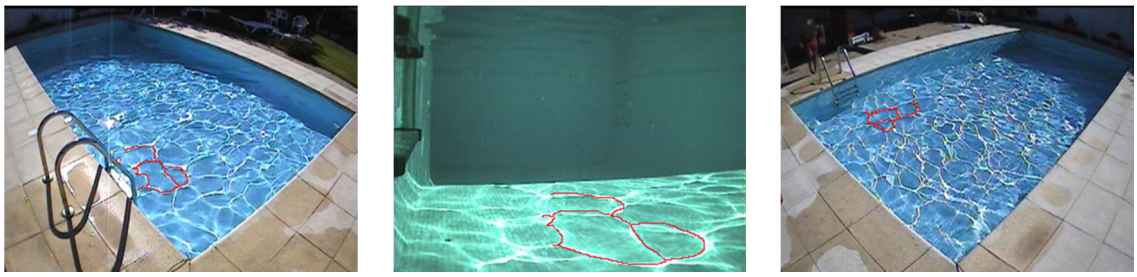
a) b) c) d)

Figura 2.2.4: As reflexões especulares causadas pela água à superfície podem assumir formas muito diferentes. Quando as reflexões são demasiado intensas, tal como nas imagens (a) e (b), o sensor CCD sofre uma saturação que causa o aparecimento das linhas verticais que passam nos pontos de reflexão.

a) Reflexão especular muito intensa, causada pela sobreposição de várias reflexões muito próximas e semelhantes às da imagem b);
b) Reflexão especular perfeita da fonte de luz. Neste caso a superfície da água apresenta uma oscilação mínima, comportando-se como um espelho;
c) Reflexão especular causada por uma oscilação de elevada frequência, mas de baixa amplitude;
d) Múltiplas reflexões especulares em localizações dispersas, devido à elevada amplitude da oscilação da água.

Outro aspecto muito importante a ter em consideração está relacionado com a densidade da água, que por ser diferente da do ar, faz com que a luz seja refractada. Esta refracção causa o aparecimento de múltiplas projecções da fonte luminosa no fundo da piscina, como se pode observar na imagem (b) da Figura 2.2.5. Além disso, e porque essas projecções são de novo reflectidas para a superfície da água, ao mudar novamente de meio, agora da água para o ar, a luz sofre nova refracção e a posição das reflexões vista de fora não coincide com a real posição no fundo da piscina. Este fenómeno é semelhante ao de um lápis inserido num copo cheio de água. É de notar o facto das reflexões especulares, mostradas na Figura 2.2.4, nada terem a ver com as projecções da luz no fundo da piscina. São na realidade dois fenómenos diferentes, mas em que

ambos resultam no aparecimento de brilhos semelhantes na superfície da água, sendo a única diferença a intensidade de luz presente numa reflexão especular ser por norma maior.



a) b) c)

Figura 2.2.5: As projecções da fonte de luz no fundo da piscina, imagem (b), assumem um padrão altamente variável de acordo com a oscilação à superfície da água. Estas projecções são causadas pela refracção da luz ao atingir um meio diferente do ar, no caso, a água. Foram marcadas a vermelho algumas projecções formando um padrão que é visível do exterior.

a) As projecções no fundo da piscina são visíveis à superfície, mas devido à refracção da luz, agora no sentido inverso, da água para o ar, a sua posição no fundo da piscina vista de fora é enganadora;
b) Projecções no fundo da piscina, captadas com recurso a uma câmara submersa;
c) O mesmo padrão pode ser encontrado num ponto de vista mais longínquo.

Todos estes aspectos relacionados com o movimento da superfície da água fazem com que seja extremamente difícil estimar a cena que prevalece a maior parte do tempo, isto é, o *background*, uma vez que o mesmo se encontra em movimento e é afectado por um vasto conjunto de condicionantes que o tornam imprevisível. Com os algoritmos tradicionais de estimação do plano de fundo (Stauffer & Grimson, 1999) o mapa binário correspondente ao *foreground* acaba por ser toda a área da piscina, razão pela qual se torna muito difícil separar os objectos presentes na mesma, ainda que estes estejam em movimento, pois na realidade tudo está em movimento. Na Figura 2.2.6 assiste-se a um caso muito comum numa piscina com nadadores, os salpicos. Estes aparecem e desaparecem de forma repentina e caracterizam-se por uma mancha branca que oculta normalmente outros objectos presentes na piscina. Os salpicos são causados pelas actividades dos nadadores presentes numa piscina e têm um impacto profundamente negativo na estimação do *background*. O problema advém principalmente do facto de os salpicos aparecerem e desaparecerem muito rapidamente, o que torna impossível, em tão curto espaço de tempo, adaptar o *background*. Deste modo, a detecção de movimento por subtracção de *background* resulta no aparecimento de falsos positivos pertencentes ao *foreground* nas zonas afectadas pelos salpicos. Em casos como o da Figura 2.2.6 (c), onde o nadador se esconde por detrás da mancha

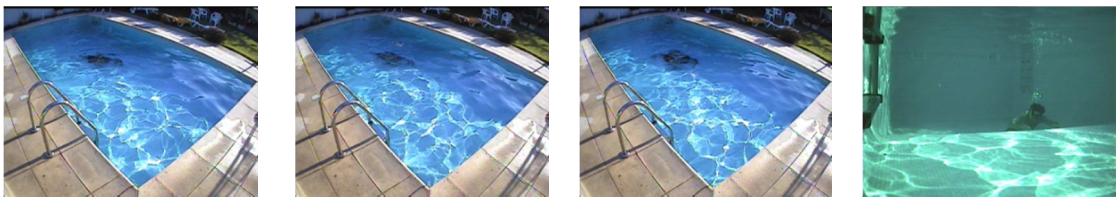
de salpicos a situação é ainda pior pois causa o desaparecimento do nadador, dificultando a tarefa de segmentação.

a) b) c)

Figura 2.2.6: Salpicos e bolhas de ar causados pelos nadadores. Estes aparecem como manchas brancas repentinas que logo desaparecem. Assemelham-se às reflexões especulares e às projecções da fonte de luz no fundo da piscina.

a) Ao nadar, normalmente o movimento dos braços de um nadador causa salpicos;
b) O nadador ao mergulhar para a piscina provoca uma quantidade enorme de salpicos;
c) Por vezes os nadadores provocam a sua própria oclusão ao criarem salpicos com os braços, sem estarem necessariamente a afogar-se.

Além dos problemas que o *background* altamente variável pode causar no processo de estimação do plano de fundo existem ainda outros problemas. Para isso atente-se nas imagens da Figura 2.2.7. Apesar do corpo submerso e estático, visto de fora, a superfície da água em oscilação faz parecer que este corpo tem movimento relativamente a ele próprio. A imagem (d) da mesma figura pretende mostrar que o corpo está estático, não só em relação à piscina, mas também em relação a ele próprio. De qualquer modo, visto de fora não é essa a ideia com que se fica.

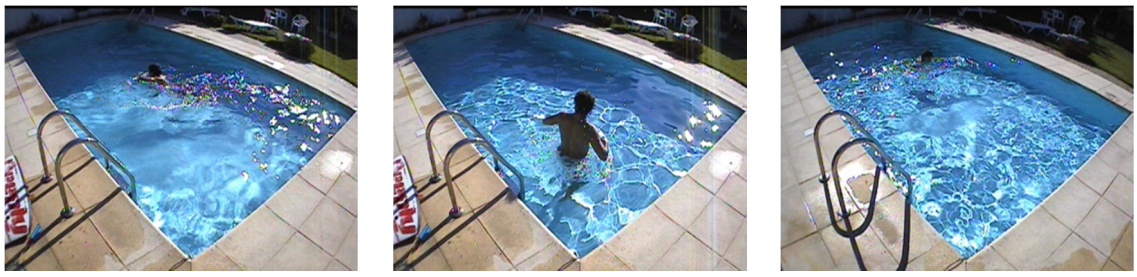


a) b) c) d)

Figura 2.2.7: A oscilação da superfície da água provoca a distorção dos corpos submersos. Nas 3 primeiras imagens os fotogramas são consecutivos mas podemos observar diferenças em todos eles relativamente ao corpo submerso e estático mostrado na última imagem. Apesar do corpo não se mover, na realidade, à superfície da água, parece existir movimento, devido à refracção e à oscilação da mesma.

a) Fotograma $\boldsymbol{n}$;
b) Fotograma $\boldsymbol{n+1}$;
c) Fotograma $\boldsymbol{n+2}$;
d) O corpo do nadador encontra-se submerso e estático no fundo da piscina durante os fotogramas $\boldsymbol{n}$ a $\boldsymbol{n+2}$.

A oscilação da água aliada à refracção faz com que objectos submersos sejam distorcidos de forma variável ao longo do tempo e que deste modo pareçam mexer-se, quando na realidade isso não está a acontecer. Num ambiente aquático desta natureza é também normalmente muito difícil visualizar a parte do corpo submersa de um nadador, devido não só à distorção causada pela água mas também a outros factores, tais como as bolhas de ar ou os salpicos e as reflexões especulares. A Figura 2.2.8 mostra exactamente esse aspecto muito importante, destacando a enorme dificuldade em detectar, mesmo para um ser humano, a parte do corpo submersa do nadador nas imagens.



a) b) c)

Figura 2.2.8: O conjunto de reflexões especulares e refracções aliado ao movimento oscilatório da água esconde as partes submersas dos corpos dos nadadores.

a) Consegue-se detectar de forma muito ténue a presença do corpo submerso. Mas as diferenças em relação à água são mínimas;
b) Mesmo com o indivíduo perto da câmara é muito difícil detectar a presença das pernas na água. Este nadador utiliza um vestuário vermelho e nem assim se consegue distinguir essa cor na água;
c) Caso extremamente difícil de segmentar, uma vez que apenas é possível observar a cabeça do nadador. Mesmo para um ser humano é impossível visualizar o resto do corpo abaixo da cabeça a partir desta imagem. No entanto um ser humano sabe que essa parte muito provavelmente está lá.

O pior dos casos acontece quando o corpo de um indivíduo se encontra submerso, tal como as imagens (a) e (b) da Figura 2.2.9 mostram e a oscilação à superfície da água é extremamente elevada. Deste modo é completamente impossível detectar o corpo submerso, imagem (b) da Figura 2.2.9, a não ser recorrendo a câmaras subaquáticas.

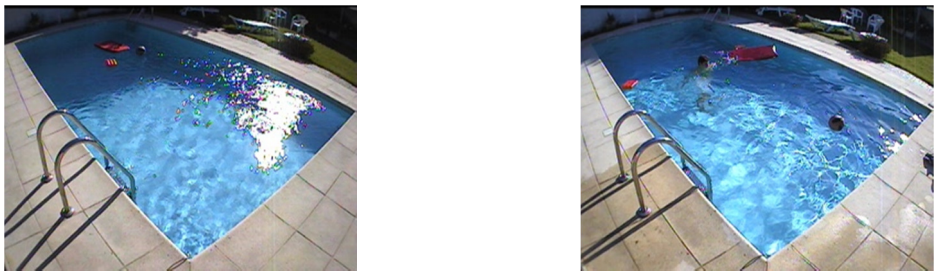
Por último, toda a oscilação da água à superfície provoca a movimentação dos objectos que nela flutuam. Quanto maior for a oscilação mais evidente será esta movimentação. A Figura 2.2.10 permite demonstrar a presença de objectos a flutuar na superfície da água da piscina.

a) b) c)

Figura 2.2.9: Quando a oscilação à superfície da água é elevada, um corpo submerso pode deixar de ser visível do exterior. No entanto de baixo de água vê-se perfeitamente e sem qualquer distorção.

a) Com pouca oscilação à superfície da água é possível visualizar o corpo submerso, ainda que este esteja bastante distorcido;
b) O corpo submerso deixa de ser detectável à superfície com a oscilação existente;
c) Corpo submerso nas imagens (a) e (b) visto por intermédio da câmara submersa na imagem (c).

Normalmente estes objectos não teriam movimento, mas desta forma, até o vento é capaz de os mover graças ao baixo atrito entre eles e a superfície da água, sendo mais uma vez classificados como *foreground* pelo algoritmo de segmentação de movimento. Em suma, os problemas aliados a um ambiente aquático são complexos e afectam negativamente a qualidade da segmentação. Além da cena ser exterior e deste modo afectada pelas condições atmosféricas e por conseguinte pela variação da luminosidade, a mesma apresenta características reflectoras e de movimento aleatório que causam uma série de ruídos indesejáveis. Entre eles estão as reflexões especulares variáveis no tempo no que respeita à sua posição e forma, as refracções provocadas pela água, distorcendo e por vezes escondendo os objectos submersos, as bolhas de ar e os salpicos entre outros acima descritos.



a) b)

Figura 2.2.10: Objectos a flutuar à superfície da água estão em movimento devido à oscilação da mesma. Quanto maior for a amplitude dessa oscilação mais evidenciado será o movimento dos objectos.

a) A oscilação da água é mínima, mas mesmo neste caso os objectos movem-se por causa do vento;
b) Caso em que a oscilação da água à superfície é superior à da imagem a), devido principalmente à presença de um indivíduo. Este indivíduo além disso ainda interage com os objectos fazendo com que os mesmos se desloquem na superfície da água de uma forma mais destacada.

2.3 Estado da arte na segmentação de movimento por subtracção de background

A segmentação de movimento através da técnica de subtracção de *background*, isto é, a subtracção da imagem actual ao *background* estimado, sujeita a um determinado *threshold*, é seguramente a mais utilizada em aplicações de vídeo vigilância, com câmara fixa, (McIvor, 2000). No entanto, os desafios desta técnica estão directamente relacionados com a modelação do *background*, ou seja, com a forma como se actualiza o mesmo ao longo do tempo de modo a comportar as suas variações para que este seja dinâmico. Além disso a resposta à pergunta, o que é o *background*, ou como se define o *background* é o ponto de partida para a concepção de um bom modelo. Entre as várias técnicas de modelação de *background* podem-se destacar as seguintes: *Running Gaussian Average*, *Temporal Median Filter*, *Mixture of Gaussians* (MoG), *Kernel Density Estimation* (KDE), *Sequential Kernel Density Approximation* (SKDA), *Cooccurence of image variations* e *Eigenbackgrounds*.

A primeira técnica mencionada, (Wren, Azarbayejani, Darrell, & Pentland, 1997), estima os parâmetros da distribuição gaussiana que melhor define cada pixel num conjunto de $\boldsymbol{n}$ imagens consecutivas, sendo os mesmos actualizados a uma taxa $\boldsymbol{\alpha}$, através da *running average* a cada novo fotograma $\boldsymbol{I}$, no instante de tempo $\boldsymbol{t}$, de acordo com a equação 2.3.1:

$$\mu_t = \alpha I_t + (1 - \alpha)\mu_{t-1} \tag{2.3.1}$$

A cada novo fotograma um pixel pode ser considerado como *foreground* se a desigualdade 2.3.2, que se segue, for verdadeira:

$$|I_t - \mu_t| > k\sigma_t \tag{2.3.2}$$

(Koller, et al., 1994) propõe a actualização selectiva do *background* baseada na equação 2.3.3:

$$\mu_t = M\mu_{t-1} + (1 - M)(\alpha I_t + (1 - \alpha)\mu_{t-1}) \tag{2.3.3}$$

Onde $\boldsymbol{M}$ é 1 caso o pixel corresponda ao *foreground* ou 0, caso contrário. Este método é de todos o que requer menores requisitos de processamento e memória, dada a sua simplicidade.

No entanto é também o método com pior precisão no processo de segmentação, sendo apenas adequado para cenas onde o *background* é pouco variável.

A metodologia de modelação do *background* baseada na média temporal dos últimos n fotogramas, *Temporal Median Filter*, é sugerida por (Lo & Velastin, 2001), sendo mais precisa que a descrita anteriormente. No entanto requer tanta mais memória e processamento quanto maior for o número de fotogramas utilizados na média temporal. Esta técnica apresenta ainda outras desvantagens no que respeita à falta de rigor estatístico e a uma medida de desvio dos valores em relação à média, fundamental para a adaptação do valor do *threshold* na subtracção.

Uma metodologia próxima da apresentada por (Lo & Velastin, 2001) consiste na modelação de cada pixel de *background* recorrendo a uma mistura de gaussianos, *Mixture of Gaussians* (MoG), uma vez que cada pixel ao longo do tempo pode tomar diferentes valores que se agrupam em classes distintas. Esta metodologia é mais flexível relativamente às variações que ocorrem no *background*. Os trabalhos de (Stauffer & Grimson, 1999) apresentam a utilização deste método não só para modelar o *background*, mas também o *foreground* correspondente a cada objecto encontrado. Assim sendo, a probabilidade do novo pixel, cujo valor é x, pertencer ao *background*, num determinado instante de tempo t, é dada pela equação 2.3.4:

$$P(x_t) = \sum_{i=1}^{K} \omega_{i,t} \eta \left(x_t - \mu_{i,t}, \sum_{i,t} \right) \tag{2.3.4}$$

Em geral, na prática o número de classes K está entre 3 e 5 e no caso em que as componentes de cor forem independentes, a matriz de co-variância $\sum_{i,t}$ é diagonal e reduz-se aos valores $\sigma_{i,t}^2 I$. Esta metodologia assume que as classes que correspondem ao *background* são as maiores e mais compactas distribuições, ou seja, aquelas que apresentam menor dispersão em torno do seu valor médio. Para estimar os parâmetros da mistura gaussiana e determinar qual a classe correspondente ao novo pixel x é utilizado o algoritmo *Expectation Maximization* (EM), desenvolvido por (Dempster, Laird, & Rubin, 1977), sobre os últimos n fotogramas. A modelação do *background* com MoG é ainda mais precisa que as técnicas descritas anteriormente e mais adaptativa no que diz respeito aos *backgrounds* variáveis, ou seja, permite que o *background* seja mais dinâmico. No entanto em termos de carga de processamento e memória utilizada é também mais exigente que os anteriores, aumentando essa exigência com o aumento do número de classes K que

compõem a mistura. Contudo, quando a frequência de variação do *background* é muito elevada, a modelação com MoG não é precisa. Ou seja, este modelo não comporta simultaneamente actualizações rápidas do *background* e uma detecção sensível do *foreground*. Deste modo, os autores (Elgammal, Hardwood, & Davis, 2000) propuseram uma modelação da distribuição do *background* por um modelo baseado no KDE (*Kernel Density Estimation*) dos últimos $\boldsymbol{n}$ fotogramas. Segundo os autores, quando existem variações rápidas no *background*, esta técnica permite estimar a densidade da distribuição a qualquer momento baseada apenas na história recente dos valores dos pixéis nessa localização. Assim sendo e considerando que uma amostra recente dos valores tomados por um pixel num dado ponto é dada por $\{x_1, x_2, \dots, x_n\}$, a função densidade de probabilidade que indica se o pixel de valor $\boldsymbol{x_t}$ pertence ou não ao *background*, utilizando o *kernel* estimador $\boldsymbol{k}$, é dada pela equação 2.3.5:

$$Pr(x_t) = \frac{1}{n}\sum_{i=1}^{n} K(x_t - x_i) \qquad \textbf{(2.3.5)}$$

Na realidade, quando o kernel $\boldsymbol{K}$ é uma distribuição normal $\eta(\,0, \sum\,)$, esta é uma generalização do MoG, com a particularidade de que cada amostra das $\boldsymbol{n}$ é considerada ela própria uma distribuição gaussiana. Esta abordagem permite um rápido esquecimento, se assim se pode dizer, sobre o passado do *background*, sendo dada uma maior importância às observações recentes dos valores do pixel. Este modelo apresenta maior sensibilidade que o MoG. Em termos de velocidade de execução e quantidade de memória requerida, este algoritmo é muito semelhante ao MoG. No entanto, o cálculo da probabilidade pode ser acelerado recorrendo a tabelas de procura, tal como os autores desta técnica referem nos seus trabalhos.

Outra técnica para modelação do *background* é baseada no algoritmo *mean-shift* e designa-se de *Sequential Kernel Density Approximation* (SKDA). A partir das amostras dos dados é possível determinar, de uma forma iterativa baseada na convergência, os valores mais frequentes da verdadeira distribuição de probabilidade. No entanto, a carga computacional é exageradamente elevada, não sendo apropriada para modelar o *background* ao nível do pixel (Comaniciu & Meer, 2002).

(Seki, Wada, Fujiwara, & Sumi, 2003) apresentaram uma técnica designada *Coocurrencce of image variations*, explicitamente direccionada para a modelação do *background*, baseada em blocos e tendo em conta a correlação espacial entre os pixéis. A ideia por detrás desta sugestão

vem do facto de que a vizinhança de um determinado pixel pertencente ao *background*, possivelmente, sofre variações similares ao longo do tempo. Esta afirmação não é, no entanto, de todo, verdadeira, uma vez que nas zonas de fronteira a variação não é similar e o algoritmo acaba por detectar *foreground* em zonas de *background*. A metodologia consiste de duas fases, uma delas de aprendizagem e outra de detecção, sendo efectuadas ambas em blocos de dimensões $N \times N$. Na primeira fase é efectuada a média temporal para cada bloco e obtidas as diferenças entre cada amostra e a média, chamadas de variações da imagem. Depois é gerada a matriz de co-variância relativamente à média e aplicada uma transformação de vector próprio (*eigenvector*) de modo a obter um conjunto de variações da imagem anterior. A segunda fase efectua para cada bloco a comparação com os seus blocos vizinhos no espaço próprio (*eigenspace*).

A abordagem sugerida por (Oliver, Rosario, & Pentland, 2000), designada por *Eigenbackgrounds*, é também ela baseada na decomposição dos valores próprios (*eigenvalues*), mas aplicada a toda a imagem ao invés de a aplicar a cada bloco, como na técnica anterior. Relativamente à modelação implementada por (Seki, Wada, Fujiwara, & Sumi, 2003), esta permite explorar extensivamente a correlação espacial entre os pixéis e evitar a formação de arrastos na imagem, causados pela partição da mesma em blocos. O algoritmo é constituído por duas fases. Na primeira fase, de aprendizagem, é calculada a média μ_b de n imagens consecutivas, com p pixéis e determinada a matriz de co-variância relativamente às diferenças entre cada uma delas e a média. A partir dessa matriz são seleccionados os melhores vectores próprios sendo armazenados numa matriz designada Φ_{Mb} de tamanho $M \times p$. Na segunda fase de classificação, para cada nova imagem disponível I é efectuada a projecção da mesma no espaço próprio de acordo com a equação 2.3.6:

$$I' = \Phi_{\mathrm{Mb}}(I - \mu_b) \tag{2.3.6}$$

Esta imagem representada no espaço próprio é de novo projectada no espaço da imagem através da equação 2.3.7 que se segue:

$$I'' = \Phi_{Mb}^{T} I' + \mu_b \tag{2.3.7}$$

Uma vez que o *eigenspace* é um bom modelo para as partes estáticas da imagem, a imagem I'' não contém os objectos em movimento. A subtracção entre I e I'' submetida a um determinado *threshold* permite evidenciar as zonas onde existe movimento na imagem.

Não tendo esta técnica melhores resultados quando comparada com o MoG, ela apresenta vantagens sobretudo na rapidez de execução (Piccardi, 2004). Os requisitos de memória e de processamento estão directamente relacionados com o número de amostras, n, utilizados no cálculo da média, bem como no número dos melhores vectores próprios.

Alguns autores, no entanto, principalmente aqueles que pretendem implementar a segmentação de movimento em ambientes aquáticos complexos, tal como o associado a uma piscina, conceberam novos algoritmos de segmentação adaptados especificamente para este tipo de ambiente. Estes novos algoritmos são na sua maioria adaptações e extensões das modelações de *background* baseadas em misturas gaussianas que proporcionam boa adaptação às mudanças do *background*. Seguidamente são abordados os algoritmos de segmentação de movimento específicos para ambientes aquáticos.

O sistema DEWS (Eng, Toh, Yau, & Wang, 2008) utiliza a técnica de subtracção de *background* para detectar os nadadores em movimento na piscina. Porém, o grande desafio desta técnica consiste na modelação do *background*, tarefa dificultada quando o mesmo é altamente variável, tal como o do ambiente associado a uma piscina. Nos trabalhos elaborados por (Eng, Wang, Kam, & Yau, 2004) os autores afirmam implementar uma detecção robusta e de elevada performance baseada na modelação efectiva do *background* aquático exterior e dinâmico comportando variações de luminosidade, brilhos e movimentações espaciais dos elementos que compõem o mesmo, aumentando a visibilidade dos nadadores que estão parcialmente escondidos devido às reflexões especulares. Para isso, conceberam um esquema de modelação do *background* constituído por um conjunto de regiões descritas por processos dinâmicos e homogéneos. Este modelo proporciona a construção de um esquema de procura que explora as dependências espaciais entre os pixéis. A concepção de um esquema de filtragem espácio-temporal permite aumentar a detecção dos nadadores que estão parcialmente escondidos devido às reflexões especulares causadas pela iluminação artificial no período nocturno. O método apresentado pelos autores difere das abordagens anteriores na medida em que efectua uma modelação baseada em regiões homogéneas dentro de cada bloco.

O algoritmo é composto por duas fases, a primeira de aprendizagem, que constrói o modelo do *background* inicial e a segunda de detecção. A fase inicial de geração do *background* começa por efectuar a média temporal da imagem correspondente a um determinado número de fotogramas, aplicando um modelo de cor da pele e uma média espacial de modo a eliminar os resíduos causados pelos nadadores. Seguidamente, a imagem do *background* inicial é convertida para o espaço de cores CIELa*b* que, segundo os autores, é a que apresenta melhores resultados. Depois, a imagem é dividida em blocos quadrados com as mesmas dimensões e aplicado o algoritmo de *clustering* hierárquico *k-means* a cada um deles de modo a encontrar a média e desvio padrão de cada região homogénea, em cada componente de cor, dentro do mesmo bloco. Os autores assumem que as componentes de cor do espaço de cores não se relacionam, motivo pelo qual as médias e desvios padrão de cada canal são determinados isoladamente. Cada região homogénea, em cada bloco, é assim modelada por uma única distribuição gaussiana multivariada, portanto, nas três componentes de cor do espaço CIELa*b*. Na prática, cada bloco é modelado por várias distribuições gaussianas multivariadas, cada uma correspondente a uma região homogénea, existindo assim uma mistura gaussiana em cada bloco. A cada nova imagem é calculada a probabilidade de cada novo pixel pertencer a cada região homogénea do *background* num dado bloco. Os autores consideram que, por utilizarem o espaço de cores CIELa*b*, as componentes do mesmo são independentes, reduzindo o cálculo da probabilidade ao produto das probabilidades de cada componente. A partir da distância euclidiana entre as distâncias de *mahalanobis* em cada componente de cor, relacionado com o *log-likelihood*, é determinada a similaridade de cada pixel do novo fotograma com o modelo de *background*, comparando depois esse valor de modo a determinar se o mesmo pertence ao *background* ou ao *foreground*. A actualização dos parâmetros das distribuições gaussianas é implementada através de um filtro de resposta infinita IIR com um determinado factor de aprendizagem. Cada região homogénea pertencente ao modelo de *background* é eliminada se não existirem pixéis que pertençam à mesma durante um certo número de fotogramas consecutivos. Caso os pixéis não pertençam a nenhuma distribuição, são criadas novas distribuições para modelar o *background*. Para aumentar a robustez do mapa binário de *foreground* os autores criaram uma técnica que utiliza *threshold* com histerese e agrupamento de pixéis conectados, calculando para isso a distância euclidiana de cada pixel nos blocos vizinhos com dois valores diferentes de *threshold*, um deles reduzido, com pouca precisão e sem ruído e outro elevado, com muita precisão e ruído. Destes dois mapas binários, os autores extraem os pixéis conectados no mapa binário com ruído na vizinhança dos pixéis de *foreground*

pertencentes ao mapa binário sem ruído. No final resulta um mapa binário com boa precisão, tal como se o *threshold* fosse elevado, mas sem ruído, tal como se o *threshold* fosse baixo. É ainda criado e mantido um modelo de *foreground* para os nadadores, baseando-se nos blocos vizinhos e depois no processo anteriormente descrito.

A metodologia concebida pelos autores, no que respeita ao uso de *threshold* com histerese, parece melhorar a precisão da segmentação de forma significativa. No entanto, a utilização do algoritmo *k-means* implica a definição, *a priori*, tal como no MoG, do número de classes a considerar na aglomeração das regiões homogéneas. Ainda assim, o maior problema associado a esta metodologia tem a ver com a actualização dos parâmetros do *background*. Na realidade, esta actualização é muito dependente das frequências de oscilação na superfície da água. Neste caso, a velocidade de aprendizagem do novo fundo é muito rápida, o que compromete a detecção de um nadador caso este permaneça parado, mesmo por pouco tempo. Num sistema deste tipo, tem que se evitar ao máximo os falsos negativos, uma vez que isso compromete a detecção dos nadadores, principalmente no caso de estes estarem parados, um caso especial a ter em atenção, porque pode estar directamente ligado a um afogamento.

Em (Tan & Lu, 2002) é apresentado um algoritmo capaz de modelar o *background* da piscina e dos nadadores explorando para isso a homogeneidade da cena e a capacidade de lidar com grandes variações nos pixéis de *background*. Este algoritmo tem a particularidade de acoplar a tarefa de segmentação à de seguimento, obtendo assim melhores resultados. A modelação do *background* é efectuada a partir de uma mistura gaussiana multivariada no espaço de cores HSV, sendo os parâmetros da mistura determinados com recurso ao algoritmo standard EM. Para determinar os parâmetros iniciais do modelo os autores utilizam o algoritmo de *clustering mean-shift* para identificar as classes dominantes da cena na área monitorizada da piscina. Para cada novo pixel é determinada a probabilidade de este pertencer a cada uma das classes, sendo depois esse valor comparado com um *threshold* predefinido. Caso o valor da probabilidade seja inferior ao *threshold* o pixel é considerado como não pertencente ao *background*. Além disso, todos os pixéis nesta situação, que partilhem alguma proximidade espacial, são agrupados utilizando para isso um algoritmo standard de etiquetagem de pixéis adjacentes (Horn B. , 1997). Se uma dada região destes pixéis tiver um tamanho suficiente então a mesma é considerada *foreground*.

Quando uma região é considerada *foreground* começa a ser construído um modelo de aparência de cor para a mesma, que possivelmente corresponderá a um nadador. À semelhança do

background, estes pixéis de *foreground* também são modelados por uma mistura gaussiana multivariada, com cerca de 3 a 5 componentes, no espaço de cores HSV. A segmentação dos nadadores é assim efectuada recorrendo ao seguimento destas regiões de *foreground* modeladas, partindo do princípio de que a localização e tamanho de um nadador variam de forma suave ao longo do tempo se a taxa de fotogramas por segundo for elevada. Deste modo, baseando-se nestas características das regiões no fotograma anterior, é definida uma janela de procura no fotograma actual de modo a encontrar os pixéis com maior probabilidade de pertencerem ao modelo de *foreground* do centro desta janela. Um pixel é assim considerado como pertencente a um nadador, se no fotograma actual a sua probabilidade calculada com a mistura gaussiana multivariada do modelo de *foreground* for superior à probabilidade calculada a partir da mistura gaussiana multivariada do modelo de *background*.

No entanto alguns dos pixéis não pertencem nem ao modelo de *background* nem ao de *foreground*. Estes podem pertencer a reflexões especulares ou brilhos, sendo assim aplicado um mecanismo de filtragem capaz de detectar estes acontecimentos e descartá-los do *foreground*. Neste caso os autores utilizam um processo heurístico para determinar quais são estes pixéis. Como a presença de um nadador faz, normalmente, decrescer a intensidade dos pixéis na sua localização, ao contrário de uma reflexão especular ou brilho, que a faz aumentar, são implementadas condicionantes no que respeita à intensidade máxima dos mesmos. Assim, um pixel só pertence ao *foreground* caso a sua intensidade seja inferior num certo nível à intensidade média da classe que representa a maior amostra do modelo de *background*.

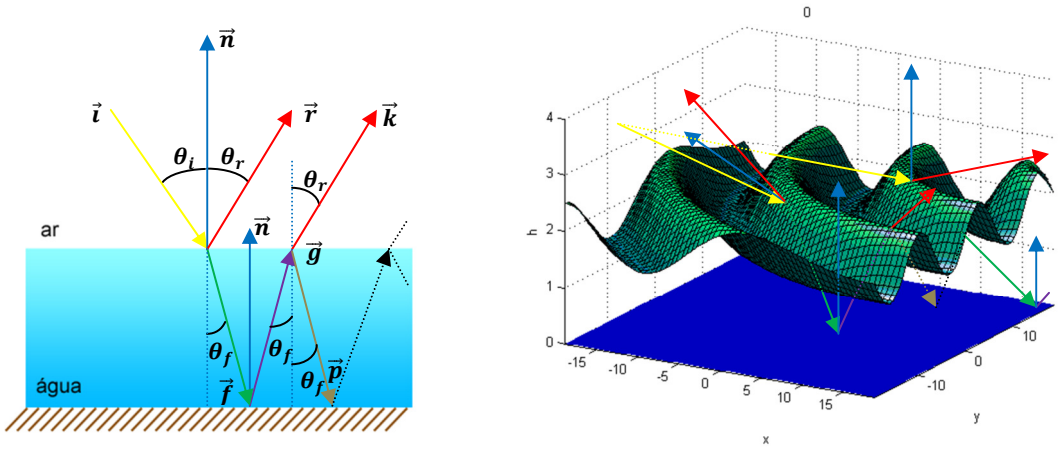
Os autores (Fei, Xueli, & Dongsheng, 2009) apresentam um algoritmo de segmentação de objectos em movimento, em imagens subaquáticas na piscina, capturadas com uma câmara monocular e fixa. A técnica baseia-se na subtracção de *background*, sendo o mesmo modelado por uma mistura gaussiana multivariada (MoG) cujos parâmetros são estimados via algoritmo EM standard. Estes autores eliminam ainda as sombras dos objectos baseados nos trabalhos efectuados por (Cucchiara, Grana, & Piccard, 2003) e (Cucchiara, Grana, & Piccardi, 2001). Os parâmetros da mistura gaussiana que descrevem o *background* são actualizados recorrendo a um filtro de resposta infinita IIR, tal como em (Stauffer & Grimson, 1999). O espaço de cores RGB é utilizado na modelação do *background*. Para remover as sombras dos objectos os autores convertem a imagem para o espaço de cores HSV. Neste espaço de cores, a componente tonalidade sofre algumas variações quando existe sombra, à semelhança da componente de saturação. Baseando-

se nessas pequenas diferenças, os autores definem uma máscara de sombra que filtra os pontos considerados sombra no *foreground*, detectados anteriormente, através da aplicação de limites às componentes intensidade, tonalidade e saturação. Se os pixéis satisfizerem as condições são considerados *foreground*, caso contrário, serão classificados como *background*.

No entanto a metodologia implementada por estes autores não contempla as reflexões internas causadas pela água, sendo por isso utilizado um limite horizontal na imagem para excluir as mesmas. Esse limite impõe um processamento apenas na parte inferior da imagem, tendo alguma vantagem na aceleração do processo de segmentação, mas uma desvantagem clara no aumento dos falsos negativos, fazendo com que os nadadores não sejam completamente detectados quando estão submersos. Além disso, a utilização de câmaras subaquáticas limita o próprio sistema no que respeita à detecção precoce do afogamento, uma vez que só é possível detectar um afogamento quando o corpo se afunda sem movimento, que não é o único caso possível. Outro problema advém do facto de os nadadores poderem ocultar as câmaras, uma vez que estas se encontram nas paredes da piscina e deste modo influenciar negativamente o funcionamento do sistema, não permitindo que este detecte os nadadores e um possível afogamento. Quanto à segmentação de movimento, pelas imagens fornecidas pelos autores, parece ter uma qualidade elevada, contemplando ainda a remoção das sombras. No entanto nada é referido relativamente às variações de luminosidade, até porque os testes são efectuados num ambiente interior, onde estas condições são bastante controladas.

2.4 Modelo de representação da superfície da água

Como se teve a oportunidade de verificar na primeira secção deste capítulo, que destaca os problemas da segmentação num ambiente aquático complexo, a superfície da água pode apresentar muita ou pouca oscilação, sendo esta oscilação causada pelos choques de pequenas ondas com diferentes características. Devido ao facto da densidade da água ser diferente da do ar, e estando esta em oscilação, parte da luz incidente é projectada no fundo da piscina e outra parte directamente reflectida, tal como se pode verificar pela análise das imagens (a) e (b) da Figura Figura 2.4.1. Ainda existe a reflexão interna total, que é originada quando a reflexão da luz do fundo da piscina para a superfície da água ultrapassa o ângulo crítico em relação à normal à superfície. No caso da água este ângulo em relação à normal à superfície é de 48.8º, o que significa que para ângulos superiores a luz não é refractada para o ar mas sim reflectida de novo para o interior da água e novamente para o fundo da piscina.



a) b)

Figura 2.4.1: Modelo de representação da superfície da água.

a) Reflexão especular perfeita causada pela superfície da água. Os ângulos entre o raio de luz incidente $\vec{i}$ e o reflectido $\vec{r}$ são iguais relativamente à normal à superfície da água $\vec{n}$. A passagem da luz de um meio para outro causa a refracção da mesma, de acordo com a lei de *Snell-Descartes*. O ângulo θ_f é inferior a θ_i. O raio de luz refractado $\vec{f}$ é reflectido pelo fundo da piscina dando origem a $\vec{g}$. A passagem da água para o ar causa nova refracção da luz $\vec{k}$ e nova reflexão, agora interna, de volta para o fundo da piscina $\vec{p}$;

b) Modelo representativo da superfície da água supondo que a sua ondulação corresponde a um seno perfeito. Na realidade a superfície da água corresponde a um somatório de diferentes ondas com diferentes amplitudes e frequências. O plano $z = 0$ pretende representar o fundo plano da piscina. Como a superfície da água não é plana devido à ondulação, a normal à superfície está continuamente a variar de direcção, fazendo com que a luz refractada e reflectida estejam também constantemente a variar a sua direcção. Os vectores deste esquema representam o mesmo que os vectores da imagem (a), sendo a sua correspondência verificada pela cor.

O ângulo crítico, no caso da água, é dado, em função da densidade n, pela equação 2.4.1, que se segue:

$$\theta_c = arcsen\left(\frac{n_{ar}}{n_{agua}}\right) \quad (2.4.1)$$

O modelo de representação da superfície da água proposto pelos físicos (Molesini & Vannoni, 2008) baseia-se em espelhos planos de área infinitesimal interligados entre si, Figura 2.4.2. Deste modo, propõe-se a aplicação deste modelo na representação da superfície da água, mas com a particularidade destas superfícies apresentarem apenas dois estados diferentes. Um estado reflector perfeito, à semelhança de um espelho, ou um estado de transparência perfeito, tal como um vidro com índice de refracção igual ao da água. No segundo estado tudo se passa como se se estivesse a observar directamente um ponto do fundo da piscina com os seus brilhos e sombras

projectados pela água à superfície. Este estado irá depender da sua posição angular relativamente ao eixo dos xx e dos yy, bem como da sua altura h, relativamente ao fundo da piscina.

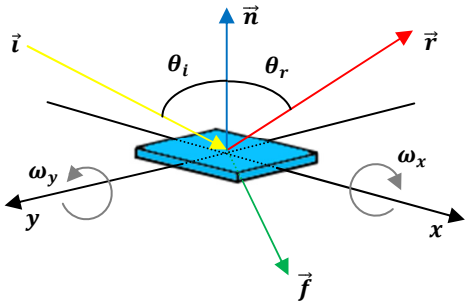


Figura 2.4.2: Área infinitesimal que representa uma ínfima parte da superfície da água da piscina. O conjunto destas superfícies interligadas entre si representa toda a superfície da água. Este modelo contempla as propriedades de reflexão e refracção da luz e além disso pode rodar em torno do eixo dos xx e dos yy com uma determinada velocidade angular ω_x e ω_y, respectivamente.

A sua posição angular θ_x e θ_y bem como a sua altura h são modeladas por equações periódicas, tendo uma determinada frequência também ela variável. Para certos valores de θ_x, θ_y e h, dependentes do posicionamento da câmara e da fonte de luz, a superfície infinitesimal pode variar o seu estado entre os dois possíveis, como se pode observar pela análise da Figura 2.4.3.

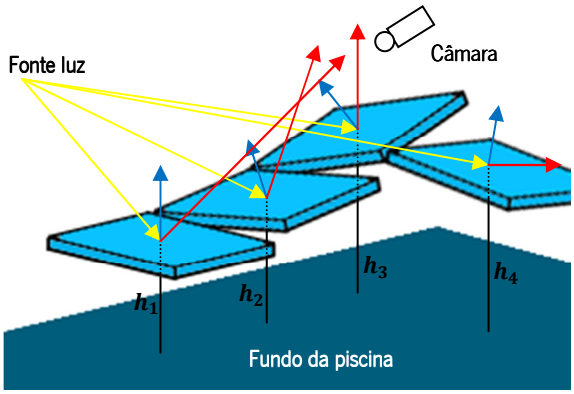


Figura 2.4.3: Superfície da água composta por áreas infinitesimais conjuntas, com diferentes alturas em relação ao fundo da piscina e diferentes ângulos de inclinação relativamente aos eixos x e y de cada uma delas. Com este modelo, a luz reflectida por cada uma destas superfícies pode ou não atingir a câmara, dependendo do posicionamento desta e da fonte de luz.

É ainda importante considerar o facto de a água apresentar uma tonalidade próxima da cor azul, mesmo que o fundo da piscina seja branco e a superfície da água não esteja a reflectir o céu.

Na realidade, estudos protagonizados por (Braun & Smirnov, 1993) demonstram que a água apresenta estes tons porque absorve a radiação visível próxima do vermelho, como se pode verificar pelo gráfico da Figura 2.4.4.

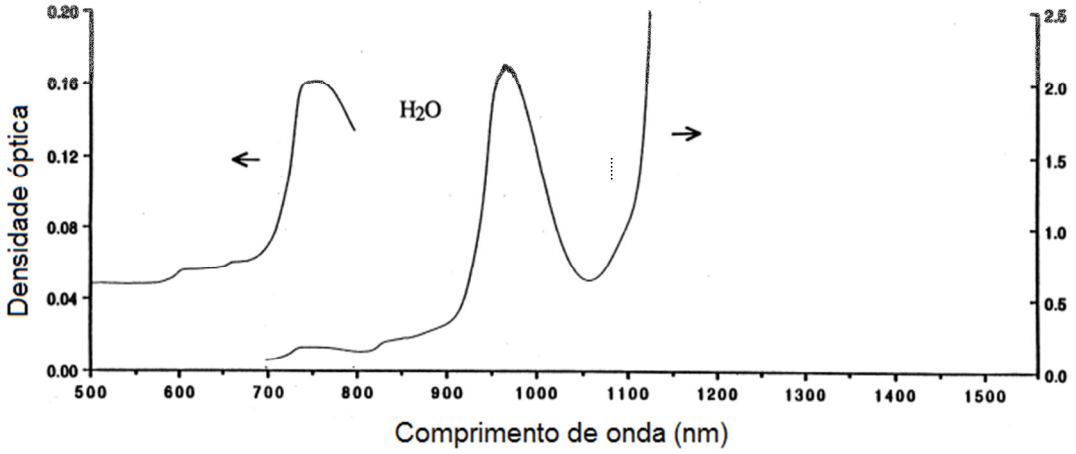


Figura 2.4.4: A absorção da luz com comprimentos de onda abaixo dos 700 nm provocada pela água pura contribui para a sua cor em tons próximos do azul. Adaptado de (Braun & Smirnov, 1993).

Quanto maior for a massa de água, isto é, a profundidade, mais intensa se torna esta absorção. Esta constatação implica que, quando a luz reflectida do fundo da piscina é refractada pela passagem da água para o ar, a mesma vem afectada pela absorção dos comprimentos de onda próximos do vermelho causada pela água. Portanto, neste modelo constituído por espelhos infinitesimais, o estado em que a superfície infinitesimal se comporta como transparente, tal como se observasse directamente o fundo da piscina, vem afectado por uma filtragem da componente vermelha da cor, daí os tons de azul serem predominantes na superfície da água. Explicado o modelo da superfície da água tendo em conta as suas características essenciais, define-se que cada pixel pertencente à área da mesma represente cada uma dessas superfícies infinitesimais e os seus dois estados possíveis. Assim é também possível obter o valor de um pixel com base no modelo de reflexão dicromático que será examinado posteriormente.

2.5 Os Modelos de representação do espaço de cores e o modelo de reflexão dicromático

A luz cromática apanha todo o intervalo do espectro electromagnético compreendido entre os comprimentos de onda de 400 e 700 nm, tal como a Figura 2.5.1 esclarece. Uma fonte de luz cromática é descrita por três valores designados por radiância (*radiance*), luminância (*luminance*) e

luminosidade (*brightness*), (Gonzalez & Woods, 2001). A radiância corresponde à quantidade total de energia que flui da fonte de luz e é normalmente medida em *watts* (W). A luminância, medida em *lumens* (lm) representa a quantidade de energia que um observador capta da fonte de luz, ou seja, a quantidade de energia realmente absorvida da energia total irradiada. Por último, tem-se a luminosidade, um valor subjectivo e impossível de medir, que representa a noção acromática da intensidade e que aparece como um factor relevante na descrição da sensação de cor. No ser humano, os cones, uma parte constituinte do olho, são os responsáveis pela visão a cores. Resultados experimentais evidenciaram que os 6 a 7 milhões de cones existentes podiam ser divididos em três categorias principais de absorção da cor, correspondendo estas ao vermelho, verde e azul, Figura 2.5.2 (b). Daí que se tenha adoptado que a representação de qualquer cor está na origem da combinação das intensidades de cada uma destas cores primárias R, G ou B, (Gonzalez & Woods, 2001), Figura 2.5.2 (a). Um espaço de representação de cor tem como objectivo quantificar formalmente a cor, sendo deste modo possível, matematicamente, medir e comparar cores.

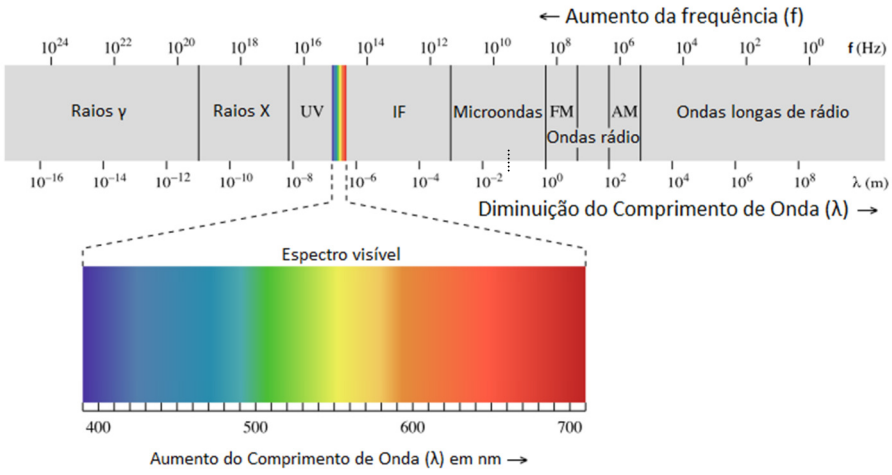
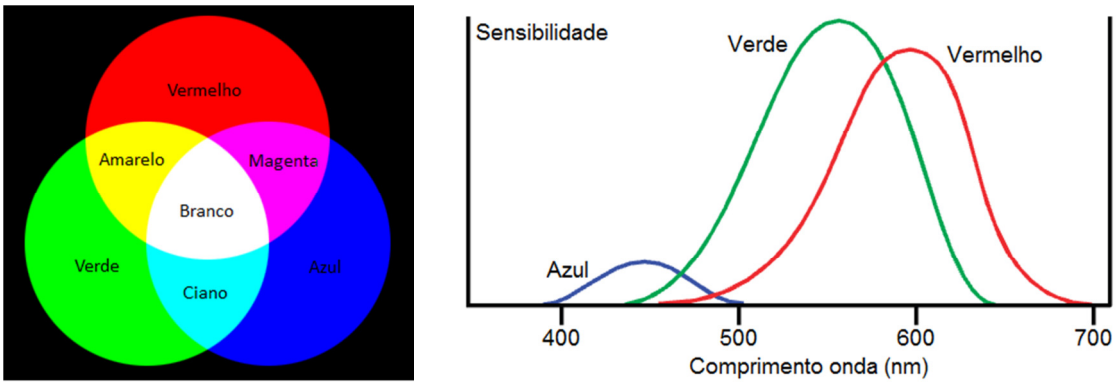


Figura 2.5.1: Cumprimentos de onda visíveis do espectro electromagnético. A luz visível apresenta um comprimento de onda que se encontra entre os 400 e os 700 nm.

Através do modelo de cor RGB baseado no funcionamento do sistema de visão humano é criado o espaço de cor RGB, onde cada cor é composta pela combinação das intensidades das três componentes primárias, vermelho, verde e azul. Este espaço é baseado num sistema de coordenadas cartesiano, tal como a Figura 2.5.3 mostra. Deste modo, a escala de cinzentos, ou intensidade luminosa, vai desde a origem do referencial, ponto (0,0,0) correspondente à cor preta,

até ao ponto de coordenadas máximo, (255,255,255), correspondente à cor branca. Ao longo dessa linha todas as componentes têm o mesmo valor. No cubo da Figura 2.5.3 todos os valores das cores foram normalizados para um máximo correspondente à unidade.



a) b)

Figura 2.5.2: Sistema de cores aditivo RGB.

a) Sistema de cores aditivo. As três cores primárias podem-se misturar dando origem a qualquer cor do espectro visível;
b) Os sensores de cor presentes no olho humano são de três tipos. O primeiro, designado por S, corresponde ao azul e capta comprimentos de onda mais baixos. Os sensores M e L correspondem ao verde e ao vermelho, respectivamente e captam comprimentos de onda mais elevados.

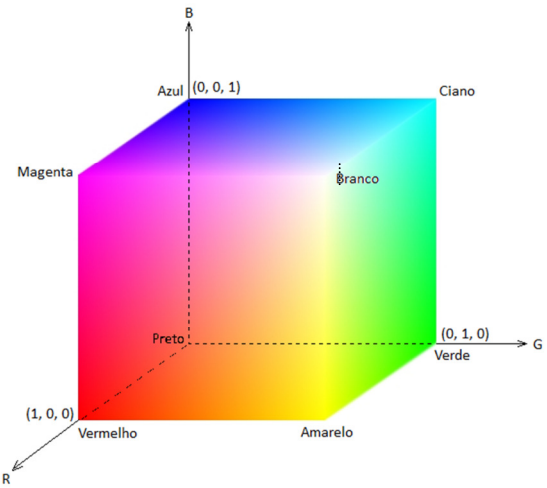


Figura 2.5.3: Espaço de cores RGB representado através de um cubo num referencial cartesiano onde cada eixo representa uma cor primária. O vértice (0,0,0) do cubo corresponde ao preto e o vértice (255,255,255) corresponde ao branco. Ao longo da diagonal que liga estes dois vértices o valor das três componentes é igual, representando a escala de cinzentos, ou intensidade luminosa.

Um *pixel* define-se assim como sendo o elemento mais pequeno de uma imagem, ou seja, um ponto, sendo composto pelas intensidades de cada um dos três canais RGB e fazendo assim parte do espaço cartesiano definido pelo cubo. Este modelo de representação de cor é normalmente utilizado na aquisição da maior parte dos sensores das câmaras de vídeo, sendo também disponibilizado à saída dos mesmos como o formato mais comum. Além disso a sua forma geométrica e a sua linearidade fazem com que a complexidade na sua utilização seja reduzida, evidenciando assim alguns benefícios.

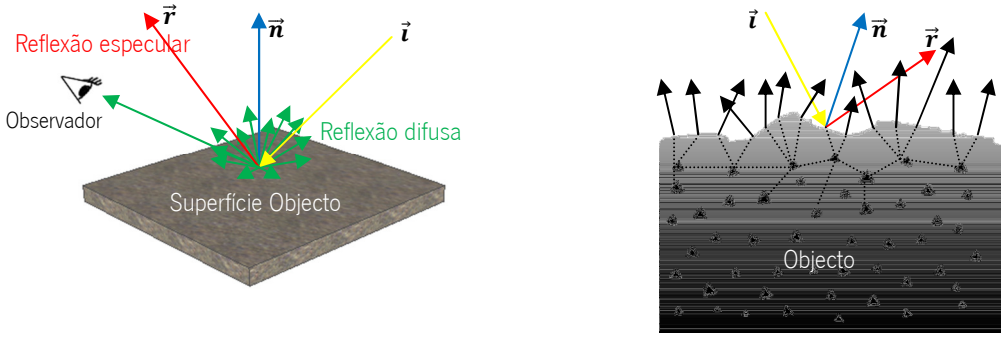
No entanto este modelo de representação de cor não é o mais indicado para utilização em operações de segmentação de imagem, uma vez que não se relaciona com a forma como um ser humano percepciona a cor relativamente a intensidade e tonalidade. Deste modo torna-se extremamente difícil comparar similaridades nas cores dos objectos, pois tanto a luminosidade das cores como a crominância estão misturadas nas três componentes, o que faz com que sombras e brilhos nos objectos da mesma cor tenham valores no cubo RGB completamente diferentes. Existe assim a necessidade de encontrar um espaço de representação de cor invariante no que respeita à posição do observador, à orientação do objecto, às variações na intensidade e direcção da fonte luminosa, aos brilhos e às sombras. De modo a encontrar um espaço de cor invariante que esteja o mais próximo possível destas características parte-se do modelo de reflexão dicromático introduzido por (Shafer, 1984) para representar o valor de um pixel no espaço de cores RGB.

O modelo de reflexão dicromático divide a luz que atinge um objecto em duas componentes, de acordo com as propriedades do objecto. Uma delas corresponde à reflexão especular ou brilho, causada pelo chamado interface do objecto e apresenta um espectro igual ao da fonte de luz que é irradiada sobre o objecto. Esta reflexão é semelhante à reflexão causada por um espelho, sendo os ângulos de reflexão e de incidência iguais, tal como se pode verificar através da análise da Figura 2.5.4 (a). A outra componente, normalmente apresenta um espectro diferente da fonte de luz, e está directamente relacionada com a cor do objecto em questão. Esta componente designa-se por reflexão difusa, pois difunde-se em todas direcções, sendo tanto mais difusa quanto mais rugoso for o objecto e é causada pelo chamado corpo do objecto, Figura 2.5.4 (b).

Supondo que uma superfície de área infinitesimal pertencente a um determinado objecto, numa cena, é atingida por uma fonte de luz uniforme e uma câmara com três sensores, um para cada componente R, G e B, com sensibilidades espectrais $f_R(\gamma)$, $f_G(\gamma)$ e $f_B(\gamma)$, respectivamente,

capta a reflexão difusa e especular dessa superfície, então, o valor desse pixel é dado pela equação 2.5.1:

$$C\{R,G,B\} = m_b(\vec{n},\vec{s})\int_\gamma f_C(\gamma)\, e(\gamma) c_b(\gamma) d\gamma + m_s(\vec{n},\vec{s},\vec{v})\int_\gamma f_C(\gamma)\, e(\gamma) c_s(\gamma) d\gamma \quad (2.5.1)$$



a) b)

Figura 2.5.4: Modelo de reflexão dicromático.

a) **Reflexões especular e difusa num objecto fosco e rugoso a um nível macroscópico. Neste caso, devido ao seu posicionamento o observador é atingido apenas pela reflexão difusa;**
b) **A uma escala microscópica o mesmo objecto apresenta uma superfície não plana. A reflexão especular $\vec{r}$ tem o mesmo ângulo que o raio de luz incidente $\vec{i}$ em relação à normal à superfície $\vec{n}$. Esta reflexão é causada pelo interface entre o ar e o objecto e a ela está inerente o fenómeno dos brilhos que aparecem nos objectos. A reflexão difusa é causada pelo corpo do objecto e dependendo das características deste absorve determinadas frequências da luz irradiada. A reflexão difusa é a que confere a cor ao objecto.**

A luz que é irradiada sobre o objecto é dada por $e(\gamma)$, sendo que, os termos m_b e m_s representam as magnitudes de reflexão ou dependências geométricas relacionadas com o corpo e o interface do objecto, respectivamente, e são independentes do comprimento de onda da luz reflectida. O termo m_b depende da normal à superfície $\vec{n}$ e da direcção da fonte do raio de luz incidente $\vec{s}$. O termo m_s depende não só da normal à superfície do objecto e da direcção da fonte de luz, mas também da direcção do observador $\vec{v}$. Por outro lado, $c_b(\gamma)$ e $c_s(\gamma)$ representam as cores da reflexão provocada pelo corpo e pelo interface do objecto, respectivamente. Estas dependem efectivamente do comprimento de onda da luz irradiada sobre o objecto, mas não dependem da sua geometria, nem da posição da fonte de luz ou do observador. Num dado ponto, e partindo do princípio de que tanto a direcção da fonte de luz bem como do observador não variam, as magnitudes de reflexão m_be m_s são constantes.

Em (Gevers & Smeulders, 1997) os autores baseiam-se no modelo de reflexão neutra (NIR) e numa iluminação branca contendo todo o espectro visível para considerar a cor da reflexão provocada pelo interface do objecto constante, ou seja, independente do comprimento de onda γ. Assim sendo os autores consideram que $e(\gamma) = e$ e $c_s(\gamma) = c_s$ e que $k_c = \int_\gamma f_c(\gamma)\, c_b(\gamma) d\gamma$, dependente da sensibilidade dos sensores e do corpo do objecto, a equação 2.5.1 vem assim simplificada e dada por:

$$C_\omega\{R, G, B\} = C_b + C_s = em_b(\vec{n}, \vec{s})k_c + em_s(\vec{n}, \vec{s}, \vec{v})c_s f$$

(2.5.2)

Os autores consideram ainda que, $\int_\gamma f_R(\gamma)\, d\gamma = \int_\gamma f_G(\gamma)\, d\gamma = \int_\gamma f_B(\gamma)\, d\gamma = f$, pois a luz é branca e as sensibilidades espectrais dos sensores são consideradas iguais.

Partindo do modelo de reflexão dicromático e das assunções feitas por (Gevers & Smeulders, 1997), em (Luckav & Plataniotis, 2007) são analisados alguns modelos de representação de cor relativamente à sua invariabilidade no que respeita à localização do observador, à geometria do objecto, à intensidade e cor da fonte de iluminação e aos brilhos. O modelo de representação de cor *rgb*, um espaço de cores RGB normalizado, é obtido a partir das equações que se seguem:

$$r = \frac{R}{R + G + B}$$

(2.5.3)

$$g = \frac{G}{R + G + B}$$

(2.5.4)

$$b = \frac{B}{R + G + B}$$

(2.5.5)

Utilizando o termo correspondente à reflexão difusa do modelo de reflexão dicromático, $C_b = em_b(\vec{n}, \vec{s})k_c$, que considera a fonte de luz e a magnitude da reflexão relativa ao corpo do objecto constantes, apenas a cor da reflexão difusa, relacionada com as propriedades do próprio objecto, e a sensibilidade do sensor têm influência em C_b. Deste modo, substituindo o referido termo nas equações que determinam os valores de *r*, *g* e *b* tem-se que:

$$r(R_b, G_b, B_b) = \frac{em_b(\vec{n}, \vec{s})k_R}{em_b(\vec{n}, \vec{s})(k_R + k_G + k_B)} = \frac{k_R}{k_R + k_G + k_B}$$

(2.5.6)

$$g(R_b, G_b, B_b) = \frac{em_b(\vec{n}, \vec{s})k_G}{em_b(\vec{n}, \vec{s})(k_R + k_G + k_B)} = \frac{k_G}{k_R + k_G + k_B}$$

(2.5.7)

$$b(R_b, G_b, B_b) = \frac{em_b(\vec{n}, \vec{s})k_B}{em_b(\vec{n}, \vec{s})(k_R + k_G + k_B)} = \frac{k_B}{k_R + k_G + k_B}$$

(2.5.8)

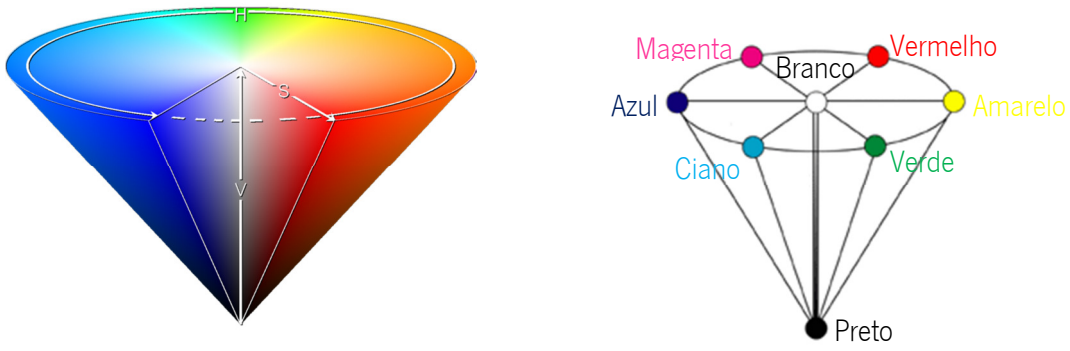
O desaparecimento da parte correspondente à fonte de iluminação e à magnitude de reflexão do corpo do objecto na operação de normalização indica que este espaço de cores é independente da orientação da superfície, na direcção e intensidade da fonte de luz. No entanto é dependente da cor do próprio objecto das características de sensibilidade dos sensores e da reflexão especular causada pelo interface do mesmo.

O espaço de representação de cor HSV, ao contrário do RGB, está directamente relacionado com a forma como um ser humano percepciona a cor, ignorando brilhos e sombras, sendo capaz de relacionar a cor com o comprimento de onda na gama do visível. Este sistema é composto por 3 componentes, a tonalidade H (*Hue*), a saturação S (*Saturation*) e a intensidade V (*Value*). A tonalidade ou matiz corresponde a um valor angular compreendido entre 0° e 360° que representa a cor, tal como se pode verificar através das imagens (a) e (b) da Figura 2.5.5. Por exemplo, a cor vermelha corresponde a um ângulo de 0°. A componente saturação corresponde à quantidade de cor que está presente, podendo ser também definida como a pureza da cor. Por último, a intensidade determina a luminosidade ou o brilho associado à cor.

No espaço de cores HSV, o valor da tonalidade é obtido a partir da seguinte equação:

$$H(R, G, B) = \arctan\left(\frac{\sqrt{3}(G - B)}{((R - G) + (R - B))}\right)$$

(2.5.9)



a) b)

Figura 2.5.5: Espaço de cores HSV.

a) Sistema de representação de cor HSV. O eixo vertical do cone corresponde à intensidade ou brilho (V), sendo mínima no vértice e máxima na base. O raio desde o centro até à periferia do cone mede a saturação da cor (S), conferindo uma pureza à cor tanto maior quanto maior for este. A tonalidade da cor ou matiz (H) é medida em graus sendo determinada pelo ângulo descrito pelo vector que representa a saturação. Adaptado de (Mundhenk, 2008).

b) Esquema de representação do espaço de cores HSV onde se podem observar as localizações das cores de acordo com o ângulo do vector saturação. À medida que se caminha desde o vértice do cone até à sua base o brilho aumenta. Adaptado de (Adrienne & John, 2010).

Aplicando o modelo de reflexão dicromático no que respeita à reflexão difusa na equação 2.5.9, tem-se que:

$$H(R_b, G_b, B_b) = \arctan\left(\frac{\sqrt{3}em_b(\vec{n},\vec{s})(k_G - k_B)}{em_b(\vec{n},\vec{s})\left((k_R - k_G) + (k_R - k_B)\right)}\right) = \arctan\left(\frac{\sqrt{3}(k_G - k_B)}{(k_R - k_G) + (k_R - k_B)}\right) \quad (2.5.10)$$

Através da equação 2.5.10 nota-se que a componente tonalidade do espaço de cores HSV é invariante relativamente à orientação da superfície, direcção e intensidade da fonte de luz. Além disso, e considerando que a cor dos brilhos é independente da cor do objecto e igual à cor da fonte de luz, no caso em que esta é branca, a reflexão especular também o é, significando que a reflexão causada pelo interface do objecto encontra-se sempre na diagonal do cubo que representa a escala de cinzentos ou intensidade. De acordo com (Shafer, 1984), a reflexão total, especular e difusa, causada por um objecto é baseada no modelo de reflexão dicromático, sendo que, deste modo, o valor de $C\{R, G, B\}$ deve estar contido no paralelogramo definido pelas componentes C_b e C_s no interior do cubo RGB. Deste modo, a componente tonalidade é também invariante a brilhos, tal como se pode verificar pela equação 2.5.11:

$$H(R_\omega, G_\omega, B_\omega) = \arctan\left(\frac{\sqrt{3}(G_\omega - B_\omega)}{(R_\omega - G_\omega) + (R_\omega - B_\omega)}\right) = \arctan\left(\frac{\sqrt{3}em_b(\vec{n},\vec{s})(k_G - k_B)}{em_b(\vec{n},\vec{s})\big((k_R - k_G) + (k_R - k_B)\big)}\right)$$
$$= arctan\left(\frac{\sqrt{3}(k_G - k_B)}{(k_R - k_G) + (k_R - k_B)}\right)$$

(2.5.11)

Relativamente à componente saturação, esta é definida pela seguinte equação:

$$S(R, G, B) = 1 - \frac{\min(R, G, B)}{R + G + B}$$

(2.5.12)

De acordo com o modelo de reflexão dicromático esta componente apenas depende dos sensores, do corpo do objecto e dos brilhos, obtendo-se então:

$$S(R_b, G_b, B_b) = 1 - \frac{\min(em_b(\vec{n},\vec{s})k_R, em_b(\vec{n},\vec{s})k_G, em_b(\vec{n},\vec{s})k_B)}{em_b(\vec{n},\vec{s})(k_R + k_G + k_B)} = 1 - \frac{\min(k_R, k_G, k_B)}{k_R + k_G + k_B}$$

(2.5.13)

O valor da intensidade no espaço de cores HSV é determinado pelo maior valor das componentes R, G e B. Este é um caso claro que depende da intensidade e direcção da fonte de luz, do corpo do objecto, dos sensores e dos brilhos causados pela reflexão especular, tal como de pode verificar pela aplicação do modelo de reflexão dicromático:

$$V(R_b, G_b, B_b) = \max(em_b(\vec{n},\vec{s})k_R, em_b(\vec{n},\vec{s})k_G, em_b(\vec{n},\vec{s})k_B)$$

(2.5.14)

Os espaços de representação de cor conhecidos como $c_1c_2c_3$, $l_1l_2l_3$ e $m_1m_2m_3$ sugeridos por (Gevers & Smeulders, 1997) também invariantes, podem ser verificados com o modelo de reflexão dicromático de modo a avaliar a sua invariabilidade.

O espaço de cores $c_1c_2c_3$ é determinado pelas seguintes equações:

$$c_1 = arctan\left(\frac{R}{max\{G, B\}}\right)$$

(2.5.15)

$$c_2 = arctan\left(\frac{G}{max\{R,B\}}\right)$$

(2.5.16)

$$c_3 = arctan\left(\frac{B}{max\{R,G\}}\right)$$

(2.5.17)

Utilizando o termo correspondente à reflexão difusa, pode-se destacar que este espaço de representação de cor também é invariante no que respeita à geometria do objecto e à direcção e intensidade da fonte de luz que atinge o mesmo.

$$c_1(R_b,G_b,B_b) = arctan\left(\frac{em_b(\vec{n},\vec{s})k_R}{em_b(\vec{n},\vec{s})max\{k_G,k_B\}}\right) = arctan\left(\frac{k_R}{max\{k_G,k_B\}}\right)$$

(2.5.18)

$$c_2(R_b,G_b,B_b) = arctan\left(\frac{em_b(\vec{n},\vec{s})k_G}{em_b(\vec{n},\vec{s})max\{k_R,k_G\}}\right) = arctan\left(\frac{k_R}{max\{k_R,k_G\}}\right)$$

(2.5.19)

$$c_3(R_b,G_b,B_b) = arctan\left(\frac{em_b(\vec{n},\vec{s})k_B}{em_b(\vec{n},\vec{s})max\{k_R,k_G\}}\right) = arctan\left(\frac{k_R}{max\{k_R,k_G\}}\right)$$

(2.5.20)

O espaço de cores $I_1I_2I_3$ é obtido através das seguintes equações:

$$l_1 = \frac{(R-G)^2}{(R-G)^2+(R-B)^2+(G-B)^2}$$

(2.5.21)

$$\boldsymbol{l_2 = \frac{(R-B)^2}{(R-G)^2+(R-B)^2+(G-B)^2}}$$

(2.5.22)

$$\boldsymbol{l_3 = \frac{(G-B)^2}{(R-G)^2+(R-B)^2+(G-B)^2}}$$

(2.5.23)

A aplicação do modelo de reflexão dicromático às equações anteriores destaca que, à semelhança do espaço de cores $c_1c_2c_3$, este modelo de representação de cor é também invariante relativamente à intensidade e direcção da fonte de luz e à geometria do objecto.

$$l_1 = \frac{\left(em_b(\vec{n},\vec{s})\right)^2(k_R - k_G)^2}{\left(em_b(\vec{n},\vec{s})\right)^2((k_R - k_G)^2 + (k_R - k_B)^2 + (k_G - k_B)^2)} = \frac{(k_R - k_G)^2}{(k_R - k_G)^2 + (k_R - k_B)^2 + (k_G - k_B)^2}$$

(3.4.24)

$$l_2 = \frac{\left(em_b(\vec{n},\vec{s})\right)^2(k_R - k_B)^2}{\left(em_b(\vec{n},\vec{s})\right)^2((k_R - k_G)^2 + (k_R - k_B)^2 + (k_G - k_B)^2)} = \frac{(k_R - k_B)^2}{(k_R - k_G)^2 + (k_R - k_B)^2 + (k_G - k_B)^2}$$

(3.4.25)

$$l_3 = \frac{\left(em_b(\vec{n},\vec{s})\right)^2(k_G - k_B)^2}{\left(em_b(\vec{n},\vec{s})\right)^2((k_R - k_G)^2 + (k_R - k_B)^2 + (k_G - k_B)^2)} = \frac{(k_G - k_B)^2}{(k_R - k_G)^2 + (k_R - k_B)^2 + (k_G - k_B)^2}$$

(3.4.26)

Analisando a Tabela 2.5.1 é possível verificar que o espaço de representação de cor HSV é invariante, relativamente à sua componente tonalidade, à direcção e intensidade da fonte de luz, à geometria do objecto e também a brilhos. Apesar de ser dependente da cor da fonte de luz que ilumina o objecto, esta componente devido às suas características de invariabilidade é a que melhores condições apresenta para ser utilizada no processo de segmentação.

Sistema de representação de cores		Posição do observador	Geometria	Cor da fonte de iluminação	Intensidade da fonte de iluminação	Brilhos
RGB		-	-	-	-	-
rgb		x	x	-	x	
HSV	H	x	x	-	x	x
	S	x	x	-	x	-
	V	-	-	-	-	-
$l_1l_2l_3$		x	x	-	x	-
$c_1c_2c_3$		x	x	-	x	-

Tabela 2.5.1: Resumo das características de invariabilidade dos diferentes espaços de representação de cor. O "x" significa que o espaço de cor é invariante nessa característica.

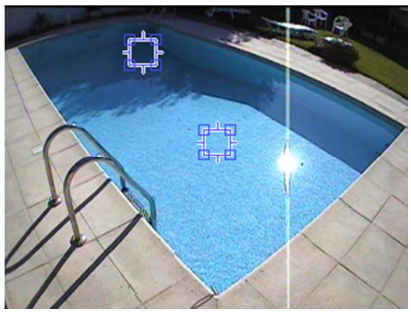
Além disso a dependência da cor da fonte de luz não é relevante uma vez que a iluminação da cena é proporcionada pela luz do sol, e esta pode ser considerada branca, uma vez que contém todas as frequências do espectro visível. Ao longo do dia é verdade que a cor da fonte de luz sofre variações de acordo com a posição do sol relativamente à cena, mas essa variação não é significativa.

No entanto, toda a formulação do modelo de reflexão dicromático assenta em objectos sólidos com superfície rugosa, daí a reflexão total do objecto corresponder à soma das reflexões especular e difusa. Nos trabalhos de (Shafer, 1984) não é feita nenhuma alusão a outro tipo de materiais, como a água, de tal modo que é necessário verificar se é possível modelar a superfície da água de acordo com o modelo de reflexão dicromático. Tal como foi mencionado anteriormente, a água pode ser modelada por um conjunto de superfícies infinitesimais interligadas entre si. Foi dito ainda que estas superfícies poderiam estar em dois estados distintos, ora reflectindo como um espelho perfeito, ora comportando-se como uma superfície transparente deixando passar os raios de luz reflectidos pelo fundo da piscina. Assim, quando a superfície da água é um reflector perfeito, a mesma reflecte os raios de luz reflectidos por outros objectos, que são modelados pelo modelo de reflexão dicromático. Mesmo quando a fonte de luz é reflectida directamente, aquilo a que se pode designar de brilho com elevada intensidade, a componente tonalidade do espaço de cores HSV é independente. No caso em que a superfície infinitesimal se comporta como uma superfície transparente, a mesma reflecte o fundo da piscina ou o objecto mergulhado na mesma, que também é modelado de acordo com o modelo de reflexão dicromático. É portanto seguro admitir que a superfície da água também pode ser modelada de acordo com o modelo de reflexão dicromático, sendo o espaço de cor HSV, devido à sua componente tonalidade, o mais adequado para efectuar a segmentação de objectos na superfície da água. É no entanto de notar que existe uma absorção das gamas de comprimento de onda maior no espectro visível, conferindo à superfície da água um tom azulado, principalmente quando esta se comporta como uma superfície transparente.

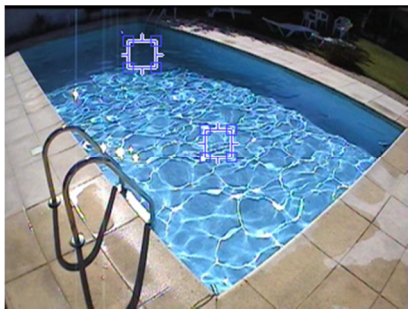
2.6 Análise experimental dos vários modelos de representação de cor

Foi efectuado um teste em duas amostras de vídeo correspondentes a uma cena com uma piscina, iluminadas nas mesmas condições, de modo a verificar a invariabilidade a brilhos, sombras e variação da intensidade luminosa sobre a cena, de diferentes modelos de representação de cor. As imagens foram recolhidas no formato RGB e têm uma dimensão de 320 por 240 pixéis. A captura foi efectuada a 25 fotogramas por segundo, durante um período de 20 segundos, o que perfaz um total de 500 fotogramas.

Observando a Figura 2.6.1 nota-se que numa das amostras a superfície da água quase não apresenta oscilação, figura (a), e que na outra a oscilação é muito elevada, figura (b). O nível de oscilação da superfície da água tem uma grande influência no aparecimento de brilhos, por isso, espaços de cor invariantes relativamente a brilhos irão atenuar este tipo de ocorrência na imagem. Por outro lado, em cada amostra, foram recolhidos os valores ao longo do tempo da média espacial da imagem completa e de duas localizações de dimensões de 20 por 20 pixéis, uma delas numa zona com reduzida luminosidade e outra numa zona com elevada luminosidade, pertencentes à superfície da água.



a)



b)

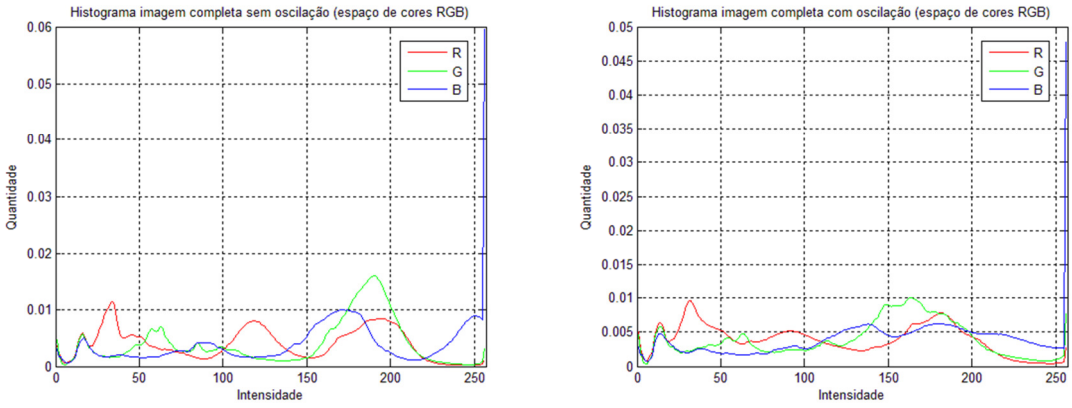
Figura 2.6.1: Amostras recolhidas nas localizações marcadas nas imagens. Uma das localizações corresponde a uma zona de baixa luminosidade e a outra a uma zona de elevada luminosidade. Ambas as imagens têm uma dimensão de 320x240 pixéis, tendo as amostras uma dimensão de 20x20 pixéis. O tempo de amostragem foi de 20 segundos correspondendo a um conjunto de 500 fotogramas sequenciais. Numa e noutra situação as amostras ocupam as mesmas posições. A quantidade de luz que é irradiada sobre a cena é muito semelhante, uma vez as amostras foram recolhidas em períodos de tempo próximos.

a) **Oscilação reduzida da superfície da água;**
b) **Oscilação elevada da superfície da água.**

Assim, para cada espaço de representação de cor foram medidas ao longo do tempo as diferentes componentes de modo a apurar quais as que apresentam menor variação nas diferentes condições de oscilação e luminosidade.

2.6.1 Espaço de representação de cor RGB

No espaço de representação de cor RGB, tal como se pode apurar pela análise dos histogramas presentes na Figura 2.6.2, a oscilação mais elevada da superfície da água, no histograma (b), fez com que os níveis de luminosidade nos três canais sofressem uma maior dispersão, ou seja, deixaram de existir picos tão destacados nas diferentes componentes, tal como existiam no histograma (a). No entanto, é nas amostras relativas apenas à superfície da água que se nota uma maior variação nas componentes de cor. Para isso atente-se nos histogramas (a) e (b) da Figura 2.6.3. Ambos pertencem à amostra de reduzida oscilação da água, mas em diferentes condições de luminosidade. Analisando os referidos histogramas nota-se um desvio nas médias de todas as componentes de cor, sinónimo do aumento da intensidade luminosa sobre a cena, nas duas situações. Comparando com os histogramas (c) e (d), o aumento da oscilação da superfície da água, por fazer aparecer mais brilhos, causou uma maior dispersão em torno das médias das diferentes componentes de cor.



a) b)

Figura 2.6.2: Histogramas das três componentes de cor RGB correspondentes à média temporal da média espacial, aplicada à imagem completa, do conjunto dos 500 fotogramas sequenciais, em duas situações distintas de oscilação da superfície da água.

a) Oscilação reduzida da superfície da água;
b) Oscilação elevada da superfície da água.

Facilmente se conclui que este espaço de representação de cor é altamente dependente da variação da intensidade luminosa sobre a cena, sendo por isso muito afectado pelas diferentes condições de oscilação da superfície da água e da intensidade luminosa irradiada sobre a cena.

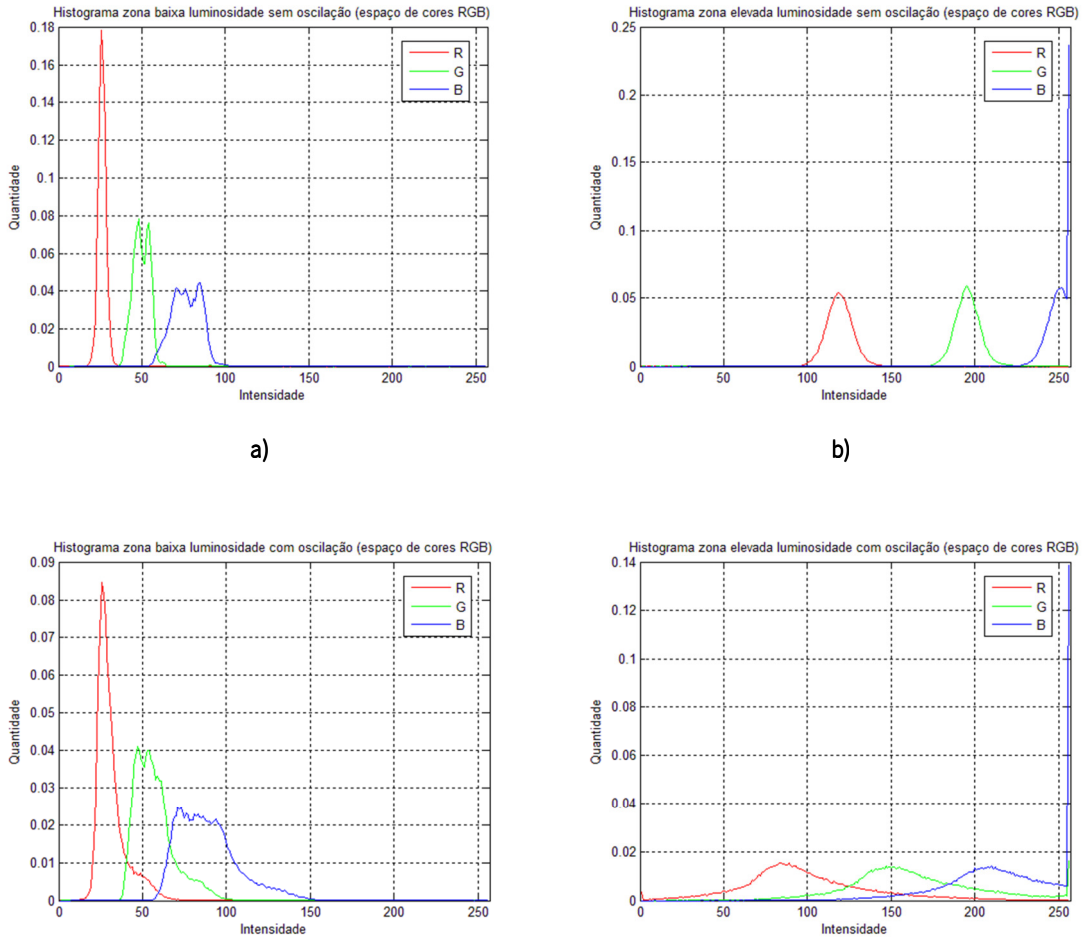


Figura 2.6.3: Histogramas das três componentes de cor RGB correspondentes à média temporal da média espacial, aplicada às duas localizações de diferente luminosidade, do conjunto dos 500 fotogramas sequenciais, em duas situações distintas de oscilação da superfície da água.

a) Oscilação reduzida e baixa luminosidade;
b) Oscilação reduzida e luminosidade elevada;
c) Oscilação elevada e baixa luminosidade;
d) Oscilação elevada e luminosidade elevada.

2.6.2 Espaço de representação de cor *rgb*

Relativamente ao espaço de cores RGB normalizado, conhecido como *rgb*, nota-se que os diferentes níveis de oscilação da superfície da água têm uma influência menor nas componentes de cor relativamente ao espaço de cor RGB. Os histogramas (a) e (b) da Figura 2.6.5 são mais semelhantes que os seus homólogos pertencentes à Figura 2.6.3. Analisando de forma mais pormenorizada as localizações pertencentes à superfície da água, Figura 2.6.4, nota-se que, a variação da intensidade luminosa ainda continua a ter influência nas componentes do espaço de cor, existindo um pequeno deslocamento das médias, ainda assim muito inferior quando

comparado com as componentes do espaço de cor RGB. Com uma oscilação mais elevada da superfície da água é possível verificar que as componentes de cor sofreram um aumento da dispersão em torno da média.

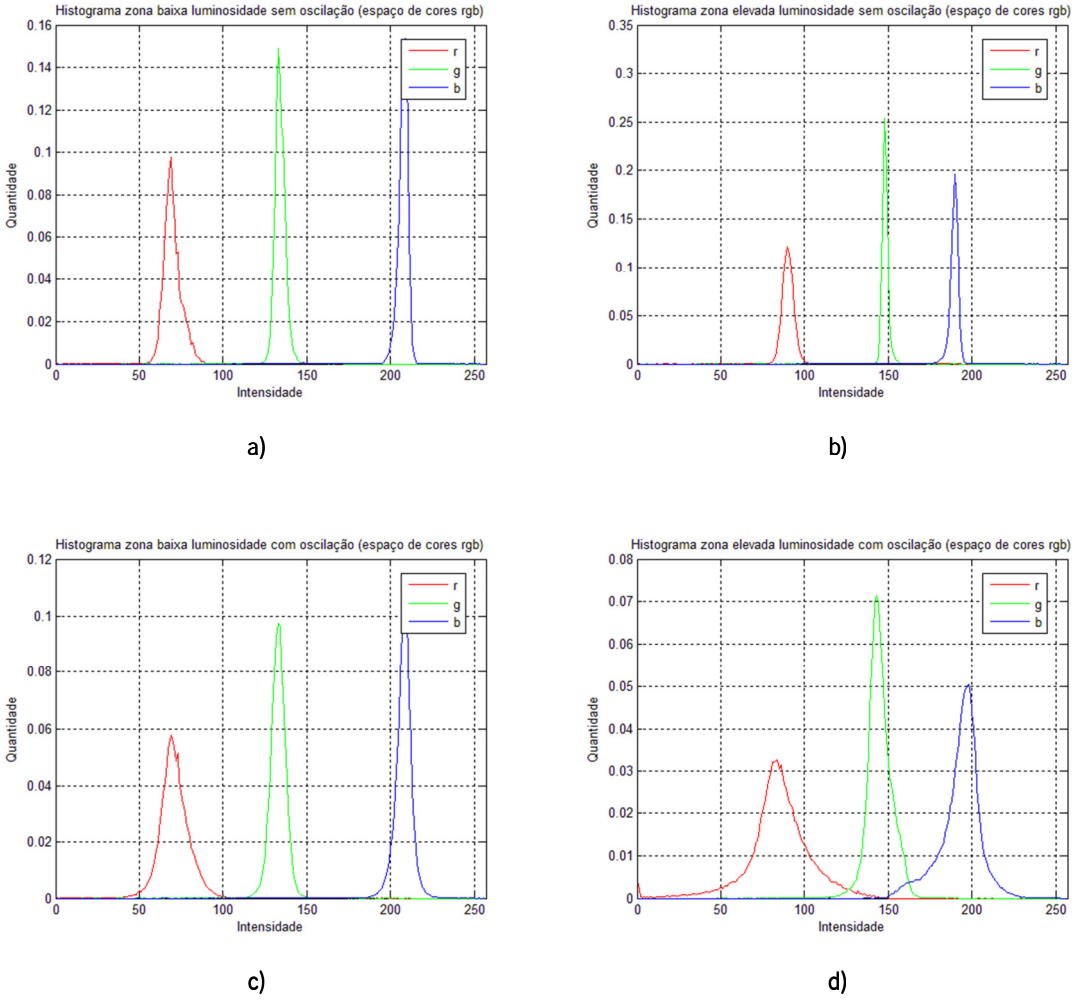
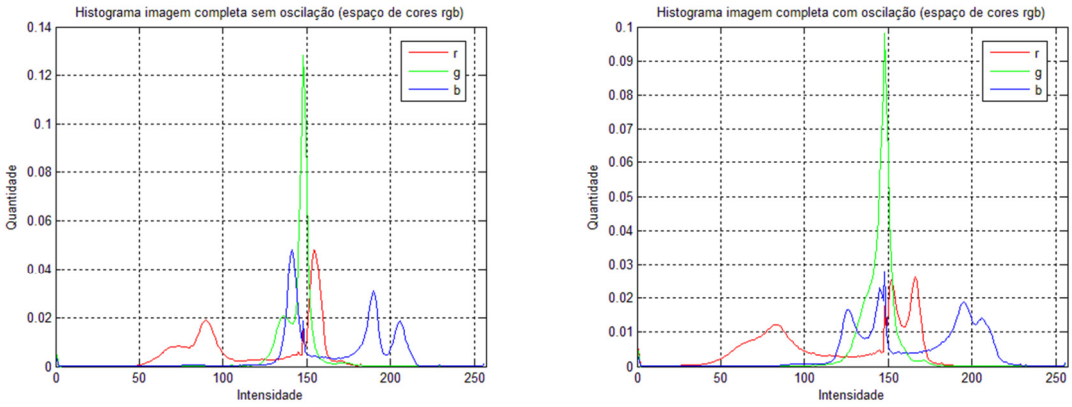


Figura 2.6.4: Histogramas das três componentes de cor rgb correspondentes à média temporal da média espacial, aplicada às duas localizações de diferente luminosidade, do conjunto dos 500 fotogramas sequenciais, em duas situações distintas de oscilação da superfície da água.

a) **Oscilação reduzida e baixa luminosidade;**
b) **Oscilação reduzida e luminosidade elevada;**
c) **Oscilação elevada e baixa luminosidade;**
d) **Oscilação elevada e luminosidade elevada.**

Apesar do espaço de cor RGB normalizado ser menos dependente das variações de luminosidade sobre a cena e oscilação da superfície da água, relativamente ao espaço de cor RGB, o mesmo apresenta, ainda assim, dependências significativas.



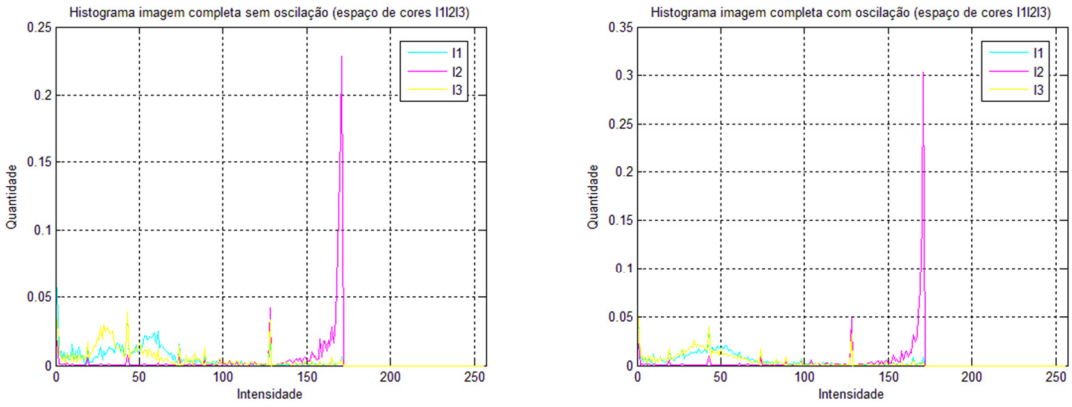
a) b)

Figura 2.6.5: Histogramas das três componentes de cor rgb correspondentes à média temporal da média espacial, aplicada à imagem completa, do conjunto dos 500 fotogramas sequenciais, em duas situações distintas de oscilação da superfície da água.

a) Oscilação reduzida da superfície da água;
b) Oscilação elevada da superfície da água.

2.6.3 Espaço de representação de cor $I_1I_2I_3$

O espaço de representação de cor $I_1I_2I_3$ apresenta poucas variações com a variação da intensidade da oscilação da superfície da água, tal como se pode verificar através da análise dos histogramas (a) e (b) da Figura 2.6.6. Existe alguma variação nas componentes I_1 e I_3, mas relativamente à componente I_2 nota-se uma total independência face à intensidade da oscilação da superfície da água.



a) b)

Figura 2.6.6: Histogramas das três componentes de cor I_1, I_2, I_3 correspondentes à média temporal da média espacial, aplicada à imagem completa, do conjunto dos 500 fotogramas sequenciais, em duas situações distintas de oscilação da superfície da água.

a) Oscilação reduzida da superfície da água;
b) Oscilação elevada da superfície da água.

Analisando apenas os histogramas das áreas correspondentes à superfície da água nota-se uma independência muito significativa da componente I_2, como seria de esperar, face aos resultados apresentados nos histogramas da Figura 2.6.6. Contudo as componentes I_1 e I_3, devido às variações na intensidade da luz, sofrem deslocações na média significativas, tal como se pode verificar pela análise dos histogramas (a), (b) e (c), (d) da Figura 2.6.7. O aumento da oscilação da superfície da água causa também uma maior dispersão destas componentes em torno da média. No entanto, a componente I_2 praticamente não é afectada pelas variações quer de luminosidade quer da intensidade da oscilação da superfície da água.

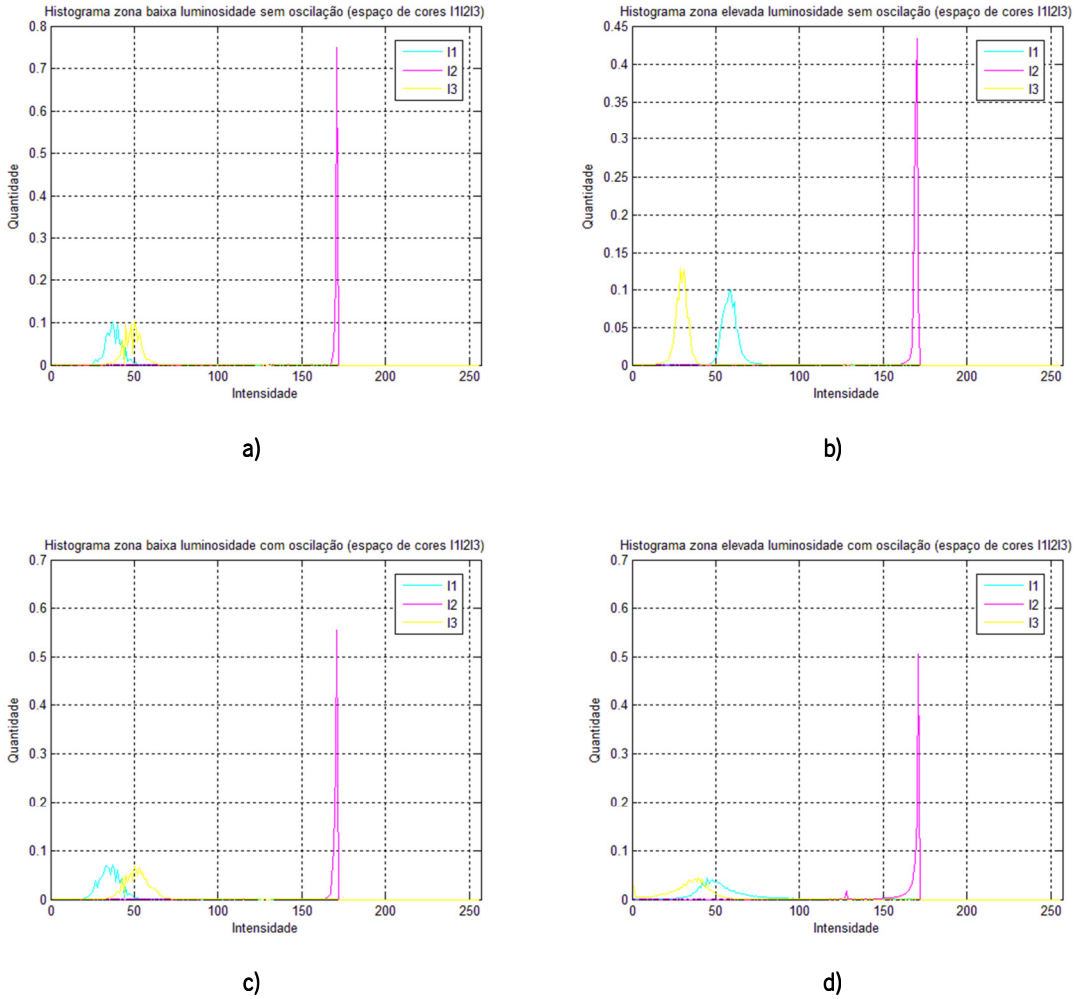
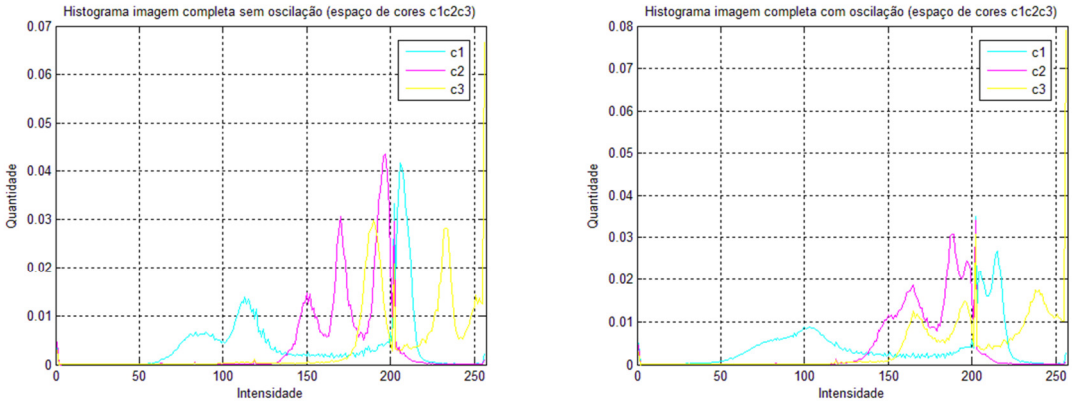


Figura 2.6.7: Histogramas das três componentes de cor I_1 I_2 I_3 correspondentes à média temporal da média espacial, aplicada às duas localizações de diferente luminosidade, do conjunto dos 500 fotogramas sequenciais, em duas situações distintas de oscilação da superfície da água.

a) Oscilação reduzida e baixa luminosidade;
b) Oscilação reduzida e luminosidade elevada;
c) Oscilação elevada e baixa luminosidade;
d) Oscilação elevada e luminosidade elevada.

2.6.4 Espaço de representação de cor $c_1c_2c_3$

O espaço de representação de cor $c_1c_2c_3$ mostra-se também muito dependente da intensidade da oscilação da superfície da água, tal como se pode verificar pela análise dos histogramas da Figura 2.6.8. Com uma oscilação da água maior os histogramas nas três componentes de cor ficam com picos mais baixos e apresentam maior dispersão em torno desses picos, o que demonstra a influência das variações da intensidade luminosa sobre a cena neste espaço de cor.



a) b)

Figura 2.6.8: Histogramas das três componentes de cor c_1 c_2 c_3 correspondentes à média temporal da média espacial, aplicada à imagem completa, do conjunto dos 500 fotogramas sequenciais, em duas situações distintas de oscilação da superfície da água.

a) Oscilação reduzida da superfície da água;
b) Oscilação elevada da superfície da água.

Fazendo uma análise aos histogramas gerados a partir dos dados provenientes das localizações na área correspondente à superfície da água é possível verificar que variações na intensidade da luz irradiada sobre a cena fazem variar as médias. Este facto é facilmente observado quando se comparam os histogramas (a) e (b) e os histogramas (c) e (d) da Figura 2.6.9. Além disso, este modelo de representação de cor é afectado pela intensidade da oscilação da água, fazendo com que o aumento da intensidade desta seja proporcional à dispersão causada em torno da média em todas as componentes, tal como se pode ver ao efectuar a comparação entre os histogramas (a) e (c) e os histogramas (b) e (d) da Figura 2.6.9.

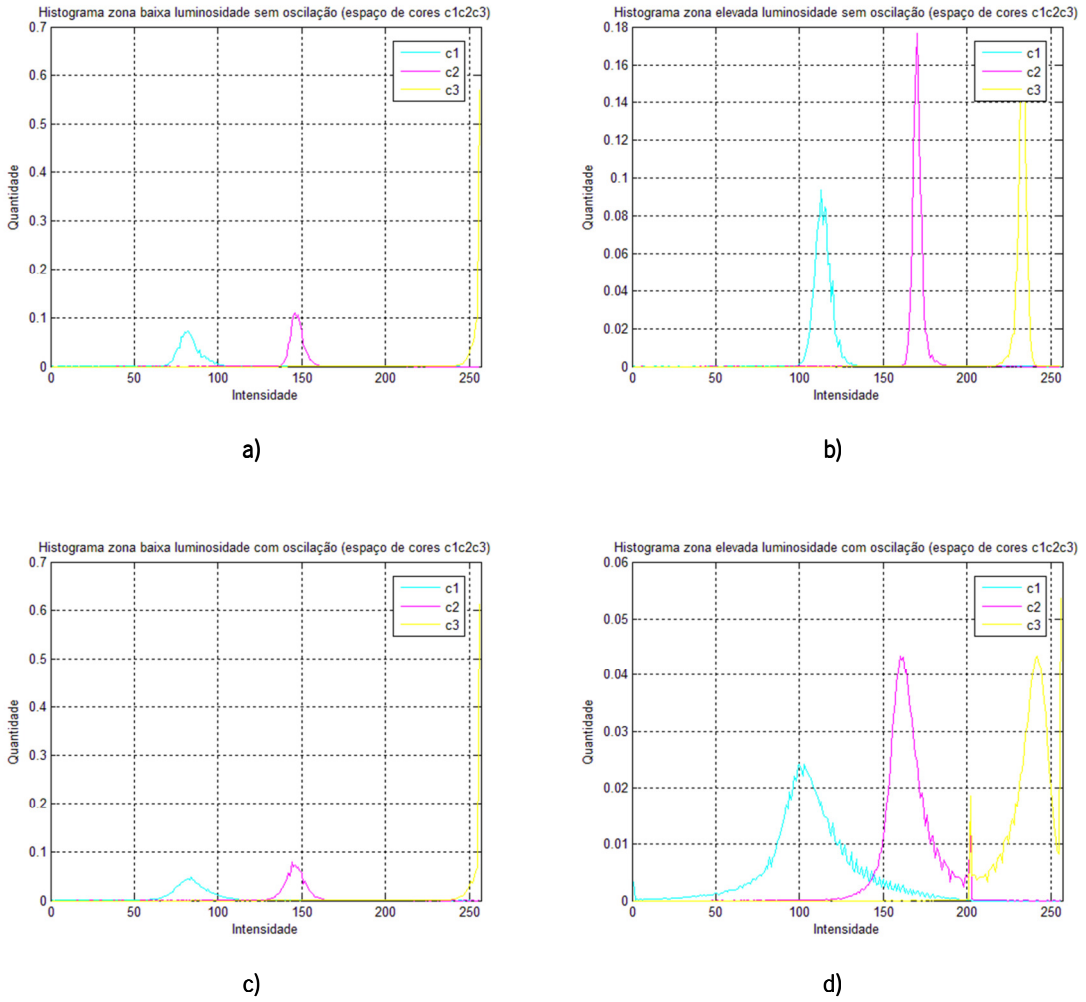
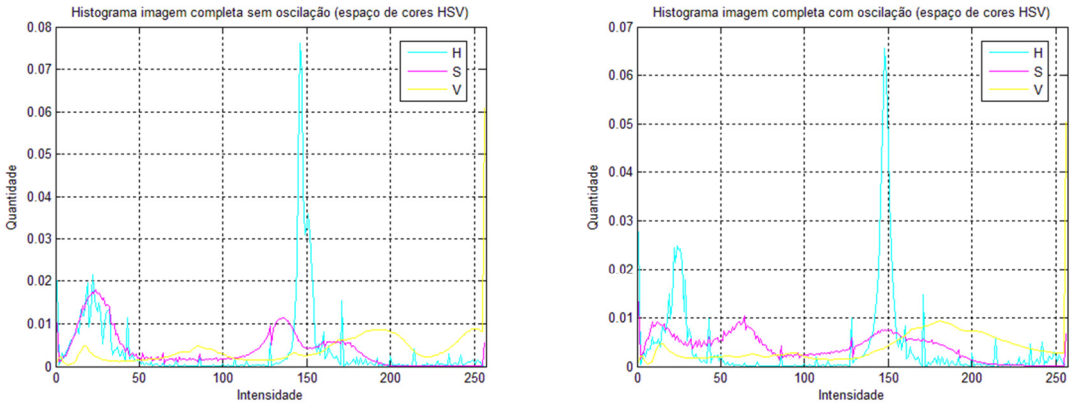


Figura 2.6.9: Histogramas das três componentes de cor c_1 c_2 c_3 correspondentes à média temporal da média espacial, aplicada às duas localizações de diferente luminosidade, do conjunto dos 500 fotogramas sequenciais, em duas situações distintas de oscilação da superfície da água.

a) Oscilação reduzida e baixa luminosidade;
b) Oscilação reduzida e luminosidade elevada;
c) Oscilação elevada e baixa luminosidade;
d) Oscilação elevada e luminosidade elevada.

2.6.5 Espaço de representação de cor HSV

Ao analisar os histogramas da Figura 2.6.10 verifica-se que a componente tonalidade do espaço de representação de cor HSV quase não sofre variação com a variação na intensidade da oscilação da superfície da água e que a componente saturação é pouco afectada. Por outro lado, a componente intensidade, como seria de esperar, apresenta uma maior dispersão quando a oscilação da água à superfície é superior.

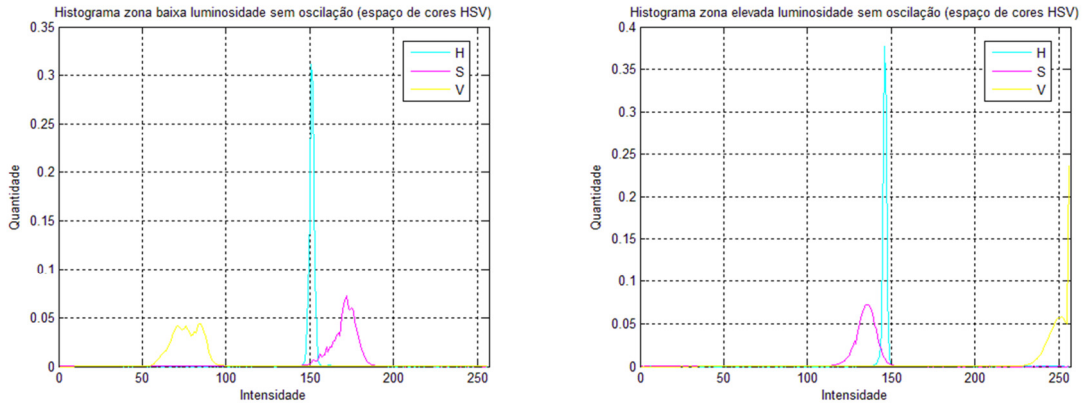


a) b)

Figura 2.6.10: Histogramas das três componentes de cor HSV correspondentes à média temporal da média espacial, aplicada à imagem completa, do conjunto dos 500 fotogramas sequenciais, em duas situações distintas de oscilação da superfície da água.

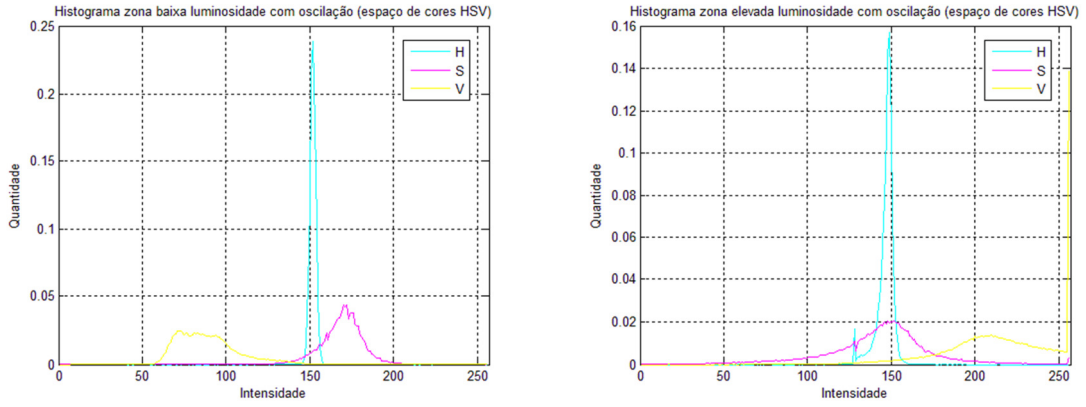
a) Oscilação reduzida da superfície da água;
b) Oscilação elevada da superfície da água.

Verificando os histogramas da Figura 2.6.11 é claramente notório a estabilidade da componente tonalidade nas diferentes situações de oscilação da superfície da água e da intensidade irradiada sobre a mesma. De facto, a variação da intensidade luminosa provoca um deslocamento diminuto na média da tonalidade, sendo também notório um aumento semelhante da dispersão em torno da média para valores mais elevados de oscilação da superfície da água. Relativamente à componente saturação, o aumento de luminosidade faz com que o valor da sua média seja deslocado no sentido da diminuição da intensidade da mesma, de uma forma mais destacada quando comparada com a componente tonalidade. O aumento da dispersão em torno da média desta componente com o aumento da oscilação da superfície da água também é evidente, sendo muito superior ao sofrido pela tonalidade. Contudo, o valor da componente intensidade acompanha não só as variações de luminosidade sobre a cena como a oscilação da água, aumentando ou diminuindo conforme a intensidade da luz aumenta ou diminui e tendo maior ou menor dispersão consoante a oscilação da superfície da água seja maior ou menor, respectivamente.



a)

b)



c)

d)

Figura 2.6.11: Histogramas das três componentes de cor HSV correspondentes à média temporal da média espacial, aplicada às duas localizações de diferente luminosidade, do conjunto dos 500 fotogramas sequenciais, em duas situações distintas de oscilação da superfície da água.

a) Oscilação reduzida e baixa luminosidade;
b) Oscilação reduzida e luminosidade elevada;
c) Oscilação elevada e baixa luminosidade;
d) Oscilação elevada e luminosidade elevada.

Os histogramas tridimensionais das componentes tonalidade e saturação, visto que são as que apresentam maior invariabilidade, são apresentados na Figura 2.6.12. Através dos mesmos podemos verificar que existe pouca dependência destas duas componentes não só da variação da intensidade luminosa como da variação da intensidade de oscilação da superfície da água. A localização da média é muito próxima nas quatro situações, sendo apenas evidenciada uma maior dispersão em torno da mesma quando a oscilação da superfície da água é maior.

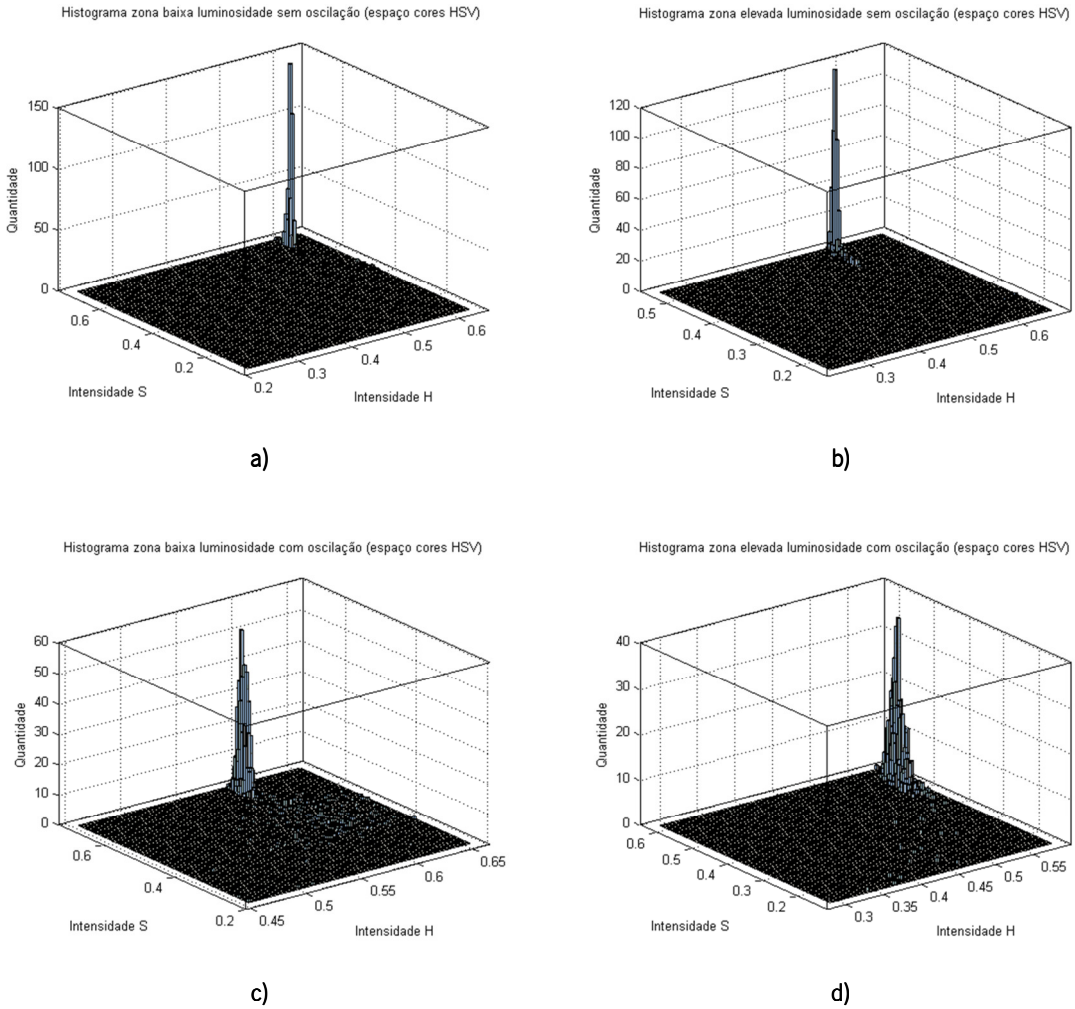


Figura 2.6.12: Histogramas tridimensionais das componentes de cor H e S do espaço de cores HSV correspondentes à média temporal da média espacial, aplicada às duas localizações de diferente luminosidade, do conjunto dos 500 fotogramas sequenciais, em duas situações distintas de oscilação da superfície da água.

a) Oscilação reduzida e baixa luminosidade;
b) Oscilação reduzida e luminosidade elevada;
c) Oscilação elevada e baixa luminosidade;
d) Oscilação elevada e luminosidade elevada.

2.6.6 Espaço de representação de cor CIELa*b*

Na Figura 2.6.13 apresentam-se os histogramas tridimensionais relativos às componentes a* e b* do espaço de cores CIELa*b*. Ao analisar os histogramas (a) e (b) repara-se numa deslocação efectiva do valor da média devido à variação de luminosidade, algo que não acontece no espaço de cor HSV. Além disso a intensidade da oscilação da superfície da água tem uma grande influência na dispersão causada em torno da média, sendo isso relevante quando se comparam os histogramas (a) com (c) e (b) com (d).

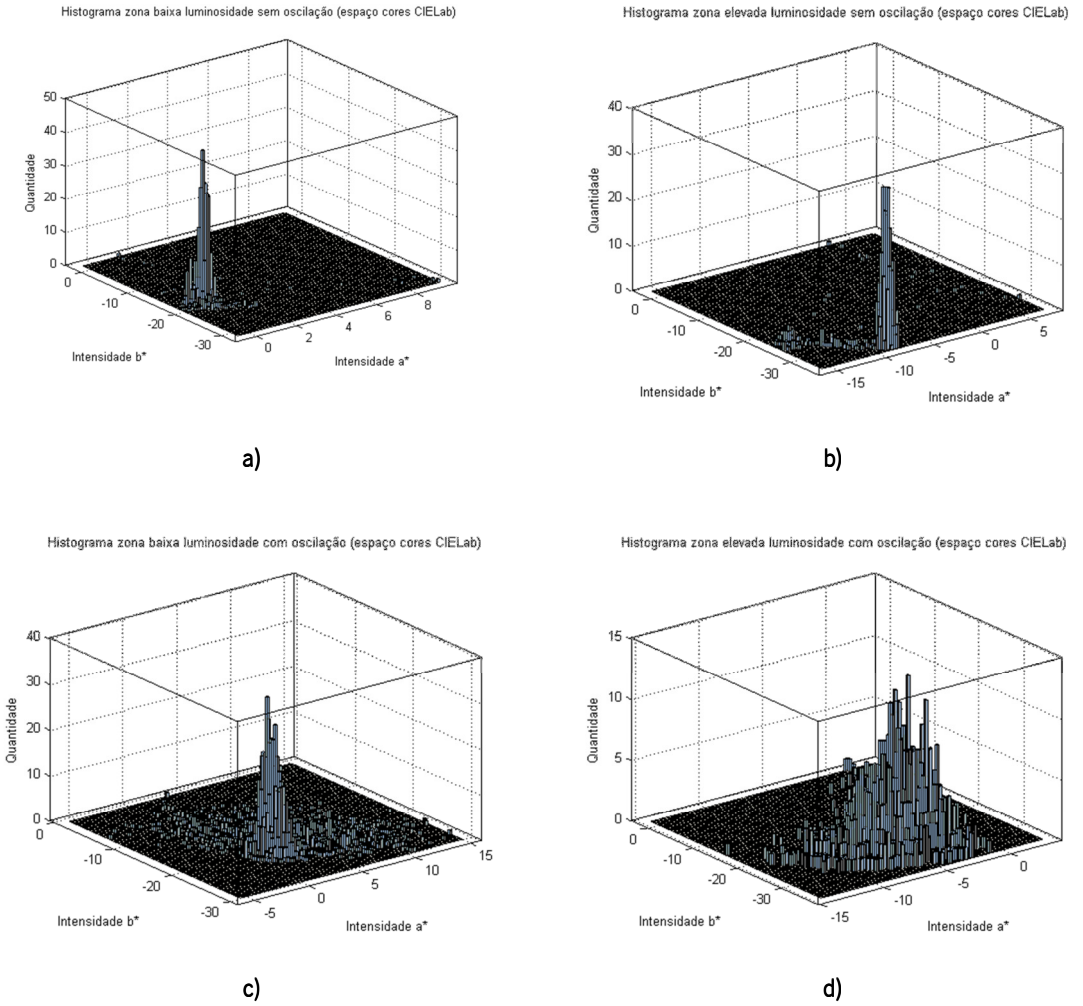


Figura 2.6.13: Histogramas tridimensionais das componentes de cor a* e b* do espaço de cores CIELab correspondentes à média temporal da média espacial, aplicada às duas localizações de diferente luminosidade, do conjunto dos 500 fotogramas sequenciais, em duas situações distintas de oscilação da superfície da água.

a) Oscilação reduzida e baixa luminosidade;
b) Oscilação reduzida e luminosidade elevada;
c) Oscilação elevada e baixa luminosidade;
d) Oscilação elevada e luminosidade elevada.

Depois de analisados os vários espaços de representação de cor quanto à sua invariabilidade a brilhos e variação da intensidade luminosa irradiada sobre a cena é possível concluir que o espaço de cor com as melhores características de invariabilidade é o HSV. De facto é aquele que em todas as situações apresenta uma dependência menor, principalmente na componente tonalidade, que define a matiz da cor. Além disso proporciona uma componente que descreve fielmente a intensidade luminosa incidente sobre a cena. Utilizando as componentes tonalidade e saturação é possível definir a cor, separando assim de forma efectiva a crominância da luminosidade. As características muito especiais deste espaço de representação de cor conferem um benefício óbvio

quanto à sua utilização na segmentação num ambiente aquático, como é o da piscina, altamente variável e complexo, como se destacou na parte inicial deste capítulo. Através da análise experimental e também analítica, prova-se que o espaço de representação de cor HSV é o mais favorável para efectuar a segmentação num ambiente aquático complexo, sendo deste modo o modelo escolhido para tal, e no qual todo o algoritmo de segmentação desenvolvido se baseia.

2.7 Segmentação híbrida de movimento num ambiente aquático complexo

Tal como foi mencionado em secções anteriores, a superfície da água apresenta diferentes padrões de brilhos e reflexões, causados pelos diferentes níveis de oscilação em que se encontra. As projecções da fonte de luz e das sombras no fundo da piscina, também altamente variáveis, em função da oscilação da superfície da água, exigem técnicas de segmentação de movimento diferentes das tradicionais, uma vez que com as mesmas, toda a área correspondente à piscina é classificada como movimento e como tal *foreground*. Deste modo, e como a restante cena envolvente da piscina não sofre dos mesmos problemas, optou-se por uma abordagem que implementa um algoritmo de segmentação híbrido, Figura 2.7.1. Existem assim dois algoritmos que executam em simultâneo, um deles na área da imagem correspondente à piscina, e o outro na restante área da imagem, i.e., a que não pertence à piscina. O primeiro algoritmo tem características especiais e foi especificamente concebido para segmentar num ambiente aquático complexo. Este algoritmo deve ser capaz de eliminar a maioria dos problemas anteriormente descritos e que estão subjacentes à oscilação da superfície da água. Quanto ao segundo algoritmo optou-se por uma solução tradicional, no caso, a mistura de gaussianos (MoG), uma vez que se trata de uma cena exterior normal e este algoritmo é um dos melhores a fazê-lo, (Piccardi, 2004). A junção dos mapas binários correspondentes ao *foreground* gerados por cada algoritmo é depois efectuada fornecendo o mapa binário global correspondente à detecção de movimento em toda a cena. No entanto, tal metodologia implica uma classificação dos pixéis da imagem pertencentes apenas a dois grupos, o grupo da piscina e os restantes. Para se efectuar esta separação foi concebido um algoritmo de detecção automática da piscina, capaz de gerar, previamente, uma máscara binária da área correspondente à mesma.

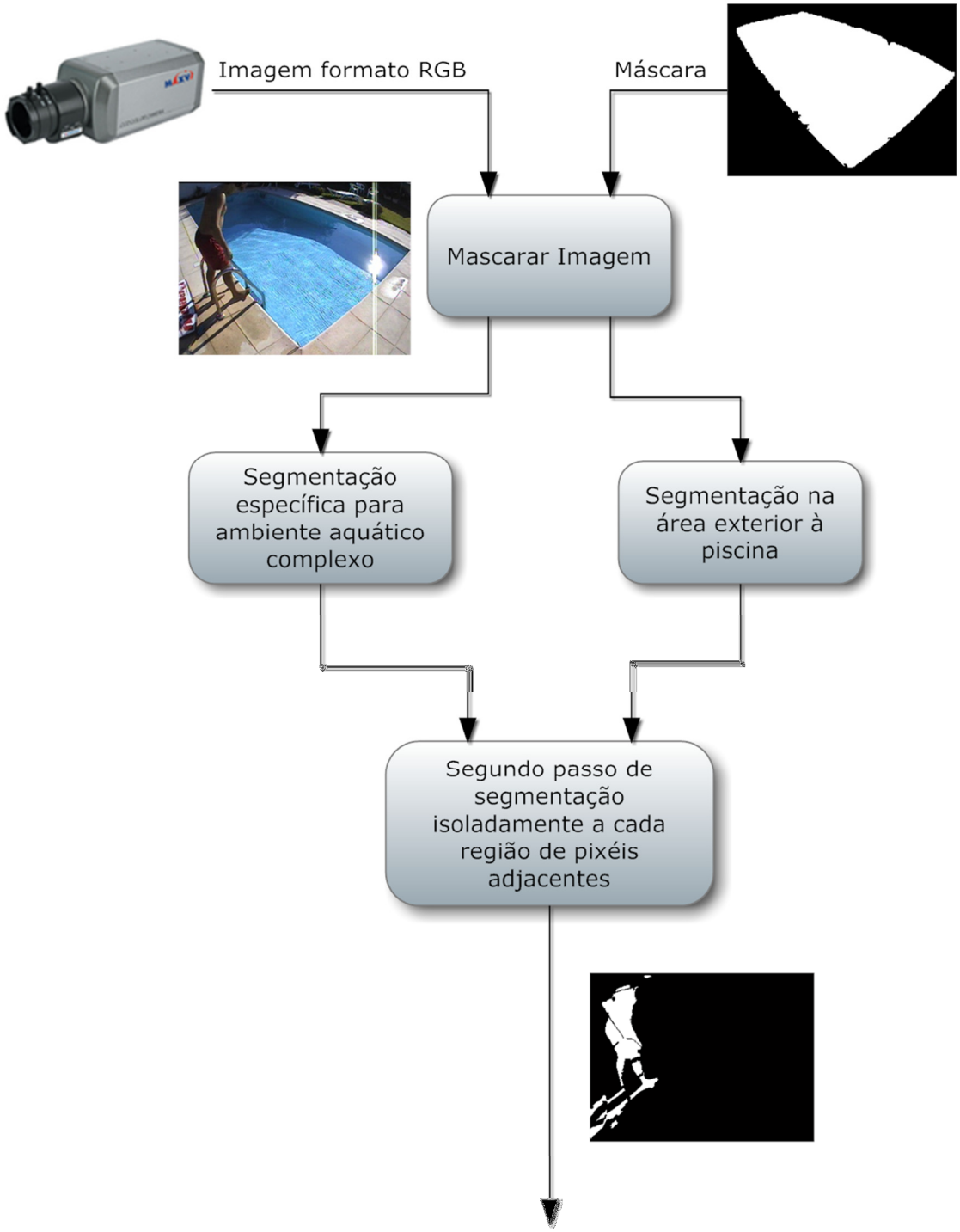


Figura 2.7.1: Esquema do algoritmo de segmentação híbrido. Depois de mascarar a imagem, o algoritmo especializado em ambientes aquáticos complexos, executa na área correspondente à piscina e o algoritmo tradicional executa na restante área. A união dos mapas binários de *foreground* de cada algoritmo resulta num mapa binário a partir do qual são extraídas as diferentes regiões e re-segmentadas com recurso ao algoritmo de *clustering k-means* gerando assim o *foreground* final da cena.

2.7.1 Algoritmo de detecção automática da piscina

Tal como foi concluído anteriormente o espaço de cores com melhores características de invariabilidade é o HSV, sendo por isso o modelo de representação de cor utilizado para a detecção automática da área da piscina. Para que a área da mesma seja detectada de forma correcta pelo algoritmo têm que ser cumpridos os seguintes requisitos:

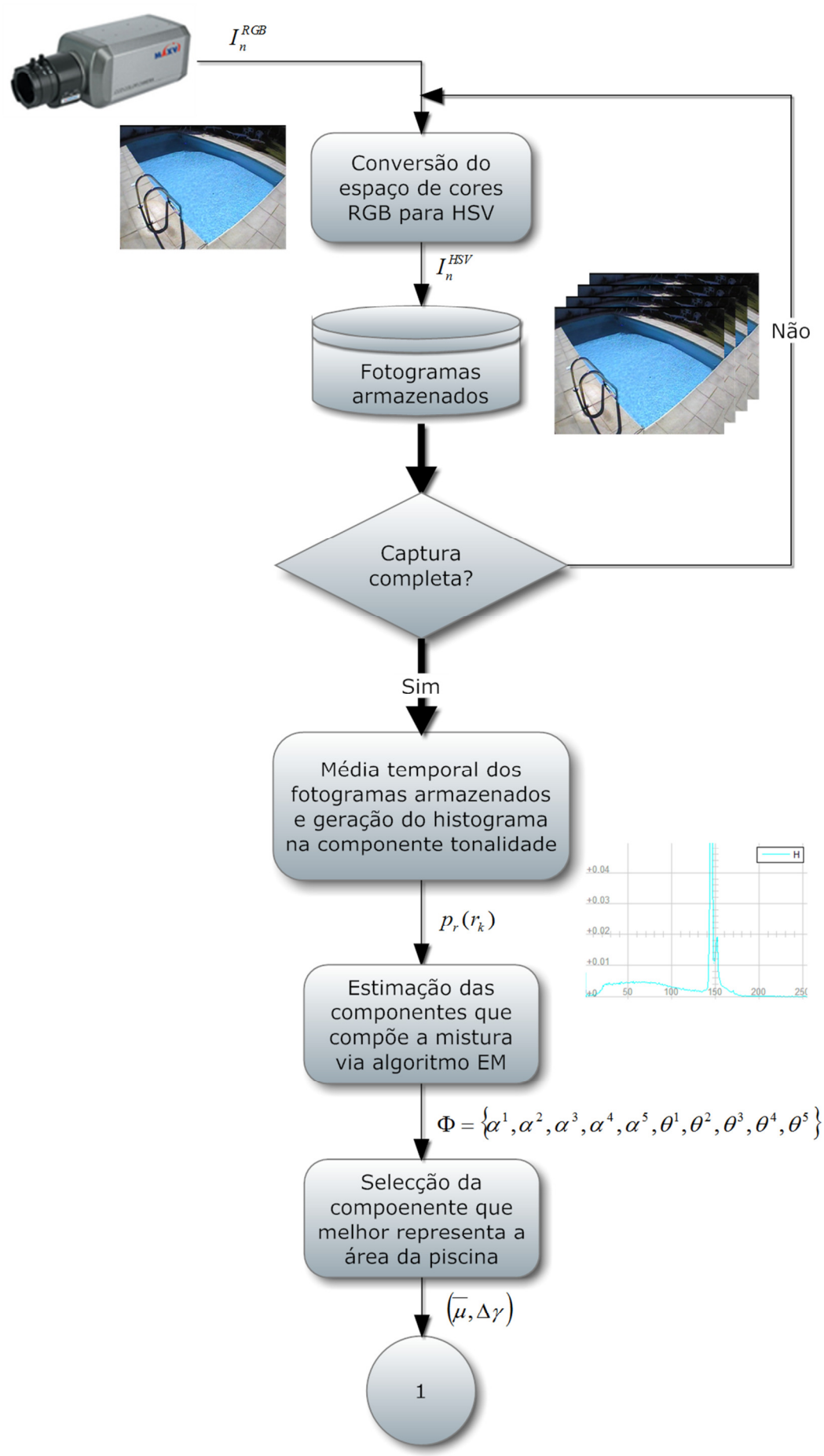
1. A área da imagem correspondente à piscina deve ocupar mais de 50% da imagem capturada. Este é um requisito que se cumpre pelas características do próprio sistema, pois o mais importante, e que estará no foco de atenção, é a piscina, pelo que a câmara deverá estar direccionada de modo que apanhe a maior parte desta;

2. Durante a execução do algoritmo de detecção automática da piscina não é aconselhável a existência de objectos ou pessoas na mesma, para que deste modo o processo seja o mais fiável possível. No entanto poderão existir, mas quanto maior for a quantidade dos mesmos menor será a qualidade do processo.

Para separar os pixéis pertencentes à piscina dos restantes, o algoritmo baseia-se na premissa de que a piscina corresponde à maior área de pixéis conjuntos e com tonalidade homogénea dentro dos limites correspondentes à matiz azul no espaço de cores HSV. Apesar de não ser totalmente verdadeira, uma vez que grandes áreas de cor homogéneas que não sejam uma piscina cumprem estas condições, a afirmação, no âmbito da aplicação do sistema e cumpridos os dois requisitos anteriormente expostos, pode ser considerada verdadeira. Desta forma, foi elaborado o algoritmo, apresentado na Figura 2.7.2, que permite descobrir automaticamente a localização da piscina, com base no pressuposto anteriormente.

Como a captura da imagem é feita no formato RGB, existe a necessidade de fazer uma conversão desta para o espaço de cores HSV, de acordo com as equações 2.5.9 e 2.5.12. Este é assim o primeiro estágio de processamento do algoritmo de detecção automática da piscina. Seguidamente a média temporal da imagem I^H, apenas na componente H, é efectuada de acordo com a equação 2.7.1:

$$\mu_n^H(x,y) = \frac{\sum_{i=0}^{n} I_i^H(x,y)}{n}, com\ n = [100, 150] \quad (2.7.1)$$


I_n^{RGB}
Conversão do espaço de cores RGB para HSV
I_n^{HSV}
Fotogramas armazenados
Não
Captura completa?
Sim
Média temporal dos fotogramas armazenados e geração do histograma na componente tonalidade
$p_r(r_k)$
H
Estimação das componentes que compõe a mistura via algoritmo EM
$\Phi = \{\alpha^1, \alpha^2, \alpha^3, \alpha^4, \alpha^5, \theta^1, \theta^2, \theta^3, \theta^4, \theta^5\}$
Selecção da compoenente que melhor representa a área da piscina
$(\bar{\mu}, \Delta\gamma)$
1

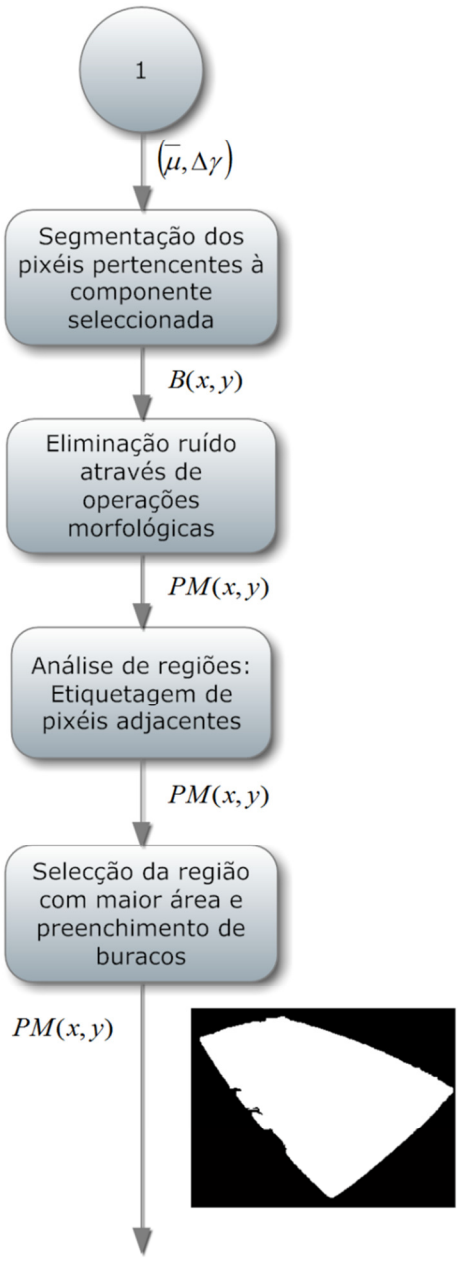


Figura 2.7.2: Esquema do algoritmo de geração automática da máscara binária correspondente à localização da piscina.

A uma velocidade de captura de 25 fotogramas por segundo, um valor de $\boldsymbol{n}$ dentro dos limites impostos pela equação 2.7.1 corresponde a cerca de 4 a 6 segundos de vídeo. Esta média temporal, μ_n^H, é importante para suavizar a imagem de modo a homogeneizar a mesma e eliminar o ruído proveniente da oscilação da água. A próxima fase de processamento consiste na criação do histograma espacial da imagem μ_n^H, de acordo com a equação que se apresenta a seguir:

$$p_r(r_k) = \frac{n_k}{N \times M}, para\ k = 0, 1, 2, \dots, L-1\ e\ L = 256 \quad (2.7.2)$$

Na equação 2.7.2 n_k corresponde ao número de pixéis existentes na imagem μ_n^H para o nível de tonalidade r_k, sendo N e M as dimensões da mesma. Como cada componente de cor RGB é quantificada por 8 bits, existem assim 256 valores possíveis para cada uma delas, tendo deste modo $L = 256$ na equação 2.7.2. Findo o processo de captura e geração do histograma espacial da média da imagem ao longo do tempo, $p_r(r_k)$, inicia-se o processo de determinação da mistura gaussiana que dá origem ao histograma encontrado. Os parâmetros desta mistura são calculados com base no algoritmo *Expectation Maximization* (EM), (Bilmes, 1998), partindo do pressuposto de que uma mistura de 5 gaussianos é normalmente suficiente para definir o histograma $p_r(r_k)$. Os parâmetros iniciais da mistura, Φ_0, são obtidos a partir do algoritmo de *clustering k-means*. Assim sendo, tem-se o seguinte:

$$\Phi_0 = \{\alpha_0^1, \dots, \alpha_0^k, \theta_0^1, \dots, \theta_0^k\}, com\ \theta_0^k = \{\mu_0^k, \sigma_0^k\}, e\ k = 5$$

Onde α_0^k são os pesos de cada classe determinados a partir do quociente entre o número de valores pertencentes à classe k, determinados pelo algoritmo *k-means*, e o conjunto total de dados $N \times M$. O algoritmo EM é um processo iterativo com duas fases, designadas por *Expectation* (*E-step*) e *Maximization* (*M-step*). De seguida explica-se sucintamente as duas fazes do algoritmo e o critério de paragem.

E-step:

Determinar o *likelihood* ou probabilidade de cada pixel da imagem $x = \mu_n^H(x, y)$ pertencer à classe k, com parâmetros θ estimados na iteração anterior, ou seja:

$$p(x|k, \mu_k, \sigma_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(x-\mu_k)^2}{2\sigma_k^2}} \quad (2.7.3)$$

Através da regra de *Bayes* determinar a probabilidade a *posteriori* da classe k representar o pixel x, de acordo com a seguinte equação:

$$p(k|x, \mu_k, \sigma_k) = \frac{\alpha_k p(x|k, \mu_k, \sigma_k)}{\sum_{i=1}^{N\times M} p(x_i|k, \mu_k, \sigma_k)} \quad (2.7.4)$$

M-step:

Maximizando o valor esperado na equação anterior, para cada classe k, obtêm-se os novos valores dos parâmetros da mistura de acordo com:

$$\alpha_k^{novo} = \frac{1}{N \times M} \sum_{i=1}^{N \times M} p(k|x_i, \mu_k, \sigma_k)$$

(2.7.5)

$$\mu_k^{novo} = \frac{\sum_{i=1}^{N \times M} x_i p(k|x_i, \mu_k, \sigma_k)}{\sum_{i=1}^{N \times M} p(k|x_i, \mu_k, \sigma_k)}$$

(2.7.6)

$$\sigma_k^{2^{novo}} = \frac{\sum_{i=1}^{N \times M} p(k|x_i, \mu_k, \sigma_k)(x_i - \mu_k^{novo})^2}{\sum_{i=1}^{N \times M} p(k|x_i, \mu_k, \sigma_k)}$$

(2.7.7)

A operação termina se a seguinte condição que relaciona o *log-likelihood* actual e anterior for inferior ao *threshold* T, tipicamente igual a 1^{-10}:

$$\left| \frac{\sum_{i=1}^{M \times N} log\left(\sum_{j=1}^{k} \alpha_j p\left(x|k, \mu_j^{novo}, \sigma_j^{novo}\right)\right)}{\sum_{i=1}^{M \times N} log\left(\sum_{j=1}^{k} \alpha_j p(x|k, \mu_j, \sigma_j)\right)} - 1 \right| < T$$

(2.7.8)

Depois de determinados os novos parâmetros da mistura volta ao E-step, até que a condição de paragem se verifique. Quando isso acontecer o algoritmo convergiu e determinou os parâmetros das classes k que representam o histograma $p_r(r_k)$.

Terminado o processo iterativo que dá origem à descoberta dos cinco gaussianos que representam o histograma são escolhidos aqueles que se encontram dentro dos limites da cor azul no que confere à tonalidade no espaço de cores HSV. Esta gama de valores de H encontra-se entre os 150° e os 270°, correspondendo aproximadamente à gama de valores compreendida entre $H_{min} = 107$ e $H_{max} = 192$ no histograma $p_r(r_k)$. O Algoritmo 2.7.1 permite determinar os limites da gama de valores, na componente tonalidade, pertencentes à piscina.

$\Delta\gamma = 0$

$\bar{\mu} = 0$

$cnt = 0$

De $i = 0$ até 5 fazer

Se $H_{min} \leq \mu_i \leq H_{max}$ então:

$\bar{\mu} = \bar{\mu} + \mu_i$

$\Delta\gamma = \Delta\gamma + \sigma_i$

$cnt = cnt + 1$

Fim

$\bar{\mu} = \frac{\bar{\mu}}{cnt}$

Algoritmo 2.7.1: Determina os limites da gama de valores, na componente tonalidade, pertencentes à piscina.

Deste modo, a segmentação da imagem é efectuada de acordo com a equação 2.7.9, sendo gerado um mapa binário correspondente aos pixéis que se encontram dentro dos limites impostos pela mistura gaussiana que define o histograma. O valor $\bar{\mu}$ define o ponto médio sobre o qual se centram a gama de valores, sendo os limites da mesma dados pela distância $\Delta\gamma$ à direita e à esquerda de $\bar{\mu}$.

$$B(x,y) = \begin{cases} 1, se\ \bar{\mu} - \Delta\gamma \leq \mu_n^H(x,y) \leq \bar{\mu} + \Delta\gamma \\ 0, \qquad\qquad caso\ contrário \end{cases}$$

(2.7.9)

O mapa binário gerado vem acompanhado de ruído proveniente da segmentação, sendo este eliminado através da aplicação de uma erosão seguida de uma dilatação com um *kernel* unitário KE quadrado de 4×4 e um kernel unitário KD quadrado de 5×5, de acordo com a equação 2.7.10.

$$PM(x,y) = (B(x,y) \ominus KE) \oplus KD$$

(2.7.10)

Depois desta operação, a maioria das áreas diminutas, constituídas por poucos pixéis conjuntos, já desapareceram, não sendo no entanto as áreas maiores afectadas pela erosão, uma vez que sofreram também dilatação. No entanto podem existir outras áreas, onde as condições anteriores se verifiquem, que aparecem no mapa binário PM. De modo a eliminar estas áreas é escolhida a

maior das regiões conjuntas de pixéis com base numa conectividade de 4. Esta última operação permite seleccionar a área correspondente à piscina sendo apenas necessário efectuar o preenchimento dos buracos no interior da mesma através da aplicação do algoritmo concebido por (Soille & Gratin, 2004). Terminado este processo, a imagem binária PM contém a máscara correspondente à área ocupada pela piscina na cena.

2.7.2 Algoritmo de segmentação híbrido com re-segmentação por objectos

O algoritmo de segmentação de movimento tem como principal objectivo detectar todos os objectos no interior da piscina, que não pertençam à mesma, e todos os objectos em movimento na área circundante desta. A saída do processo de segmentação deve assim gerar um mapa binário com as áreas dos objectos detectados na cena de acordo com os requisitos expostos anteriormente. Deste modo, e para que a segmentação seja mais eficiente, o algoritmo será na realidade composto por dois algoritmos, cada um deles executando em áreas diferentes da imagem. Estas áreas são determinadas pela máscara binária resultante do algoritmo da Figura 2.7.2, que são geradas previamente na calibração do sistema. A união dos dois mapas binários é sujeita a uma análise das regiões de *foreground* sendo as mesmas re-segmentadas com o algoritmo de *clustering k-means*. Desse processo resulta um mapa binário de *foreground* com elevada precisão e qualidade, praticamente imune aos diferentes ruídos presentes na cena.

2.7.2.1 Algoritmo de segmentação para ambientes aquáticos complexos

O algoritmo de segmentação específico para ambientes aquáticos complexos é baseado, fundamentalmente, na componente tonalidade do espaço de cores HSV. Na Figura 2.7.3 mostra-se um esquema do algoritmo de segmentação. Nele pode ser observado que a geração do mapa binário de *foreground* é efectuada por intermédio da combinação da subtracção de *background* com subtracção entre fotogramas na componente tonalidade. O *background* é criado dinamicamente através da média temporal da imagem, sendo depois corrigida por um filtro passa baixo espacial. Por último é efectuada uma correcção ao mapa binário de *foreground* resultante de modo a eliminar as áreas menores da imagem que apresentam ruído gerado no processo de segmentação.

Tal como o esquema da Figura 2.7.3 esclarece, o algoritmo inicia-se com a conversão do espaço de cores RGB para HSV de acordo com as equações 2.7.11, 2.7.12 e 2.7.13.

$$H = \begin{cases} \frac{\left(\frac{G-B}{MAX-MIN}\right)}{6}, se\ R = MAX \\ \frac{\left(2 + \frac{B-R}{MAX-MIN}\right)}{6}, se\ G = MAX \\ \frac{\left(4 + \frac{R-G}{MAX-MIN}\right)}{6}, se\ B = MAX \end{cases}$$

(2.7.11)

$$S = \frac{MAX - MIN}{MAX}$$

(2.7.12)

$$V = MAX$$

(2.7.13)

Assumindo que a imagem tem dimensões $N \times M$ e que depois de aplicada a mascara a mesma é dada por $I_n^{RGB}(x, y)$, num dado momento n, depois de convertida para o espaço de cores HSV, tem-se que, a imagem é definida por $I_n^{HSV}(x, y)$. Como o algoritmo de segmentação usa apenas a componente tonalidade, a imagem é dada por $I_n^H(x, y)$. A primeira fase do algoritmo de segmentação, efectuada pela equação 2.7.14, corresponde à subtracção entre fotogramas consecutivos de modo a extrair o movimento dos objectos não pertencentes à água.

$$IAD_n(x, y) = |I_n^H(x, y) - I_{n-1}^H(x, y)|$$

(2.7.14)

Desta subtracção resulta a imagem $IAD_n(x, y)$ que é depois sujeita a uma dilatação com um kernel unitário UK de 3×3 resultando na imagem $IADD_n(x, y)$, tal como a equação 2.7.15 sugere. Esta operação morfológica tem como objectivo intensificar os pontos onde existe movimento na imagem.

$$IADD_n(x, y) = IAD_n(x, y) \oplus UK$$

(2.7.15)

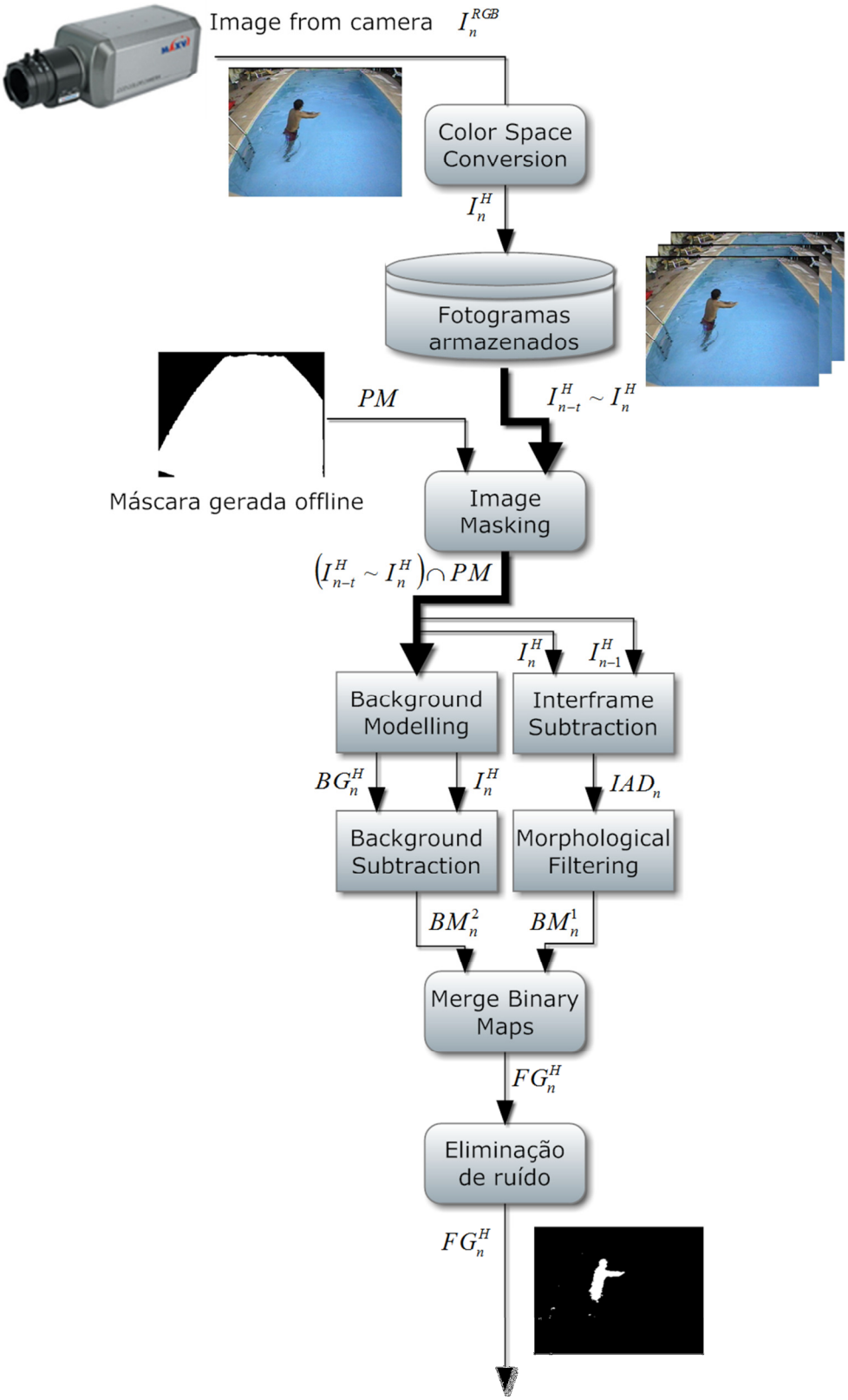


Figura 2.7.3: Esquema do algoritmo de segmentação específico para ambientes aquáticos complexos. O mapa binário de *foreground* é gerado com base na união dos mapas binários resultantes da subtracção entre fotogramas e da subtracção do *background* na componente tonalidade.

A esta imagem é efectuada uma média espacial com um kernel unitário de 5×5, designado por $W(s,t), -2 \leq s \leq 2 \wedge -2 \leq t \leq 2$, de modo a aumentar ainda mais as zonas onde existe movimento, sendo depois aplicado uma *threshold* τ, dando origem ao primeiro mapa binário de *foreground* $BM_n^1(x,y)$, tal como a equação 2.7.16 pretende demonstrar.

$$BM_n^1(x,y) = \begin{cases} 1, se \ \frac{\sum_{s=-2}^{2}\sum_{t=-2}^{2} W(s,t) \cdot IADD_n\,(x+s, y+t)}{\sum_{s=-2}^{2}\sum_{t=-2}^{2} W(s,t)} \geq \tau \\ 0, \qquad\qquad caso\ contrário \end{cases}, com\ 0.05 \leq \tau \leq 0.10$$

(2.7.16)

A gama de valores do *threshold* τ foi determinada experimentalmente, sendo verificada uma maior eficácia dentro dos limites declarados.

O segundo mapa binário, dado por $BM_n^2(x,y)$, resultante da subtracção da imagem actual $I_n^H(x,y)$ ao *background* $BG_n(x,y)$, é dado pela equação 2.7.17.

$$BM_n^2(x,y) = \begin{cases} 1, se|BG_n(x,y) - I_n^H(x,y)| \geq \tau \\ 0, \qquad\qquad caso\ contrário \end{cases}, com\ 0.05 \leq \tau \leq 0.10$$

(2.7.17)

A união entre os dois mapas binários $BM_n^1(x,y)$ e $BM_n^2(x,y)$, dada pela equação 2.7.18 permite combinar os resultados obtidos pelos dois processos, que assim se complementam, dando origem ao mapa binário final de *foreground* $FG_n^H(x,y)$.

$$FG_n^H(x,y) = BM_n^1(x,y) \cup BM_n^2(x,y)$$

(2.7.18)

A estimação do *background* é baseada na média temporal dos últimos t fotogramas sequenciais da imagem $I^H(x,y)$, tal como a equação 2.7.19 pretende demonstrar.

$$BG_n^H(x,y) = \frac{\sum_{i=n-t}^{n} I_i^H(x,y)}{t}$$

(2.7.19)

No entanto, estimado desta forma, principalmente quando existem objectos na piscina que não pertencem à mesma, o *background* dado pela equação 2.7.19 aparece corrompido pelos mesmos, existindo assim a necessidade de o corrigir. Desta feita, calculando a média espacial e desvio padrão da imagem $BG_n^H(x,y)$, equações 2.7.20 e 2.7.21, respectivamente, é possível

implementar uma correcção nos pixéis que caiam fora da gama compreendida entre o valor médio acrescido e decrescido de k vezes o desvio padrão.

$$\mu_n^H = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} BG_n^H(i,j)}{N \times M} \tag{2.7.20}$$

$$\sigma_n^H = \sqrt{\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(BG_n^H(i,j) - \mu_n^H)^2}{N \times M - 1}} \tag{2.7.21}$$

A condição 2.7.22 estipula que todos os pixéis da imagem $BG_n^H(x,y)$ que a satisfaçam serão substituídos pelo valor correspondente à média espacial da imagem definida por μ_n^H.

$$(BG_n^H(x,y) < \mu_n^H - k \cdot \sigma_n^H) \vee (BG_n^H(x,y) > \mu_n^H + k \cdot \sigma_n^H) \tag{2.7.22}$$

Experimentalmente, observa-se que, com um valor de $k = 2.5$ os resultados da segmentação são muito bons na grande maioria dos casos, sendo por isso um parâmetro que não é necessário calibrar durante a operação.

2.7.2.2 Algoritmo de segmentação da parte exterior da piscina

Para o exterior da piscina foi utilizado um algoritmo tradicional, no caso o MoG, tal como é descrito em (Power & Schoonees, 2002), uma vez que o *background* não tem as especificidades da água, mas é exterior e por isso afectado pelas variações de luminosidade e movimentos esporádicos do *background*, devido por exemplo ao movimento das árvores causado pelo vento. Este algoritmo executa apenas no mapa binário inverso da área correspondente à piscina definida na Secção 2.7.1. A imagem depois de mascarada é convertida para o espaço de representação de cor HSV, sendo utilizadas apenas as componentes H e S do referido espaço. Depois de efectuar a captura de 25 fotogramas, o equivalente a 1 segundo de vídeo, é aplicado o algoritmo *k-means clustering* a cada pixel da imagem, com $k = 5$, de modo a encontrar as classes que os descrevem na sequência dos últimos 25 fotogramas. Determinados os parâmetros μ_k e σ_k da gaussiana bivariada que define cada classe encontrada e o seu peso na mistura ω_k, os novos parâmetros são actualizados a cada novo fotograma, no instante de tempo t, a uma taxa $\alpha_t = 1/t$, de acordo com as equações seguintes:

$$\omega_{k,t}^{novo} = (1 - \alpha_t)\omega_{k,t} + \alpha_t P\left(k|X_t, \omega_{k,t}, \mu_{k,t}, \sigma_k\right)$$

(2.7.23)

$$\mu_{k,t}^{novo} = \left(1 - \rho_{k,t}\right)\mu_{k,t} + \rho_{k,t}X_t$$

(2.7.24)

$$\sigma_{k,t}^{novo} = \left(1 - \rho_{k,t}\right)\sigma_{k,t}^2 + \rho_{k,t}((X_t - \mu_{k,t}^{novo}) \circ (X_t - \mu_{k,t}^{novo}))$$

(2.7.25)

Onde

$$\rho_{k,t} = \frac{\alpha_t P\left(k|X_t, \omega_{k,t}, \mu_{k,t}, \sigma_k\right)}{\omega_{k,t}^{novo}}$$

(2.8.26)

A probabilidade *a posteriori*, $P\left(k|X_t, \omega_{k,t}, \mu_{k,t}, \sigma_k\right)$, é 1 caso o valor do novo pixel X_t esteja desviado da média da classe k um valor inferior a 2,5 vezes o desvio padrão dessa mesma classe e 0 caso contrário. Se existir mais do que uma classe para a qual a condição anterior é verdadeira, então é escolhida a classe com o maior valor $\omega_{k,t}/\sigma_{k,t}$. Organizando as classes por ordem decrescente de acordo com o valor $\omega_{k,t}/\sigma_{k,t}$, as que representam o *background* são assim as primeiras B que têm os maiores valores de ω_k, de acordo com a seguinte condição:

$$B = \underset{b}{\operatorname{argmin}}\left(\sum_{k=1}^{b} \omega_k > T\right)$$

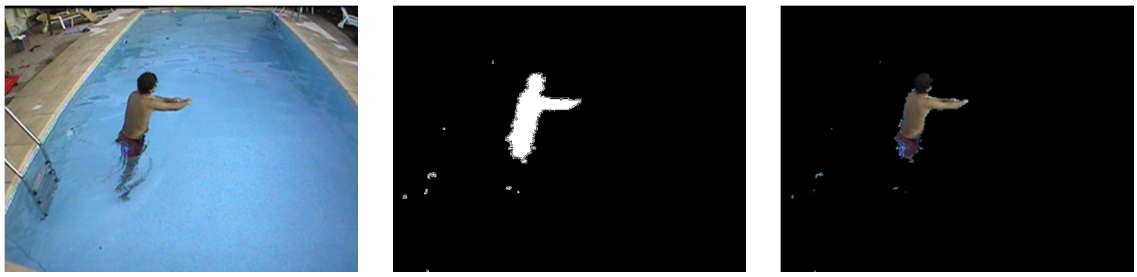
(2.7.27)

Deste modo, para cada novo fotograma, é verificada a classe à qual pertence cada pixel do mesmo, sendo considerado *background* ou *foreground* de acordo com a classificação que é dada à classe à qual este pertence.

2.7.2.3 Re-segmentação do mapa binário de *foreground*

O mapa binário gerado através da união dos mapas provenientes das zonas exterior e interior da piscina, embora seja imune ao ruído causado pelas reflexões especulares derivadas do movimento da água, não tem a precisão pretendida, tal como a Figura 2.7.4 pretende demonstrar. Na maioria dos casos as partes mais finas do corpo dos indivíduos são consideradas *background*, como por

exemplo as pernas e os braços, que muitas vezes estão submersos e são confundidos com a água no processo de segmentação proporcionado pelo algoritmo descrito na Secção 2.7.2.1.

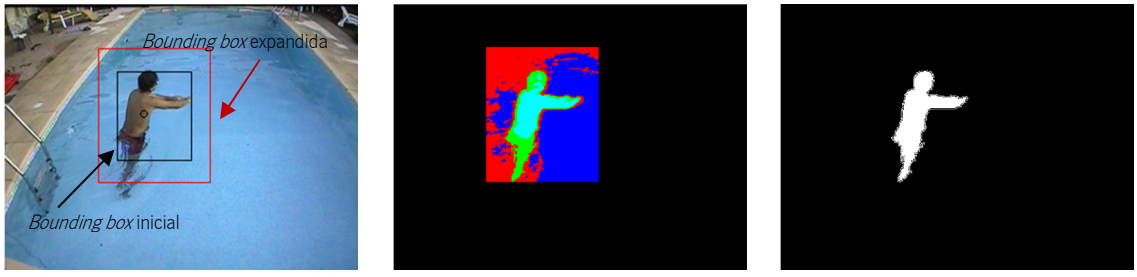


a) b) c)

Figura 2.7.4: O algoritmo de segmentação especialmente concebido para executar na área da piscina gera um mapa binário de *foreground* que não captura completamente o indivíduo. Isto deve-se ao facto de parte do corpo do mesmo estar submerso sendo confundido com a água e sendo assim classificado como *background*.

a) Imagem real;
b) As pernas do indivíduo não aparecem no mapa binário de *foreground*;
c) A sobreposição do mapa binário de *foreground* à imagem real mostra que as pernas do indivíduo foram consideradas *background*.

A forma encontrada para resolver o problema consiste na análise das regiões do mapa binário de *foreground* inicial, através da aplicação do algoritmo *connected components* (Horn B. , 1997) e efectuar a re-segmentação da área correspondente à *bounding box* expandida de cada região encontrada, tal como o esquema da Figura 2.7.6 evidencia.A expansão da *bounding box* permite capturar partes do corpo perdidas no processo inicial de segmentação, pois a parte correspondente ao *foreground* é tão significativa como a área correspondente ao *background*, Figura 2.7.5.



a) b) c)

Figura 2.7.5: Aplicação do algoritmo de re-segmentação na área da bounding box expandida de modo a aumentar a qualidade do mapa binário de foreground.

a) Expansão da *bounding box* na região detectada no mapa binário de *foreground* resultante da primeira fase de segmentação;
b) Resultado da aplicação do algoritmo de *clustering k-means*, no espaço de cores HSV, na área correspondente à *bounding box* expandida;
c) O mapa binário final de *foreground* contém as pernas do indivíduo perdidas na primeira fase de segmentação. Este mapa é gerado com base nas classes determinadas pelo algoritmo de *clustering* k-means.

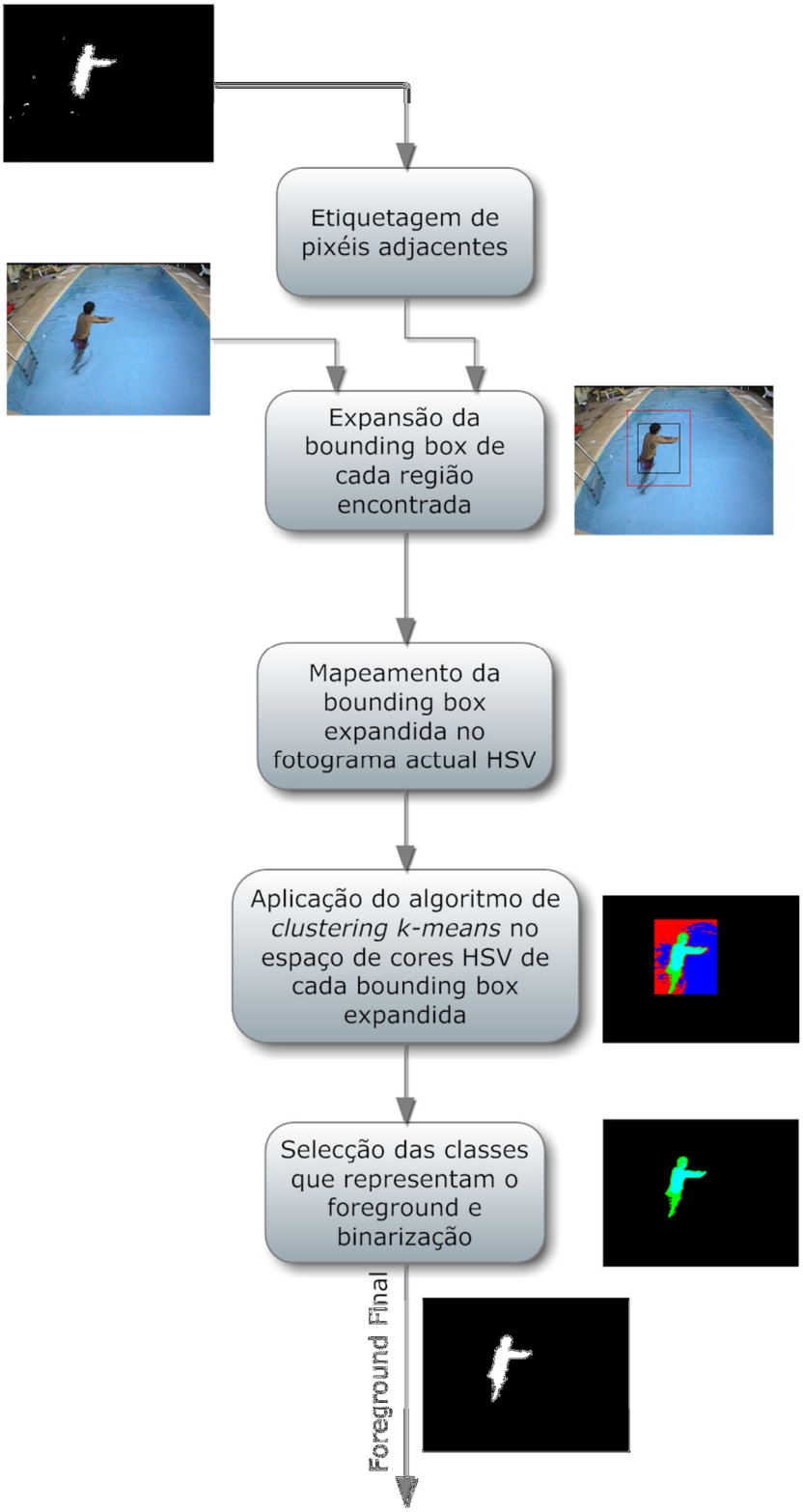


Figura 2.7.6: Esquema do algoritmo de re-segmentação. Os processos a partir da "Expansão das *bounding box*", até à "Selecção das classes pertencentes ao *foreground* ", inclusive, são efectuados para cada região descoberta pelo algoritmo de etiquetagem de pixéis adjacentes.

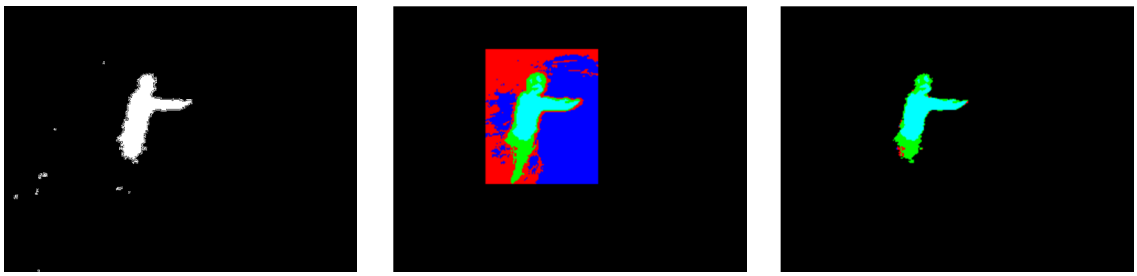
A aplicação do algoritmo de *clustering* k-means nas 3 dimensões do espaço de cores HSV permite aglomerar em 4 classes os vários pixéis na área expandida, sendo depois classificados de acordo com a percentagem localizada na região do objecto e nas bordas da *bounding box*. Os pixéis pertencentes às bordas da *bounding box* pertencem normalmente ao *background*, também porque apresentam um valor médio na componente H mais próximo do *background* global na área da piscina, Figura 2.7.7.

a) b)

Figura 2.7.7: Os pixéis pertencentes à borda da bounding box são contabilizados de modo a estabelecer um ranking das classes com maior número de pixéis nessa zona.

a) Resultado da aplicação do algoritmo de *clustering* k-means, no espaço de cores HSV, na área correspondente à *bounding box* expandida;
b) Classes dos pixéis pertencentes à borda da *bounding box* com elevada probabilidade de pertencerem ao *background*.

As classes de pixéis com maior número de incidências na região do objecto são classificadas como *foreground*, tal como a Figura 2.7.8 permite demonstrar.



a) b) c)

Figura 2.7.8: O mapeamento dos pixéis de cada classe pertencente à bounding box expandida no mapa binário de foreground inicial determina um ranking de classificação das classes mais prováveis de pertencerem ao verdadeiro foreground.

a) Mapa binário de *foreground* inicial, resultado do algoritmo apresentado na secção 2.7.2.1;
b) Resultado da aplicação do algoritmo de *clustering* k-*means*, no espaço de cores HSV, na área correspondente à *bounding box* expandida;
c) As classes que melhor representam o *foreground* são as representadas pelas cores verde e azul ciano.

Esta técnica permite aumentar a precisão da segmentação, capturando os pormenores perdidos na primeira fase. Além disso, devido à expansão da *bounding box* é possível aglomerar partes perdidas dos objectos separadas na primeira fase do processo de segmentação. Depois desse passo é executada nova análise das regiões dentro dessa área, através do algoritmo *connected components*, resultantes da segunda fase do processo de segmentação, sendo eliminadas as que pertencem ao *background* e que foram erradamente classificadas como *foreground* no processo. O mapa final de *foreground* é baseado na união das várias re-segmentações feitas a cada objecto e apresenta uma elevada qualidade e precisão, não sendo afectado pelo ruído causado pela oscilação da água à superfície, tal como se pode verificar pela análise da Figura 2.7.9.

a) b)

Figura 2.7.9: Comparação entre o mapa binário de foreground inicial e depois da re-segmentação. Houve um aumento efectivo da precisão do *foreground*, que é agora praticamente coincidente com o indivíduo, apanhando mesmo os pormenores, absolutamente necessários para o próximo estágio do *pipeline* de processamento. Além disso o nível de ruído foi também diminuído de forma significativa, reduzindo a percentagem de falsos positivos.

a) **Resultado do mapa binário de *foreground* correspondente à primeira fase de segmentação comparado directamente na imagem real;**
b) **Resultado do mapa binário de *foreground* correspondente à fase de re-segmentação comparado directamente na imagem real.**

2.8 Implementação

Todos os algoritmos desenvolvidos foram implementados no *Simulink*, uma das ferramentas que faz parte do *Matlab*. No Capítulo 5, mais concretamente na Secção 5.2, são fornecidos todos os detalhes acerca da implementação do sistema completo e o porquê da utilização destas ferramentas. Por agora apresenta-se a imagem do diagrama de blocos do algoritmo de detecção automática da piscina implementado no *Simulink*, Figura 2.8.1. No esquema da Figura 2.8.2 mostra-se a implementação do algoritmo de segmentação híbrido, capaz de extrair os mapas binários de *foreground* correspondentes aos objectos em movimento na cena.

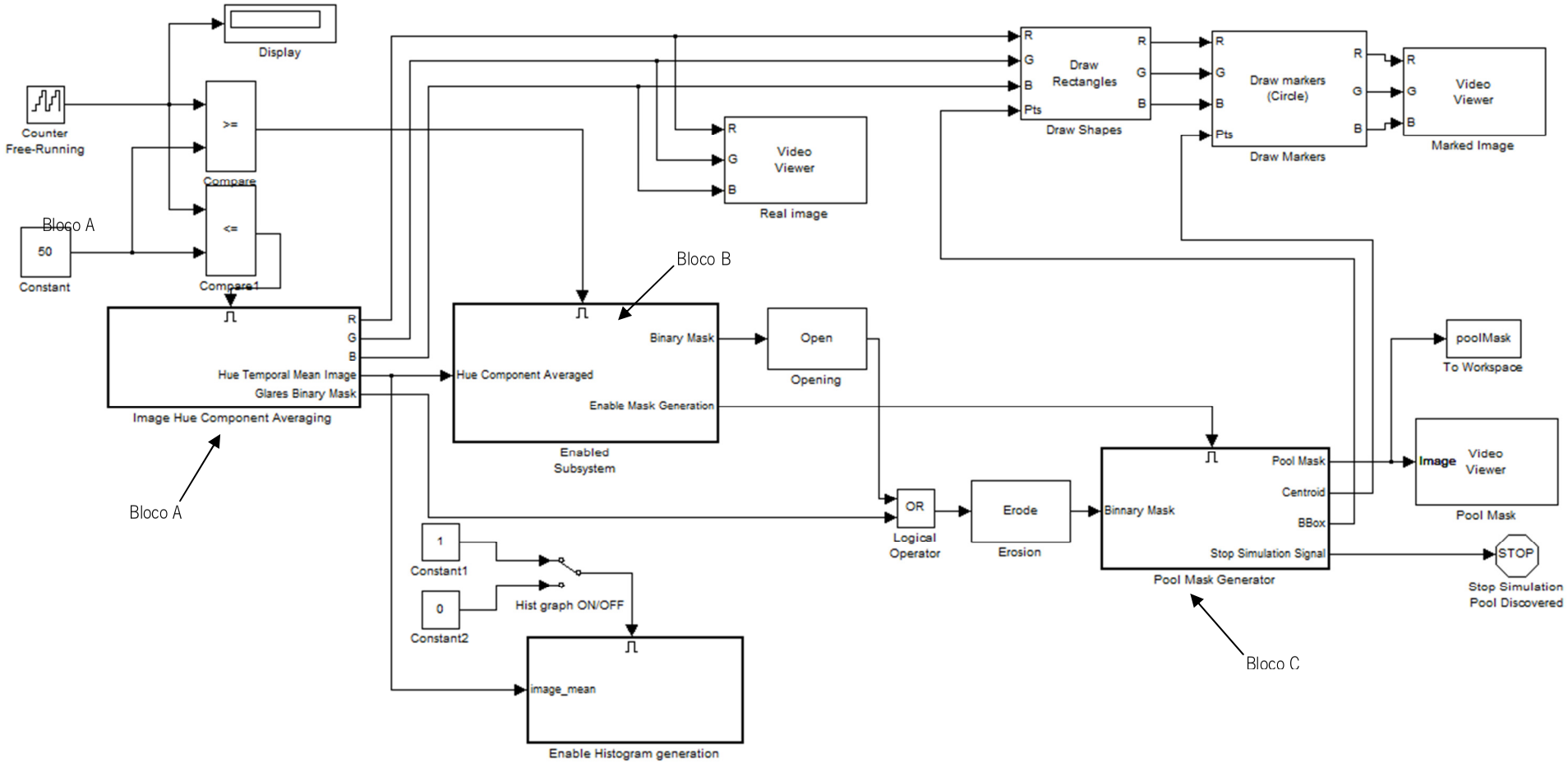


Figura 2.8.1: Algoritmo de detecção automática da piscina e geração da máscara binária correspondente à área da mesma. O bloco A implementa a reprodução do vídeo armazenado, a conversão RGB-HSV e efectua a média temporal da imagem na componente tonalidade. O bloco B executa o algoritmo EM que estima a mistura gaussiana a partir do histograma da média da imagem e implementa o *threshold*. Por último, o bloco C efectua a análise das regiões conjuntas de pixéis seleccionando a maior delas dentro dos parâmetros determinados no bloco B.

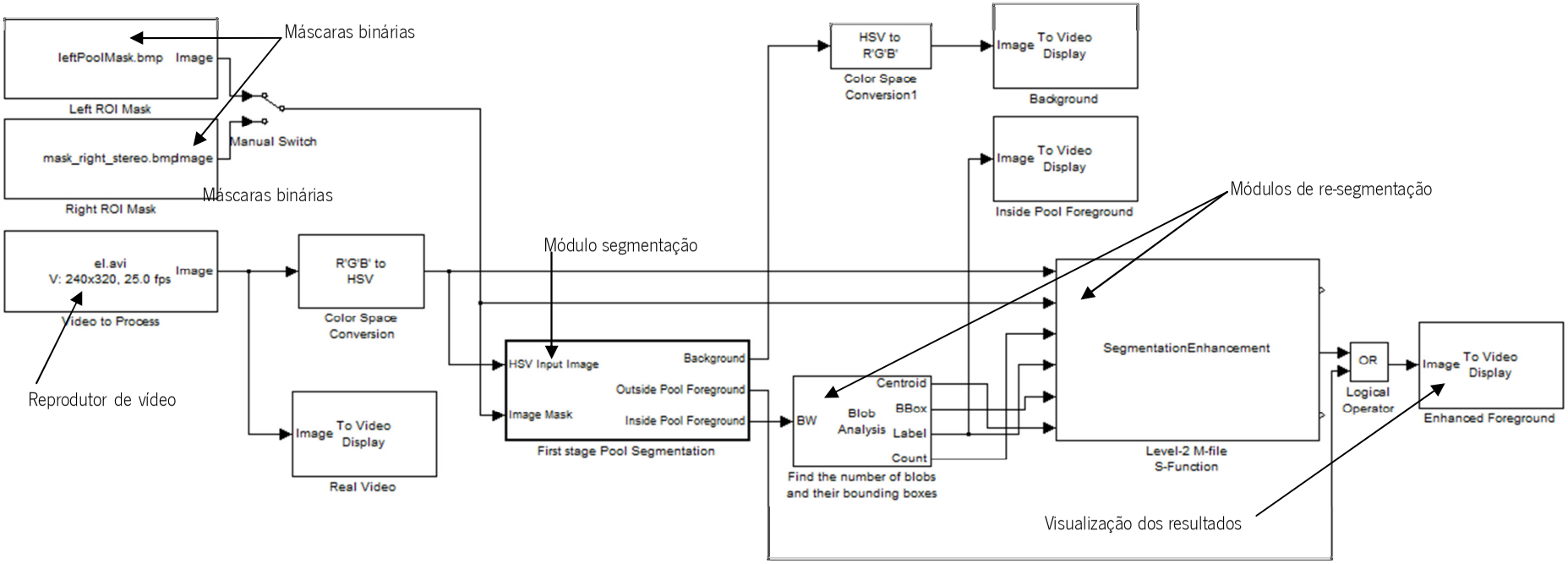


Figura 2.8.2: Algoritmo de segmentação de objectos em movimento na cena. O algoritmo é composto pelos módulos de segmentação na área da piscina, na área exterior à mesma e pelo módulo de re-segmentação. Os mapas binários de *foreground* resultantes da execução em cada módulo são combinados à saída e subsequentemente mostrados.

Esta abordagem de implementação permite provar rapidamente a funcionalidade do algoritmo, não existindo preocupação com a velocidade de processamento, fortemente afectada pelo *overhead* desta linguagem de alto nível. A implementação em "código m" de determinados blocos, tal como os da re-segmentação, permite a implementação de qualquer algoritmo constituindo este um módulo fechado, com interface de entrada e saída, diminuindo assim a complexidade das operações por disposição em camadas e de forma modular.

2.9 Resultados Experimentais

Medir a qualidade do processo de segmentação numa imagem tem sido, desde há muito tempo, alvo de intensa investigação por parte da comunidade científica, (Ge, Wang, & Liu, 2007). Existem métodos supervisionados, baseados no chamado *ground-truth* da imagem, uma referência segmentada manualmente por um ser humano. E existem métodos não supervisionados que não necessitam de referência, recorrendo a medições baseadas na homogeneidade das regiões segmentadas e das diferenças encontradas entre regiões vizinhas. Estes últimos, apesar de não serem tão eficazes, na medida em que, não existindo nenhum algoritmo genérico de segmentação, também não é possível avaliar a mesma de forma automática, são sobretudo mais rápidos e portanto exequíveis. Para isso repare-se que, a segmentação manual exige muito tempo para cada imagem, o que não confere um método viável para medição em tempo real da qualidade da segmentação. Além disso, de ser humano para ser humano, os critérios de avaliação podem mudar ligeiramente, uma vez que cada indivíduo tem os seus próprios standards de classificação, (Zhang, Fritts, & Goldman, 2008). No entanto, a metodologia supervisionada continua a ser a mais utilizada uma vez que confere um grau de crença elevado na imagem de referência como verdadeiro *foreground*, sendo por isso viável utilizar a mesma para efectuar comparações e extrair resultados credíveis de similaridade e, por conseguinte, da qualidade do processo de segmentação.

Para avaliar a qualidade do processo de segmentação do algoritmo híbrido para ambientes aquáticos complexos recorreu-se a um método supervisionado baseado na medição da discrepância entre o *foreground* gerado automaticamente e o *foreground* gerado manualmente para a mesma imagem. Deste modo, foram escolhidas algumas amostras de vídeos gravados numa piscina e procedeu-se à comparação visual das mesmas com a imagem de referência, segmentada manualmente e considerada o *ground-truth*. Além desta comparação visual, em diferentes situações que afectam negativamente a segmentação, foram extraídas 200 imagens consecutivas do vídeo capturado e segmentadas manualmente de modo a extrair dados concretos sobre a

quantidade de falsos negativos e falsos positivos do *foreground* gerado automaticamente em relação às imagens de referência. Estes dados são calculados de acordo com as equações 2.9.1 e 2.9.2, apresentadas nos trabalhos de (Butler, Bove Jr., & Sridharan, 2005).

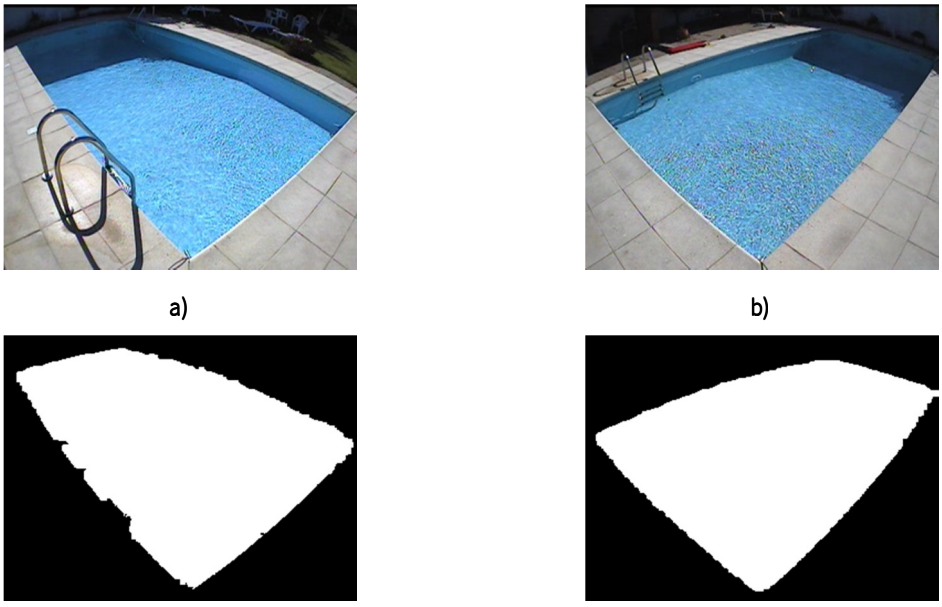
$$FAR(\%) = \frac{\#\ Falsos\ Positivos}{\#\ Pixeis\ pertencentes\ verdadeiro\ background} \tag{2.9.1}$$

$$FRR(\%) = \frac{\#\ Falsos\ Negativos}{\#\ Pixeis\ pertencentes\ verdadeiro\ foreground} \tag{2.9.2}$$

A primeira equação mede a percentagem de falsos positivos relativamente ao verdadeiro *background*, *False Acceptance Rate* (FAR). A segunda equação mede a percentagem de falsos negativos relativamente ao verdadeiro *foreground*, *False Rejection Rate* (FRR). Quanto mais baixos forem estes valores melhor será o processo de segmentação, relativamente à referência gerada manualmente.

2.9.1 Algoritmo de geração automática da máscara

Os resultados do algoritmo de geração automática da máscara da piscina são aqui apresentados em diferentes situações, não só da quantidade de luminosidade da fonte de luz sobre a cena, mas também do nível de oscilação da superfície da água e com diferentes pontos de vista na captura da cena utilizando várias câmaras. Na Figura 2.9.1 foram utilizadas imagens da piscina em condições de oscilação reduzida da superfície da água, existindo zonas de sombra, causadas pelas paredes da piscina. A aplicação do algoritmo de geração automática da máscara da piscina em apenas 6 segundos de vídeo permite gerar uma máscara que corresponde efectivamente à localização da área da piscina na cena. Já na imagem (a) da Figura 2.9.2 existe uma forte reflexão especular que também não afecta o funcionamento do algoritmo, sendo gerada uma máscara binária perfeitamente coincidente com a localização da piscina na cena, tal como se pode verificar ao analisar na imagem (c) da mesma figura. Neste caso a piscina apresenta uma oscilação quase inexistente, sendo neste caso também um benefício na segmentação da imagem.

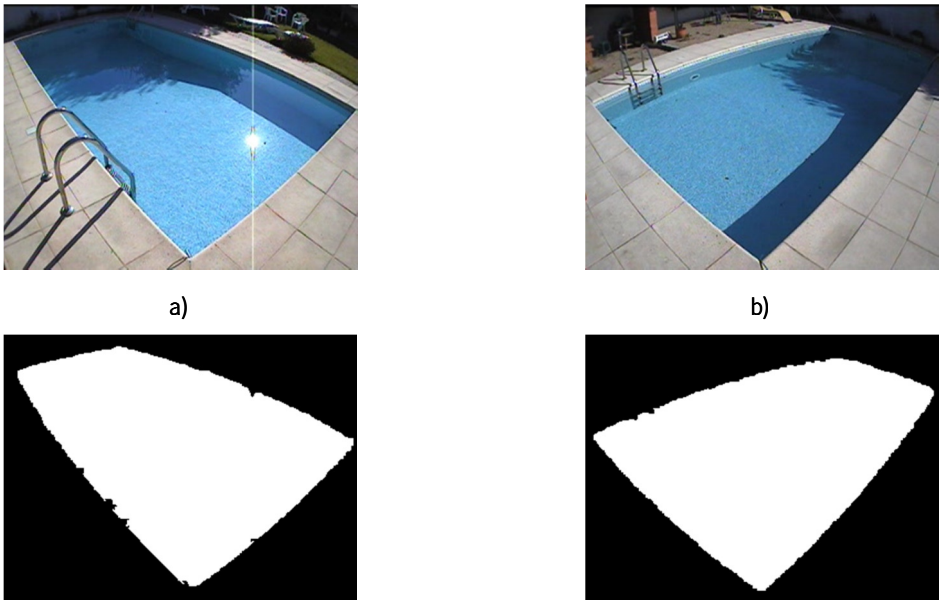


a) b)

c) d)

Figura 2.9.1: Geração da máscara binária com reduzida oscilação da água pouco depois do meio-dia. Repare-se na existência de sombra do lado da piscina oposto ao da localização das câmaras.

a) Imagem real da câmara exterior esquerda;
b) Imagem real da câmara exterior direita;
c) Mascara binária da piscina correspondente à câmara exterior esquerda;
d) Mascara binária da piscina correspondente à câmara exterior direita.



a) b)

c) d)

Figura 2.9.2: Geração da máscara binária com oscilação da água quase inexistente. Comporta-se como um espelho.

a) Imagem real da câmara exterior esquerda;
b) Imagem real da câmara exterior direita;
c) Mascara binária da piscina correspondente à câmara exterior esquerda;
d) Mascara binária da piscina correspondente à câmara exterior direita.

Um caso mais complexo é apresentado nas imagens (a) e (b) da Figura 2.9.3. Existem zonas de sombra e de luminosidade elevada, misturadas com elevada oscilação da superfície da água. Neste caso o algoritmo apresenta piores resultados que nos casos anteriores. Ainda assim não é crítico a máscara sair para além da área da piscina. Crítico seria se a máscara não contivesse toda a área da piscina, ficando o algoritmo de segmentação exterior a executar na área desta, não estando preparado para a complexidade do *background* podendo originar falsos positivos nessas zonas, de acordo com o nível de oscilação da superfície da água.

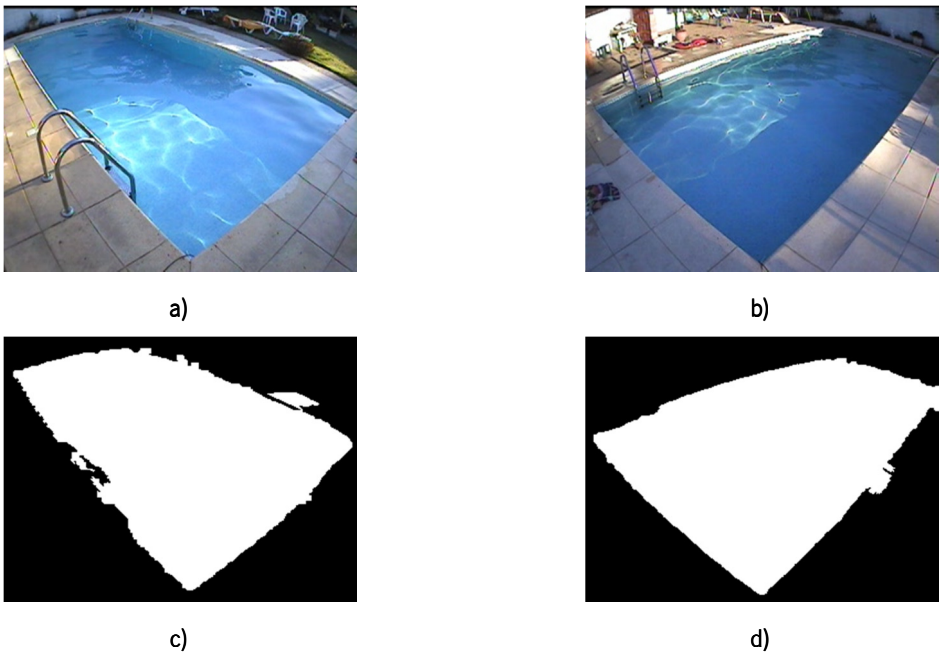
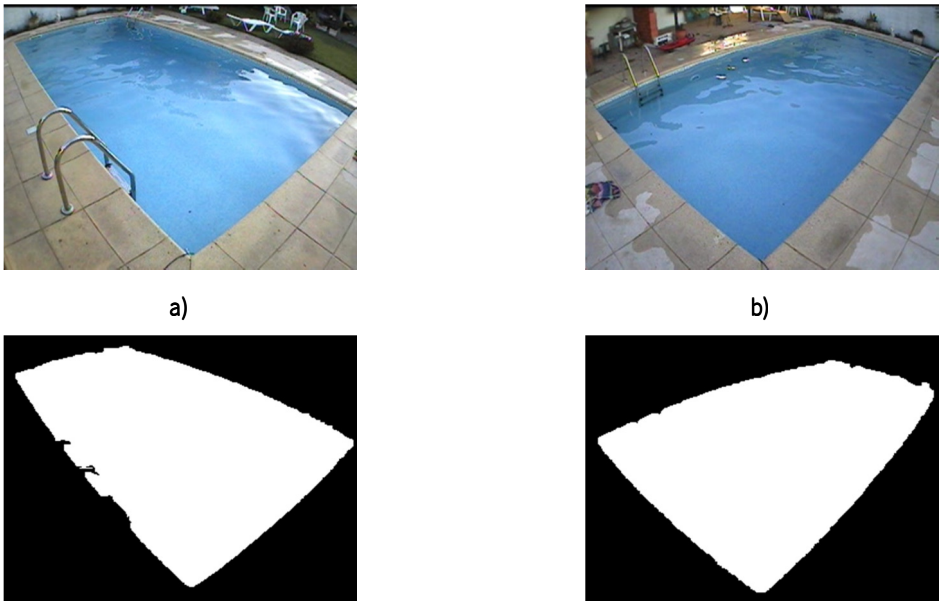


Figura 2.9.3: Geração da máscara binária com oscilação elevada da água à superfície. A existência de sombras e zonas de reflexão da luz solar torna a área da piscina altamente variável, daí que este processo tenha sido menos eficaz nestas condições.

a) Imagem real da câmara exterior esquerda;
b) Imagem real da câmara exterior direita;
c) Mascara binária da piscina correspondente à câmara exterior esquerda;
d) Mascara binária da piscina correspondente à câmara exterior direita.

Os melhores resultados alcançados pelo algoritmo remetem naturalmente para cenas onde a oscilação da água seja reduzida, onde a fonte de luz não incida directamente na mesma, sendo esta muito mais homogénea sem sofrer perturbações de sombras e reflexões especulares. Nas imagens (c) e (d) da Figura 2.9.4 podem ser observados os resultados da execução do algoritmo nestas condições. São óptimos resultados, pois a mascara binária corresponde perfeitamente à localização da área da piscina na imagem.

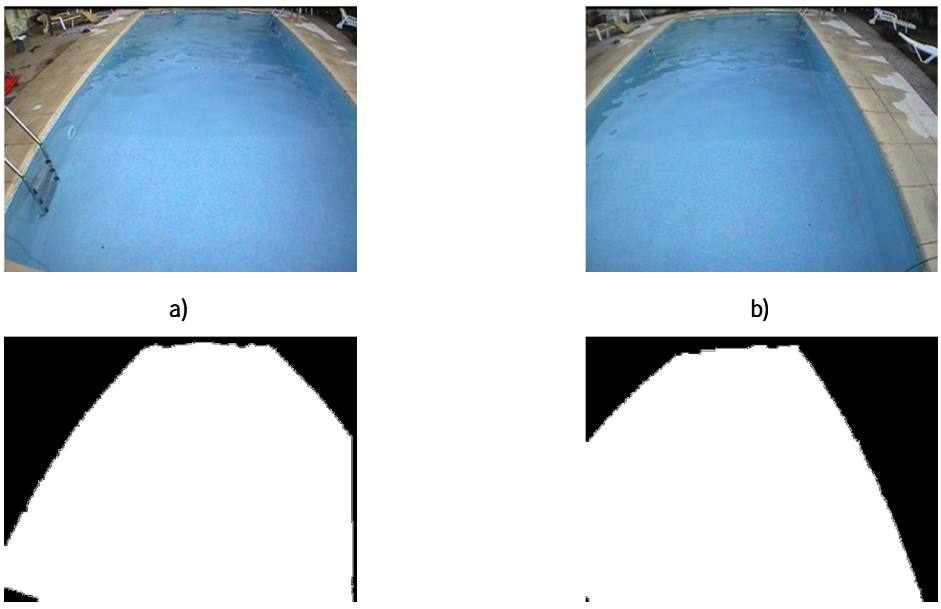


a) b)

c) d)

Figura 2.9.4: Geração da máscara binária com alguma oscilação da superfície da água. Nesta situação a fonte de luz tem menor intensidade e não se encontra definida num único ponto.

a) Imagem real da câmara exterior esquerda;
b) Imagem real da câmara exterior direita;
c) Mascara binária da piscina correspondente à câmara exterior esquerda;
d) Mascara binária da piscina correspondente à câmara exterior direita.



a) b)

c) d)

Figura 2.9.5: Geração da máscara binária numa disposição *stereo* da localização das câmaras.

a) Imagem real da câmara exterior esquerda;
b) Imagem real da câmara exterior direita;
c) Mascara binária da piscina correspondente à câmara exterior esquerda;
d) Mascara binária da piscina correspondente à câmara exterior direita.

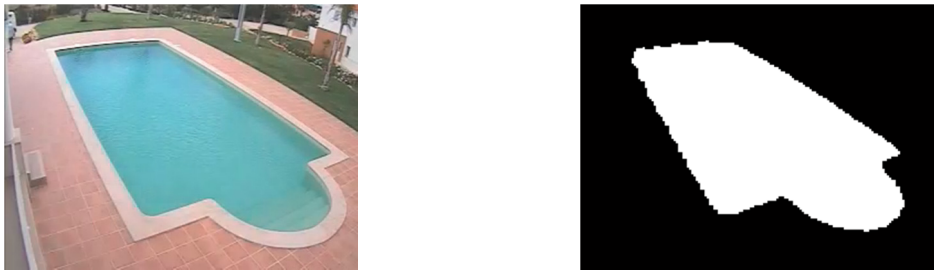
Nas mesmas condições das imagens da figura anterior, foram capturadas imagens das câmaras centrais na disposição *stereo*, que podem ser vistas na Figura 2.9.5. Tal como no caso anterior, os resultados da segmentação são excelentes, mostrando que a localização da câmara não interfere com os resultados da execução do algoritmo, desde que os requisitos impostos sejam cumpridos, nomeadamente a ocupação por parte da piscina da maior área da imagem. Com a câmara central, Figura 2.9.6, obtêm-se resultados também excelentes, mesmo em condições de oscilação da água à superfície e com a existência de sombras e reflexões. Repare-se nas diferenças significativas na tonalidade da cor da piscina na imagem (a) da Figura 2.9.5 e na imagem (b) da Figura 2.9.6.

a) b)

Figura 2.9.6: Aplicação do algoritmo à imagem captada pela câmara central.

a) Imagem real;
b) Máscara binária da piscina em (a).

Na Figura 2.9.7 podem ser vistos os resultados da aplicação do algoritmo numa piscina com um formato diferente das anteriores. À semelhança dos outros casos a máscara gerada corresponde perfeitamente à localização da piscina.



a) b)

Figura 2.9.7: Aplicação do algoritmo em outra piscina com formato diferente da anterior.

a) Imagem real;
b) Máscara binária da piscina em (a).

Graças à utilização do algoritmo EM sobre o histograma na componente tonalidade é possível determinar as classes que representam a piscina na imagem global da cena, com elevada imunidade aos ruídos causados pela oscilação da superfície da água, pelas sombras e reflexões especulares em diferentes horas do dia e em diferentes condições de luminosidade e com localizações diferentes do dispositivo de captura.

2.9.2 Algoritmo de segmentação híbrido

Os resultados obtidos na segunda fase de segmentação a cada objecto detectado na primeira fase são aqui apresentados de modo a verificar o aumento da qualidade do processo, em termos gerais. Através da análise da Figura 2.9.8 é possível verificar que, a re-segmentação aumentou a qualidade do processo de segmentação, sendo o mapa binário observado na imagem (c) muito mais semelhante ao apresentado na imagem (d), quando comparado com o da imagem (b). Com o aumento do tamanho da *bounding box*, na região ocupada pelo objecto e a consequente segmentação através da aplicação do algoritmo *k-means* foi possível capturar mais pixéis pertencentes ao *foreground* que tinham sido descartados no primeiro passo de segmentação, em virtude de eliminar o ruído causado pela oscilação da superfície da água.

a) b) c) d)

Figura 2.9.8: Comparação entre os resultados obtidos sem re-segmentação e com re-segmentação relativamente à imagem de referência gerada manualmente.

a) Imagem real;
b) Mapa binário de *foreground* gerado antes do processo de re-segmentação com recurso ao algoritmo *k-means*;
c) Mapa binário de *foreground* final, depois do processo de re-segmentação;
d) Mapa binário de *foreground* de referência.

O mesmo pode ser visto nas imagens da Figura 2.9.9. Repare-se que, após a re-segmentação aumentou o número de pixéis pertencentes à região de *foreground*, sendo mais uma vez a similaridade do mapa binário da figura (c) mais próxima da imagem de referência (d) do que o mapa binário da figura (b).

a) b) c) d)

Figura 2.9.9: Comparação entre os resultados obtidos sem re-segmentação e com re-segmentação relativamente à imagem de referência gerada manualmente.

a) Imagem real;
b) Mapa binário de *foreground* gerado antes do processo de re-segmentação com recurso ao algoritmo *k-means*;
c) Mapa binário de *foreground* final, depois do processo de re-segmentação;
d) Mapa binário de *foreground* de referência.

Contudo, estas são amostras que não justificam claramente a conclusão de que um segundo passo de segmentação melhora efectivamente a qualidade do processo, uma vez que são insuficientes e não contêm uma vasta variedade de casos. Para isso foram utilizados como amostra 100 fotogramas consecutivos, pertencentes aos vídeos de teste na situação evidenciada pelas imagens (a) da Figura 2.9.8 e da Figura 2.9.9. Como a captura foi efectuada a uma velocidade de 25 fotogramas por segundo, esta quantidade de imagens corresponde exactamente a 4 segundos de vídeo. Para cada um destes fotogramas foram calculados os valores de FAR e FRR, de acordo com as equações 2.9.1 e 2.9.2, respectivamente.

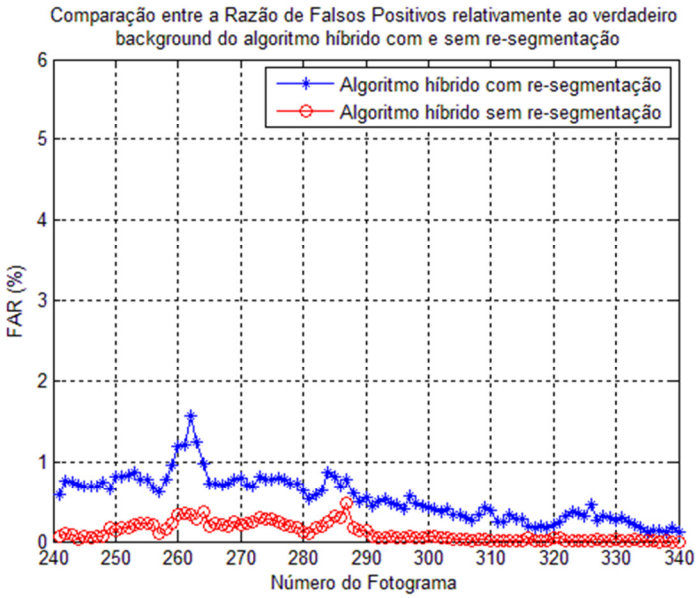


Figura 2.9.10: Comparação da percentagem de falsos positivos relativamente ao verdadeiro background do algoritmo híbrido de segmentação com um só passo e com re-segmentação por aplicação do algoritmo *k-means*, num conjunto de 100 fotogramas consecutivos.

A análise do gráfico da Figura 2.9.10 mostra que, ao longo dos 100 fotogramas consecutivos, o número de falsos positivos aumentou quase meio ponto percentual com a introdução da re-segmentação quando comparado com o algoritmo sem re-segmentação. Isto deve-se ao facto do segundo passo de segmentação, por aumentar o número de pixéis pertencentes ao *foreground*, introduzir também mais falsos positivos, o que é normal, face ao aumento brutal de verdadeiros positivos que ocorre, como se pode constatar através do gráfico da Figura 2.9.11.

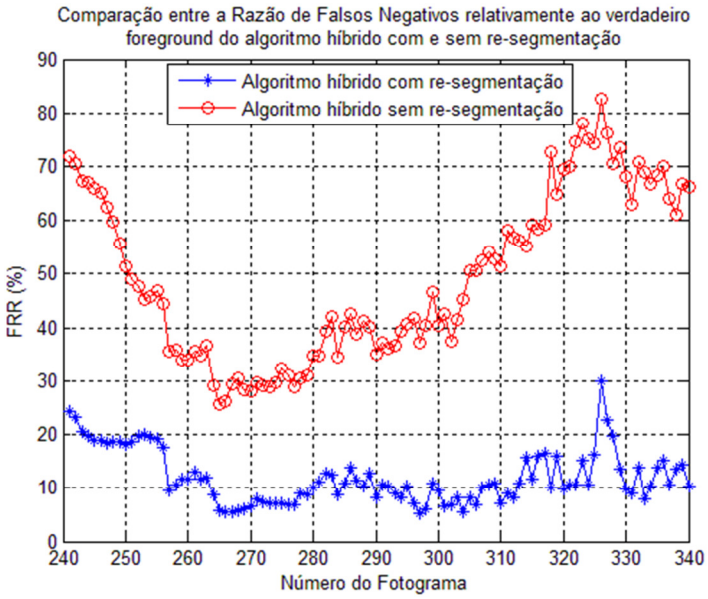
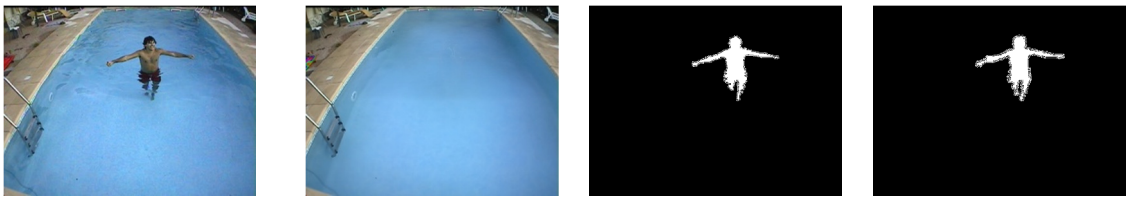


Figura 2.9.11: Comparação da percentagem de falsos negativos relativamente ao verdadeiro foreground do algoritmo híbrido de segmentação com um só passo e com re-segmentação por aplicação do algoritmo *k-means*, num conjunto de 100 fotogramas consecutivos.

Repare-se na diferença das duas curvas do gráfico da Figura 2.9.11 relativamente ao número de falsos negativos. Com a aplicação da re-segmentação o número de falsos negativos diminuiu drasticamente, apontando para uma maior precisão do mapa binário de *foreground*, já que muitos dos pixéis marcados como *background* eram na realidade *foreground*. Esta diferença mostra claramente uma maior correspondência entre o mapa binário de *foreground* e a imagem de referência, ainda que, o número de falsos positivos tenha aumentado. Na verdade o aumento de falsos positivos é insignificante quando comparado com o ganho obtido no aumento do número de verdadeiros positivos.

De modo a observar a qualidade da segmentação efectuada pelo algoritmo híbrido, já com a re-segmentação por objectos, foram capturadas várias amostras da imagem de *background* e do mapa binário de *foreground*, bem como da imagem de referência segmentada manualmente de modo a comparar visualmente as mesmas, em várias situações distintas. Na Figura 2.9.12, que se segue, a amostra extraída corresponde a uma situação mais favorável em termos de reflexões especulares sendo notório uma clara similaridade, constatada visualmente, entre o mapa binário de *foreground* (c) e a imagem de referência (d).

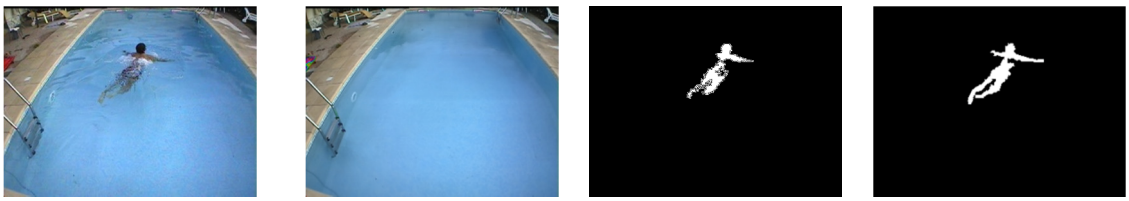


a) b) c) d)

Figura 2.9.12: Comparação entre a imagem capturada e o background estimado e entre o mapa binário de foreground e o mapa binário de referência, numa cena minimamente afectada por reflexões especulares.

a) Imagem real;
b) *Background* estimado;
c) Mapa binário de *foreground* gerado pelo algoritmo de segmentação;
d) Mapa binário de *foreground* gerado através de segmentação manual.

Já na imagem (a) da Figura 2.9.13, a cena, apesar de apresentar boas condições de iluminação, no que respeita a reflexões especulares e sombras, contém um nadador que causa algumas bolhas de ar e salpicos, devido à movimentação do seu corpo na água. No entanto, neste caso esses salpicos são filtrados pelo algoritmo de segmentação, o que permite obter um mapa binário de *foreground*, imagem (c), tão semelhante ao mapa binário de referência, imagem (d).

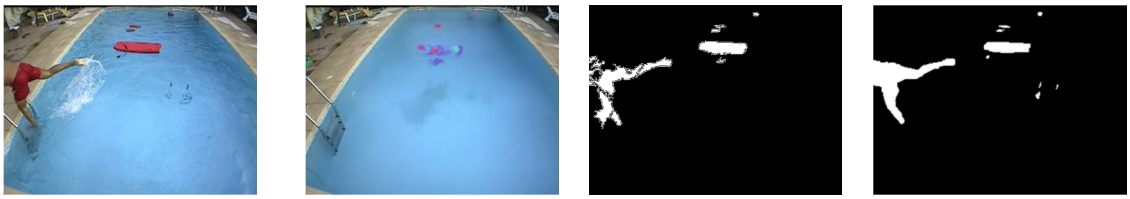


a) b) c) d)

Figura 2.9.13: Comparação entre a imagem capturada e o *background* estimado e entre o mapa binário de *foreground* e o mapa binário de referência, numa cena minimamente afectada por reflexões especulares, mas com existência de salpicos e bolhas causados pelo nadador.

a) Imagem real;
b) *Background* estimado;
c) Mapa binário de *foreground* gerado pelo algoritmo de segmentação;
d) Mapa binário de *foreground* gerado através de segmentação manual.

Situação semelhante ocorre na amostra da Figura 2.9.14. Repare-se que, mais uma vez, os salpicos são completamente filtrados. No entanto, existem objectos, que por serem tão pequenos, não são considerados como pertencentes ao *foreground*. Outro ponto interessante remete para o modelo do *background*, imagem (b), que apresenta sinais de absorção dos objectos que estão parados há muito tempo na superfície da água. Sempre que uma região de *foreground* aparece muito tempo na mesma posição sem variações significativas na sua área e forma, a mesma passa a fazer parte do modelo de *background*. Repare-se também que, a parte do corpo do indivíduo fora da área ocupada pela máscara correspondente à localização da piscina não apresenta tanta qualidade, no que respeita ao mapa binário de *foreground* (c), quando comparado com a imagem de referência (d). Este facto não é muito relevante na medida em que a acção mais importante da cena se passa no interior da piscina, pelo que, a qualidade da segmentação deverá estar orientada para obter valores máximos nessa área.

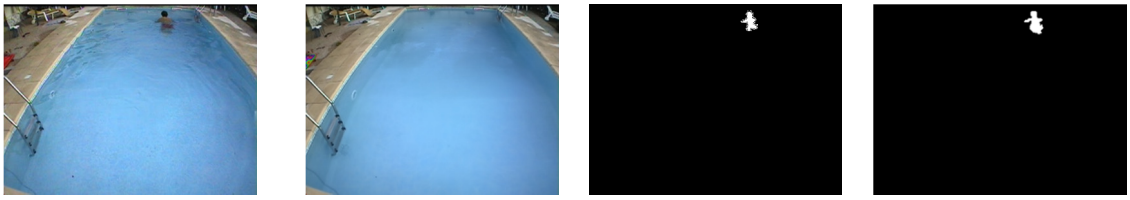


a) b) c) d)

Figura 2.9.14: Comparação entre a imagem capturada e o background estimado e entre o mapa binário de *foreground* e o mapa binário de referência, numa cena minimamente afectada por reflexões especulares, mas com existência de salpicos causados pelo individuo e vários objectos na superfície da água.

a) Imagem real;
b) *Background* estimado;
c) Mapa binário de *foreground* gerado pelo algoritmo de segmentação;
d) Mapa binário de *foreground* gerado através de segmentação manual.

Em condições muito semelhantes, a Figura 2.9.15 pretende demonstrar que, mesmo quando os objectos estão bastante longe da câmara e, por conseguinte, a quantidade de pixéis que os representam são menores, o algoritmo consegue eficazmente marcar no mapa binário de *foreground* os mesmos com elevada qualidade. Para isso compare-se visualmente as imagens (c) e (d) da referida figura. A similaridade entre as duas é bastante elevada, apesar do tamanho da região segmentada.

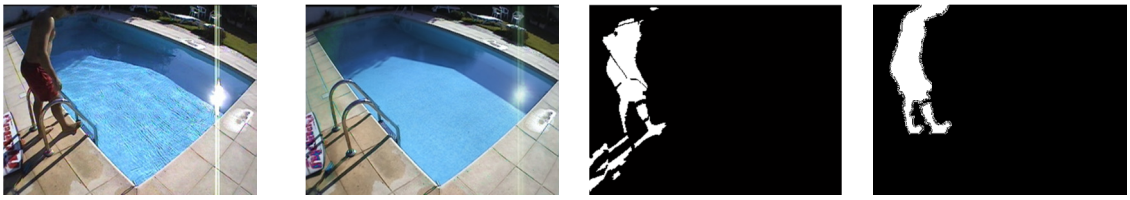


a) b) c) d)

Figura 2.9.15: Comparação entre a imagem capturada e o *background* estimado e entre o mapa binário de *foreground* e o mapa binário de referência, numa cena minimamente afectada por reflexões especulares e com o nadador muito afastado da localização da câmara.

a) Imagem real;
b) *Background* estimado;
c) Mapa binário de *foreground* gerado pelo algoritmo de segmentação;
d) Mapa binário de *foreground* gerado através de segmentação manual.

Na Figura 2.9.16 apresenta-se uma amostra em que as condições de iluminação são mais adversas, uma vez que já existem fortes reflexões especulares na superfície da água e sombras, causadas não só pelas paredes da piscina, mas também pelo indivíduo em movimento. Repare-se que o algoritmo conseguiu eliminar eficazmente a reflexão especular, até porque a mesma começa a fazer parte do modelo de *background*. No entanto, nota-se uma pequena zona de quebra do mapa binário de *foreground* na zona correspondente à fronteira da máscara da localização da piscina. A presença da sombra do indivíduo na imagem (c) é também um ponto negativo no que respeita à qualidade do algoritmo de segmentação, contribuindo para o aumento dos falsos positivos.

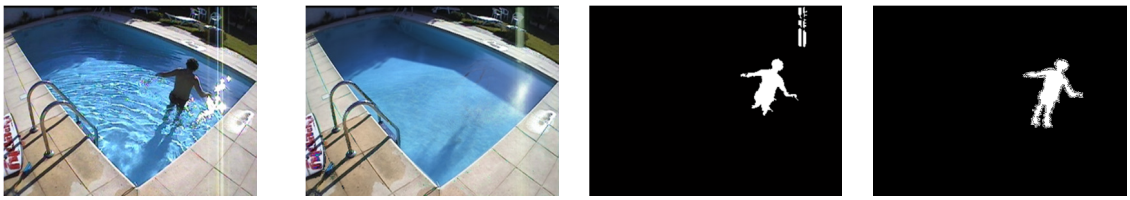


a) b) c) d)

Figura 2.9.16: Comparação entre a imagem capturada e o *background* estimado e entre o mapa binário de *foreground* e o mapa binário de referência, numa cena afectada por uma reflexão especular muito intensa, capaz de causar a saturação do CCD.

a) Imagem real;
b) *Background* estimado;
c) Mapa binário de *foreground* gerado pelo algoritmo de segmentação;
d) Mapa binário de *foreground* gerado através de segmentação manual.

Uma situação muito semelhante à anteriormente descrita, mas neste caso com o indivíduo totalmente no interior da piscina e, por conseguinte, com um nível de oscilação da superfície da água muito maior, é apresentada na Figura 2.9.17. Repare-se na qualidade da segmentação da região correspondente ao indivíduo, quando comparada com a imagem de referência. É no entanto de notar o aparecimento de uma região considerada *foreground*, originada por uma saturação do sensor CCD da câmara, mas apenas na área exterior à localização da piscina. Isto deve-se ao facto de o algoritmo que processa nessa área ser menos tolerante a reflexões especulares e sombras, ao contrário do algoritmo específico para ambientes aquáticos complexos.

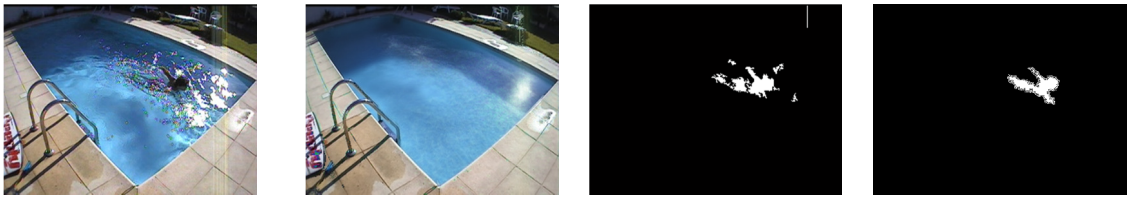


a) b) c) d)

Figura 2.9.17: Comparação entre a imagem capturada e o *background* estimado e entre o mapa binário de *foreground* e o mapa binário de referência, numa cena afectada por várias reflexões especulares de elevada intensidade causando a saturação do CCD.

a) **Imagem real;**
b) ***Background* estimado;**
c) **Mapa binário de *foreground* gerado pelo algoritmo de segmentação;**
d) **Mapa binário de *foreground* gerado através de segmentação manual.**

O caso apresentado na Figura 2.9.18 é muito grave, na medida em que é afectado por inúmeras reflexões especulares, por salpicos e bolhas de ar causadas pelo movimento do nadador e por um elevado ruído proveniente do sensor CCD da câmara na zona envolvente do corpo do indivíduo, devido às pequenas reflexões especulares. Atente-se na imagem (a), onde é notória a falta de visibilidade, mesmo para um ser humano, da parte do corpo submersa. Ainda assim, o algoritmo confere um resultado aceitável, uma vez que deixa algumas reflexões especulares serem consideradas *foreground*, aumentando a quantidade de falsos positivos. O ruído em volta do corpo do nadador só desaparece porque constitui regiões demasiado pequenas para serem consideradas *foreground*. No entanto, os pixéis próximos do corpo que sejam ruído são considerados *foreground*, o que não é crítico. Na Figura 2.9.19 a situação é análoga à anterior, mas neste caso não existe nenhum indivíduo na piscina, apenas objectos. A oscilação da água à superfície apresenta baixa amplitude, mas uma elevada frequência, o que faz com que apareça uma mancha enorme causada pelas reflexões especulares.

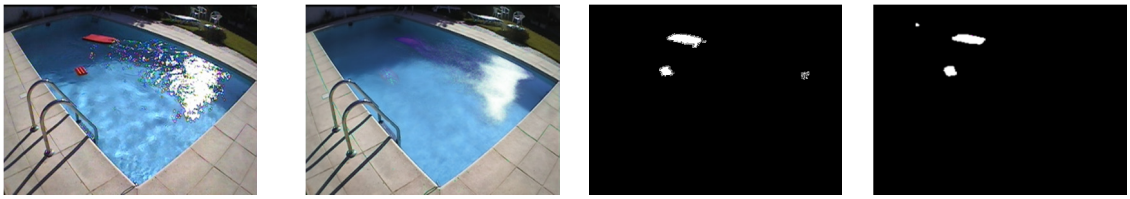


a) b) c) d)

Figura 2.9.18: Comparação entre a imagem capturada e o *background* estimado e entre o mapa binário de *foreground* e o mapa binário de referência, numa cena afectada por uma reflexão especular muito intensa, saturação do CCD em vários pontos originando muito ruído e escondendo parcialmente o nadador.

a) Imagem real.
b) *Background* estimado.
c) Mapa binário de *foreground* gerado pelo algoritmo de segmentação.
d) Mapa binário de *foreground* gerado através de segmentação manual.

Na imagem (b) desta figura pode notar-se que o modelo de *background* já absorveu essa mesma mancha e que o ruído existente entre o objecto e a mesma também não é considerado *foreground* porque não apresenta tamanho suficiente para tal. No entanto, esporadicamente, aparecem pequenas regiões pertencentes ao *foreground*, tal como a que pode ser vista na imagem (c), na zona da mancha, que não são relevantes no processo de seguimento, uma vez que aparecem e logo desaparecem. A comparação dos resultados da segmentação, imagem (c), com a imagem de referência em (d) mostra um grau de similaridade elevado.



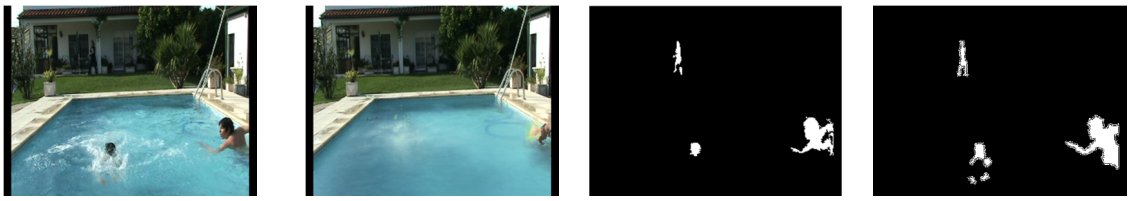
a) b) c) d)

Figura 2.9.19: Comparação entre a imagem capturadae o *background* estimado e entre o mapa binário de *foreground* e o mapa binário de referência, numa cena afectada por uma reflexão especular muito intensa e espalhada pela superfície da água ocupando uma área significativa.

a) Imagem real;
b) *Background* estimado;
c) Mapa binário de *foreground* gerado pelo algoritmo de segmentação;
d) Mapa binário de *foreground* gerado através de segmentação manual.

A amostra da Figura 2.9.20 foi capturada por uma câmara de vídeo de elevada qualidade e resolução, tendo também uma localização diferente da câmara utilizada nas amostras anteriores. Na cena pode ser vista uma elevada agitação da água causada por um dos indivíduos. O outro

indivíduo começa já a ser absorvido pelo modelo de *background*, sendo no entanto eficazmente assinalado no mapa binário de *foreground*. Repare-se ainda numa terceira pessoa, no exterior da piscina, que apesar de bastante longe do dispositivo de captura é bem segmentada e diferenciada do *background*. As similaridades do mapa binário de *foreground* (c), gerado pelo algoritmo híbrido de segmentação, com a imagem de referência (d), são notórias.

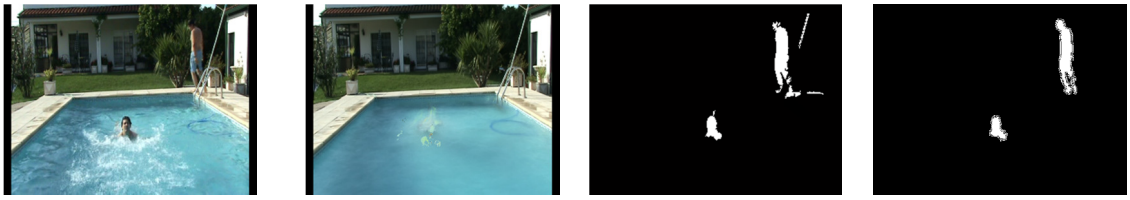


a) b) c) d)

Figura 2.9.20: Comparação entre a imagem capturada e o *background* estimado e entre o mapa binário de *foreground* e o mapa binário de referência, numa cena caracterizada pelos salpicos e bolhas de ar presentes à volta de um dos indivíduos.

a) Imagem real;
b) *Background* estimado;
c) Mapa binário de *foreground* gerado pelo algoritmo de segmentação;
d) Mapa binário de *foreground* gerado através de segmentação manual.

Na Figura 2.9.21 a situação é análoga à anterior, mas com uma quantidade superior de bolhas de ar e salpicos junto do nadador. Todos esses ruídos são filtrados pelo algoritmo, sendo no entanto de assinalar a presença da sombra da pessoa no exterior da piscina e de parte da vara no mapa binário de *foreground*.

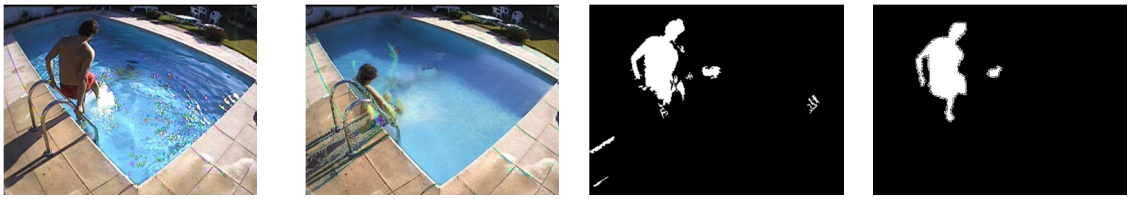


a) b) c) d)

Figura 2.9.21: Comparação entre a imagem capturada e o *background* estimado e entre o mapa binário de *foreground* e o mapa binário de referência, numa cena marcada pela elevada agitação da água em volta do nadador, causando vários salpicos e bolhas de ar.

a) Imagem real;
b) *Background* estimado;
c) Mapa binário de *foreground* gerado pelo algoritmo de segmentação;
d) Mapa binário de *foreground* gerado através de segmentação manual.

Um dos casos problemáticos referido no início deste capítulo apontava para o facto da oscilação da superfície da água encobrir parcialmente ou totalmente um corpo submerso. Na Figura 2.9.22 apresenta-se uma amostra com esta situação muito particular. Repare-se que, o indivíduo submerso é pouco perceptível, mesmo para um ser humano. No entanto, o algoritmo de segmentação consegue detectá-lo e marcá-lo como pertencente ao mapa binário de *foreground*. Mesmo a segmentação manual confere uma região igualmente diminuta ao corpo submerso. Esta corresponde à situação limite de utilização de câmaras exteriores para a tarefa de detecção de objectos na piscina. Com uma oscilação mais elevada e a uma profundidade pouco superior a 2 m já não é possível detectar o corpo submerso, pois não se vê de modo nenhum, sendo necessário recorrer a câmaras subaquáticas para complementar a segmentação.



a) b) c) d)

Figura 2.9.22: Comparação entre a imagem capturada e o *background* estimado e entre o mapa binário de *foreground* e o mapa binário de referência, numa cena afectada por reflexões especulares e salpicos, com a superfície da água fortemente agitada causando o desaparecimento do indivíduo submerso.

a) Imagem real;
b) *Background* estimado;
c) Mapa binário de *foreground* gerado pelo algoritmo de segmentação;
d) Mapa binário de *foreground* gerado através de segmentação manual.

Por último, um caso igualmente importante tem a ver com o funcionamento do algoritmo durante o período nocturno, Figura 2.9.23. Com uma ausência total de luz não seria possível capturar a imagem, mas como a maioria das câmaras de vigilância vêm equipadas com LED's infravermelhos é possível iluminar a cena com luz num comprimento de onda capaz de ser capturado pelo CCD da câmara. Deste modo, resulta uma imagem semelhante à obtida numa escala de cinzentos, capaz de ser igualmente segmentada, ainda que com menor eficácia, quando comparada com os casos anteriores.

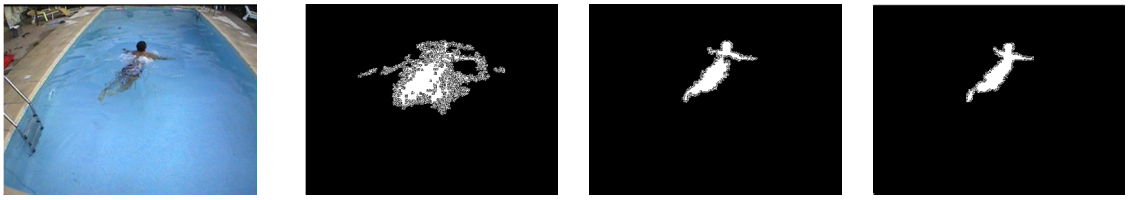
De modo a avaliar os resultados obtidos pelo algoritmo híbrido de segmentação de movimento foram efectuadas segmentações, das mesmas sequências de imagens, através do algoritmo MoG, tal como descrito nos trabalhos de (C. Stauffer, 2000).

a) b) c) d)

Figura 2.9.23: Comparação entre a imagem capturada e o *background* estimado e entre o mapa binário de *foreground* e o mapa binário de referência, numa cena iluminada por luz infravermelha, uma vez que foi capturada durante o período nocturno.

a) Imagem real;
b) *Background* estimado;
c) Mapa binário de *foreground* gerado pelo algoritmo de segmentação;
d) Mapa binário de *foreground* gerado através de segmentação manual.

Na Figura 2.9.24 pode-se observar uma amostra segmentada com os dois algoritmos e o respectivo mapa binário de referência, gerado através de segmentação manual. Apesar de este ser um dos casos mais favoráveis, uma vez que não existem sombras nem reflexões especulares, o movimento da água causa, ainda assim, o aparecimento de uma enorme quantidade de falsos positivos. O algoritmo MoG foi afinado de modo a obter a maior precisão possível com a geração do menor número de falsos positivos. No entanto, é extremamente difícil obter uma boa precisão, ou seja, segmentar correctamente o indivíduo na piscina mantendo um baixo valor de falsos positivos. O verdadeiro problema do algoritmo reside na taxa de aprendizagem do fundo, uma vez que, sendo esta muito rápida, o objecto desaparece muito rapidamente do mapa binário de *foreground*, apesar do movimento da água ser bastante atenuado. Por outro lado, taxas de aprendizagem muito baixas, apesar de segmentarem correctamente os objectos, não são imunes ao movimento da água, uma vez que essas zonas são consideradas como pertencentes ao *foreground*. Avaliando a quantidade de falsos positivos relativamente ao verdadeiro *background* é possível verificar, tal como o gráfico da Figura 2.9.25 mostra, que o algoritmo MoG apresenta valores muito mais elevados. A justificação para os mesmos advém do facto do movimento da água contribuir para a classificação errada dessas zonas como *foreground*. Quanto à quantidade de falsos negativos, gráfico da Figura 2.9.26, as diferenças entre os dois algoritmos não são tão significativas, uma vez que, com uma quantidade elevada de falsos positivos existente na região do objecto faz com que este valor seja reduzido, pois a maior parte dos pixéis pertencentes ao verdadeiro *foreground* encontram-se no mapa binário gerado pelo MoG.



a) b) c) d)

Figura 2.9.24: Comparação dos mapas binários de *foreground* do algoritmo híbrido de segmentação de movimento, do MoG e da referência, numa cena minimamente afectada por reflexões especulares e brilhos.

a) Imagem real;
b) Mapa binário de *foreground* gerado através do algoritmo MoG;
c) Mapa binário de *foreground* gerado pelo algoritmo de segmentação híbrido;
d) Mapa binário de *foreground* gerado através de segmentação manual.

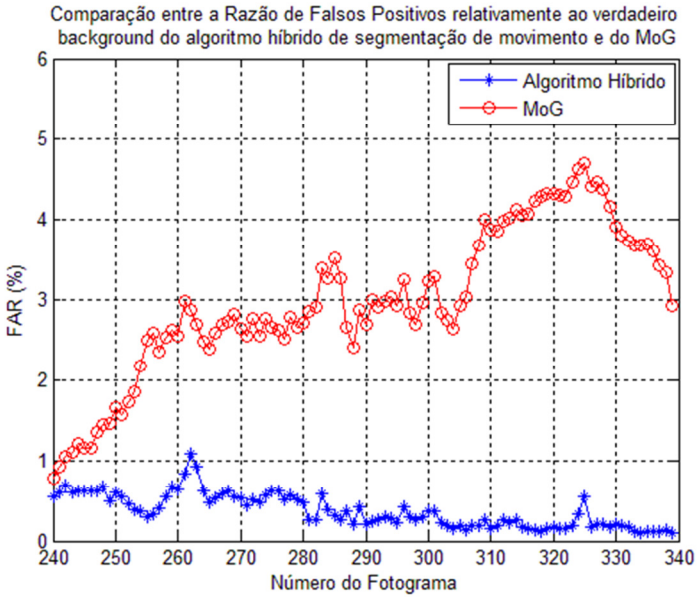


Figura 2.9.25: Comparação da percentagem de falsos positivos relativamente ao verdadeiro *background* do algoritmo híbrido de segmentação e da Mixtura de Gaussianos (MoG), num conjunto de 100 fotogramas consecutivos.

Na Figura 2.9.27, atente-se na amostra retirada e na existência de reflexões especulares, sombras, salpicos e bolhas de ar causadas pelo nadador. Este é um caso ainda mais problemático para o MoG e ao analisarmos o mapa binário de *foreground* gerado pelo mesmo, imagem b), constata-se que a oscilação da água à superfície faz aparecer regiões classificadas como *foreground*, aumentando mais uma vez a quantidade de falsos positivos. Como seria de esperar, uma análise ao gráfico da Figura 2.9.28 permite concluir que o algoritmo híbrido de segmentação de movimento lida melhor com a oscilação da água e as reflexões especulares e brilhos causadas por esta que o MoG. A quantidade de falsos positivos gerada pelo MoG é muito maior quando comparada com a quantidade gerada pelo algoritmo híbrido.

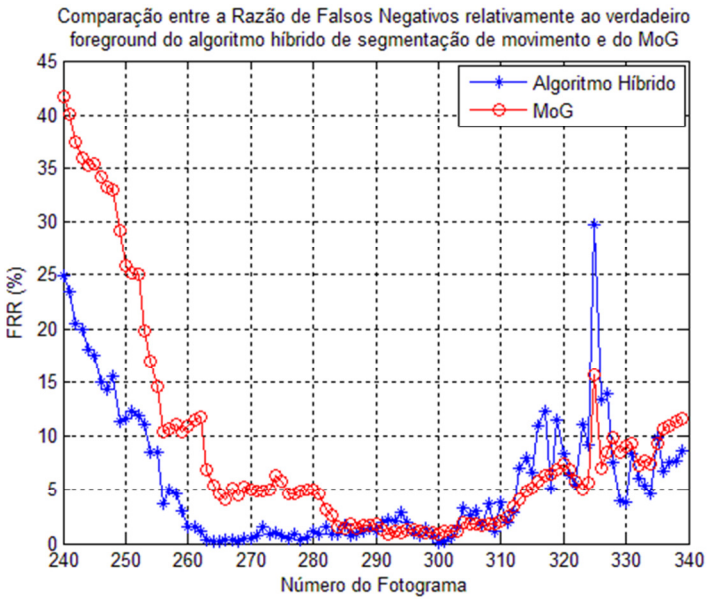
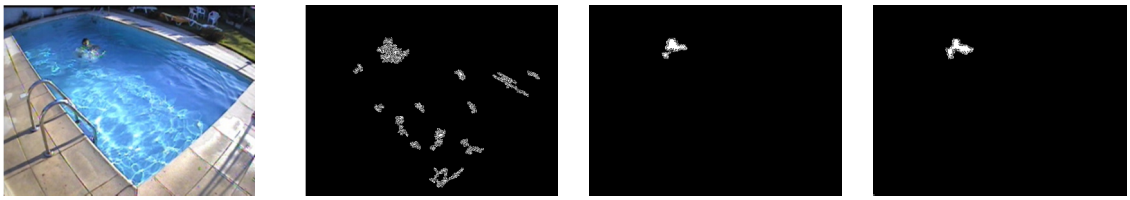


Figura 2.9.26: Comparação da percentagem de falsos negativos relativamente ao verdadeiro *foreground* do algoritmo híbrido de segmentação e da Mixtura de Gaussianos (MoG), num conjunto de 100 fotogramas consecutivos.



a) b) c) d)

Figura 2.9.27: Comparação dos mapas binários de *foreground* do algoritmo híbrido de segmentação de movimento, do MoG e da referência, numa cena fortemente afectada por reflexões especulares e brilhos.

a) Imagem real;
b) Mapa binário de *foreground* gerado através do algoritmo MoG;
c) Mapa binário de *foreground* gerado pelo algoritmo de segmentação híbrido;
d) Mapa binário de *foreground* gerado através de segmentação manual.

Como a sequência de fotogramas utilizada neste teste descreve uma acção de afogamento do indivíduo, este encontra-se parado na cena, existindo movimento apenas dos seus braços. Este facto faz com que, devido à rápida aprendizagem do *background* protagonizada pelo MoG, o nadador seja considerado como pertencente ao *background* de uma forma muito precoce, o que leva ao aumento da quantidade de falsos negativos, tal como o gráfico da Figura 2.9.29 pretende demonstrar. No entanto, nota-se que, por vezes também o algoritmo híbrido de segmentação de movimento apresenta valores elevados de falsos negativos, devendo-se esta situação ao facto do indivíduo ao esbracejar durante o afogamento se encobrir quase completamente pela água agitada, gerando uma enorme quantidade de salpicos e bolhas de ar.

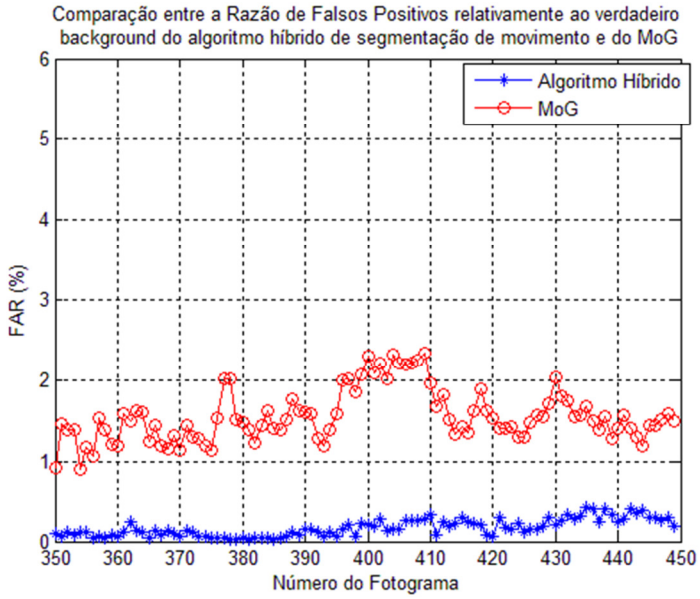


Figura 2.9.28: Comparação da percentagem de falsos positivos relativamente ao verdadeiro *background* do algoritmo híbrido de segmentação e da Mixtura de Gaussianos (MoG), num conjunto de 100 fotogramas consecutivos.

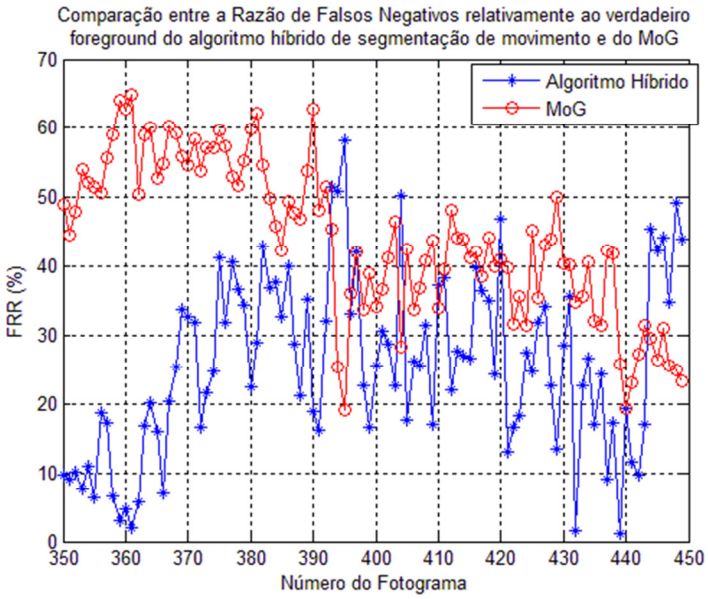


Figura 2.9.29: Comparação da percentagem de falsos negativos relativamente ao verdadeiro *foreground* do algoritmo híbrido de segmentação e da Mixtura de Gaussianos (MoG), num conjunto de 100 fotogramas consecutivos.

2.10 Discussão

O processo de segmentação é o primeiro do *pipeline* e tem como objectivo gerar um mapa binário com os objectos em movimento que não pertencem à cena, nomeadamente à piscina. Este mapa binário de *foreground* será utilizado pelos estágios seguintes de processamento, nomeadamente pelos módulos de seguimento, de reconhecimento de objectos e de análise de comportamento. É por isso fundamental que a saída do módulo de segmentação apresente baixos valores de falsos positivos e falsos negativos, fornecendo dados fiáveis acerca da posição e da forma dos objectos detectados. Observou-se também que, a oscilação da água à superfície, os brilhos e as reflexões especulares de tamanho e posição variáveis ao longo do tempo, bem como, a agitação da água causada pelos nadadores com a formação de bolhas de ar e salpicos, juntamente com todos os factores de variabilidade inerentes a uma cena exterior impõem restrições muito apertadas aos algoritmos de segmentação de movimento existentes. Estes falham nestas condições, não oferecendo uma saída suficientemente eficaz para ser utilizada nos módulos de processamento seguintes e no tipo de sistema, em geral. Daí a necessidade de se encontrar uma nova solução capaz de lidar com estes problemas.

A adopção de dois algoritmos, cada um deles executando em áreas diferentes da cena, permite diferenciar as características muito próprias de um ambiente aquático complexo e conceber um algoritmo específico para lidar com as mesmas. Esse algoritmo tira directamente partido do espaço de cores HSV, uma decisão fundamental na concepção do mesmo, pois através dele é possível isolar a componente tonalidade das cores e aproveitar o grande nível de homogeneidade que a água apresenta. A análise dos resultados mostra claramente que grande parte dos problemas foram eliminados ou, pelo menos, minimizados, ao ponto de não influenciarem erradamente os módulos seguintes. Ficaram no entanto por resolver algumas questões como o problema das sombras que influencia negativamente o processo de segmentação, embora o mesmo só seja notório na área exterior à piscina, o que não é de todo um ponto fundamental, uma vez que a acção de interesse se desenrola no interior da mesma. Além disso, a melhor forma de resolver o problema das sombras aponta para a utilização de mais que uma câmara para se concluir acerca do volume do objecto. Como uma sombra não tem volume, por se tratar de uma projecção, duas câmaras poderiam detectar o fenómeno com fiabilidade suficiente.

Outra questão muito importante tem a ver com a análise dos resultados obtidos pelos algoritmos de segmentação. A dificuldade em quantificar a qualidade do mapa binário obtido continua ainda a

ser uma área de investigação muito relevante, não existindo ainda nenhuma metodologia, verdadeiramente eficaz, capaz de avaliar automaticamente os resultados. No entanto, a análise visual, apesar de subjectiva, permite apurar a qualidade do mapa binário de *foreground* e no caso de serem geradas imagens de referência é possível quantificar a qualidade relativamente a vários factores. O facto de não existirem imagens de referência para cenas como as que aqui são exploradas remete para a obrigação de comparar com segmentações manuais, uma tarefa morosa e demasiadamente lenta, todavia necessária.

Alguns problemas relativos ao aparecimento de reflexões especulares podem ser evitados com recurso a lentes polarizadas nas câmaras. Estes filtros físicos permitem eliminar, ou pelo menos minimizar a intensidade das reflexões especulares, sem ter nenhum custo em termos de software, o que é uma vantagem muito significativa. Todos os problemas relativos à segmentação poderiam ser, no entanto, resolvidos, recorrendo a câmaras térmicas. Com este tipo de dispositivos, as variações de luminosidade, a oscilação da superfície da água e as reflexões especulares, bem como as sombras iriam desaparecer. Ou seja, a água da piscina estaria normalmente a uma temperatura inferior à dos indivíduos, pelo que seria muito fácil segmentar a imagem. Mesmo os objectos que não têm interesse, mas que por vezes estão na piscina, seriam filtrados, desde que estivessem à temperatura ambiente. Poderia no entanto existir o problema de objectos no exterior da piscina aquecerem devido à radiação solar, mas que se resolveria facilmente com a máscara da localização da piscina. Em suma, a utilização das câmaras térmicas iria simplificar e aumentar muito a robustez do sistema, no que respeita à segmentação. No entanto, este tipo de câmaras têm um custo mínimo, em média, de cerca de trinta vezes o de uma câmara de vigilância normal, o que torna este tipo de solução impraticável. Quando o custo deste tipo de dispositivos se tornar atractivo, o caminho a seguir será claramente neste sentido. A utilização destas câmaras só teria influência no módulo de segmentação, sendo apenas necessário conceber um módulo diferente com um simples *threshold* de temperatura. Todos os restantes módulos do *pipeline* permaneceriam iguais em qualquer uma das configurações.

Capítulo 3

3 Seguimento de Objectos

3.1 Introdução

O seguimento de objectos consiste na estimação da trajectória dos mesmos no plano da imagem, ao longo do tempo, localizando a sua posição em cada fotograma e garantindo que, cada um deles é marcado de forma consistente e diferenciada dos demais. Esta tarefa é no entanto bastante complexa, devido não só às variações na aparência dos objectos, provocada pela perda de informação na projecção do mundo 3D para o plano de imagem 2D, mas também às variações de luminosidade que ocorrem na cena e ao ruído inerente à mesma. Problemas como os movimentos complexos dos objectos no plano da imagem, associados a vários tipos de oclusões parciais ou totais, e mesmo a não rigidez dos mesmos, bem como o aparecimento e desaparecimento do campo de visão são outro entrave a este processo, (Yilmaz, Javed, & Shah, 2006).

Um bom algoritmo de seguimento deverá assim ser capaz de lidar com os problemas expostos, de forma transparente, de modo que, os estágios seguintes de processamento, nomeadamente, a análise comportamental dos objectos, possam funcionar de forma eficaz. Num sistema de detecção precoce de afogamento é essencial conter um módulo de seguimento dos objectos previamente detectados de modo a analisar o comportamento singular de cada objecto relevante.

No presente capítulo serão destacados os algoritmos mais importantes nesta matéria, sendo analisados de acordo com as características mencionadas anteriormente. Seguidamente, será proposto um algoritmo de seguimento capaz de lidar com os problemas inerentes ao sistema de detecção precoce de afogamento. Mais tarde serão mencionados os resultados obtidos, pelo seguidor implementado, com dados sintéticos e dados reais. O capítulo terminará com a discussão dos resultados obtidos pelo algoritmo, no que respeita ao cumprimento efectivo da sua tarefa, no âmbito do sistema global.

3.2 Revisão do estado da arte no seguimento de objectos

De acordo com (Yilmaz, Javed, & Shah, 2006) o seguimento de objectos pode ser dividido em três tipos de categorias, *Point Tracking*, *Kernel Tracking* e *Silhouette Tracking*. Na primeira os objectos são representados por um único ponto ou por vários pontos e a sua posição ao longo do tempo é medida de fotograma em fotograma de modo a fazer corresponder os mesmos. No seguimento baseado num *Kernel*, ou seja, numa forma geométrica, em cada fotograma é calculado o movimento do *kernel* na forma de transformações paramétricas, tais como a rotação e a translação. O seguimento baseado na silhueta dos objectos utiliza um modelo da forma ou aparência dos mesmos fazendo depois a correspondência entre eles em fotogramas consecutivos.

3.2.1 Point Tracking

O *Point Tracking* pode ainda ser subdividido em métodos determinísticos, baseados na heurística de movimento dos mesmos e métodos probabilísticos que estabelecem a correspondência entre objectos, baseados nas medições das suas características tendo em conta as incertezas associadas. Os autores (Sethi & Jain, 1987) utilizaram um método determinístico baseado nas restrições de proximidade e rigidez dos objectos, ou seja, um objecto não pode mudar, significativamente, de posição e de forma de um fotograma para o seguinte. Para resolver a correspondência entre objectos, os autores utilizam um algoritmo de procura extensiva que visa minimizar o custo entre os vizinhos mais próximos. Esta abordagem, no entanto, não leva em consideração oclusões e entrada e saída de objectos do campo de visão. Este problema foi resolvido pelos autores (Sethi & Salari, 1990) ao adicionar um número de pontos hipotéticos representativos dos objectos em falta, depois de efectuar a correspondência entre os pontos existentes. Para objectos definidos por múltiplos pontos, os trabalhos de (Veenman, Reinders, & Backer, 2001) introduziram a restrição *common motion*, para efectuar a correspondência entre objectos. Esta restrição indica que os pontos próximos, ou numa pequena vizinhança, devem ter velocidades similares. No entanto, e apesar de tratar as oclusões, o número de objectos na cena deve-se manter constante, não tratando entradas e saídas dos mesmos. O algoritmo proposto pelos autores, ao contrário dos anteriores, utiliza um método de optimização, no caso, o *Hungarian algorithm*. Outros autores, como (Shafique & Shah, 2003), propuseram uma abordagem *multiframe* tirando partido de uma janela temporal, durante a qual os objectos devem manter posições e velocidades similares. Esta abordagem permite ainda tratar entradas e saídas de objectos na cena bem como oclusões, quando o tempo de duração destas for inferior à janela

temporal. Já o autor (Lowe, 2004) propôs uma transformação das características do objecto, *Scale-Invariant Feature Transform* (SIFT), baseada nos seus pontos, sendo esta invariante relativamente à translação, rotação e variação da escala do objecto.

Os métodos probabilísticos, por outro lado, assentam em abordagens relacionadas com modelos de espaço de estados, tendo em conta as incertezas relativamente ao processo e às medições das posições dos objectos na imagem. Os filtros de *Kalman*, (Kalman, 1960), e os filtros de partículas, (Tanizaki, 1987), permitem modelar um único objecto ao longo do tempo tolerando a existência de ruído na imagem. O filtro de *Kalman* é utilizado para estimar o estado de um sistema linear definido por uma distribuição Gaussiana. A sua execução dá-se em dois passos, um de previsão, onde o estado actual do objecto é utilizado para determinar o estado seguinte e outro de correcção, onde os valores medidos permitem actualizar o verdadeiro estado do mesmo. Contudo, o filtro de *Kalman* considera que as variáveis que compõem o estado do sistema são caracterizadas por distribuições normais, fazendo com que a estimação seja pouco eficaz quando não se trata desse caso. Nos filtros de partículas já esse problema não ocorre, uma vez que são utilizadas várias distribuições gaussianas e as partículas com diferentes pesos são assim capazes de representar mais eficazmente o estado do objecto. Um elevado número de partículas pode, porém, comprometer o funcionamento em tempo real. Para que estes filtros possam ser utilizados é necessário antes efectuar uma correspondência dos objectos. Existem várias técnicas de associação estatística de dados para esse efeito, entre elas a *Joint Probability Data Association Filter* (JPDAF) e a *Multiple Hypothesis Tracking* (MHT), (Reid, 1979), sendo estas as mais utilizadas. A primeira foi utilizada por (Hager & Rasmussen, 2001) para efectuar o seguimento de regiões. Este método apresenta uma grande limitação, pois não permite variações no número de objectos, não comportando deste modo oclusões, entradas e saídas dos mesmos na cena. O funcionamento assenta na previsão do estado de cada objecto a partir de um filtro de *kalman*, sendo depois medidos os valores actuais das variáveis que representam o seu estado. Logo que se tenha os valores reais das variáveis de cada objecto faz-se uma correspondência baseada na melhor aproximação entre os valores previstos e os realmente medidos, podendo aplicar o segundo passo do filtro de *kalman*, a actualização do estado com base nos valores medidos para cada objecto. Os trabalhos efectuados por (Jilkov, Huimin, Li, & Nguyen, 2006) são baseados em múltiplos pontos representativos do objecto, extraídos por técnicas de processamento de imagem. O objecto rígido é composto por um conjunto de juntas com movimento associado, sendo este estimado por intermédio de um filtro não linear, no caso o *Unscented Kalman Filter* (UKF), por ser

mais eficiente que o Extended Kalman Filter (EKF), (Rosales & Sclaroff, 1999), ou o filtro de partículas, embora seja computacionalmente mais pesado. O algoritmo MHT, por outro lado, comporta oclusões, entrada e saída de objectos do campo de visão. Este adopta uma abordagem de correspondência que não se baseia em dois fotogramas consecutivos, mas sim em vários, determinando a trajectória final através do conjunto de correspondências mais similares dentro do período de observação. Este algoritmo é muito pesado, não só em termos de requisitos de memória como também de processamento, sendo esta carga exponencial relativamente ao número de objectos na cena. Deste modo, (Streit & Luginbuhl, 1994), propuseram uma abordagem probabilística ao MHT, designada por PMHT, de modo a reduzir a carga computacional. Esta metodologia elimina a necessidade de efectuar uma enumeração exaustiva de associações no espaço de hipóteses, considerando as mesmas estatisticamente independentes.

3.2.2 Kernel Tracking

Este tipo de algoritmos pode ser dividido em duas categorias, relativamente à forma de representação da aparência dos objectos. Alguns utilizam *templates* e *density-based appearence models* e outros *multiview appearence models*. Os primeiros ainda podem ser divididos em duas subcategorias, os que permitem o seguimento de apenas um objecto e os que permitem o seguimento de vários objectos. O método mais utilizado no seguimento de um único objecto designa-se de *template matching* e permite, através da procura exaustiva na imagem actual, encontrar a região similar ao *template* do objecto na imagem anterior. Além das limitações já conhecidas desta técnica, as características escolhidas para efectuar a medida de similaridade são um factor determinante, uma vez que a intensidade da imagem pode variar ao longo do tempo devido às variações do nível de luminosidade que afecta a cena. Outro problema tem a ver com a procura na imagem total do *template* do objecto. Este tipo de abordagem tem um custo computacional elevado, pelo que alguns autores, tais como, (Wu, Bell, & Schweitzer, 2002), desenvolveram técnicas de procura apenas na região próxima do objecto, tendo em conta a sua velocidade e direcção anteriores.

Na categoria dos algoritmos *density-based appearence models*, (Fieguth & Terzopoulos, 1997) utilizaram a média encontrada dentro da região rectangular que envolve o objecto na imagem anterior para comparar com a média encontrada em oito regiões vizinhas. Aquela que apresentar um valor mais próximo é seleccionada para novo *template*. Existem ainda autores, (Comaniciu, Ramesh, & Meer, 2003), que fazem uso do histograma definido na região ocupada pelo objecto

implementando *mean-shift*. Este tipo de abordagem permite comparar iterativamente a similaridade do histograma do *template* com os histogramas nas localizações próximas do mesmo, utilizando o coeficiente de *Bhattacharya*. Esta técnica é substancialmente mais rápida do que as baseadas na procura exaustiva, uma vez que a procura é efectuada por convergência, normalmente em menos de seis iterações, nas zonas próximas do objecto. Outras abordagens, (Jepson, Fleet, & Elmaraghi, 2003), assentam na mistura de três componentes, uma representando as características estáveis da aparência do objecto e as outras as características transitórias e o ruído do processo, respectivamente. Através do algoritmo EM são determinados os parâmetros que compõem a mistura sendo depois utilizadas as componentes estável e transitória para encontrar o objecto na próxima imagem levando em consideração uma transformação por translação, rotação e escala do objecto. (Shi & Tomasi, 1994) recorrem ao cálculo do fluxo óptico nas regiões centradas nos pontos de interesse pertencentes à região geométrica envolvente do objecto para deste modo determinar a translação sofrida pela *bounding box* do mesmo. Um algoritmo de seguimento de múltiplos objectos baseado na modelação da imagem global por um conjunto de camadas e fazendo uso de *density-based appearence models* é proposto por (Tao, Sawhney, & Kumar, 2002). O *background* é representado por uma única camada e cada um dos objectos existentes é representado por outras camadas independentes. Cada uma delas é definida por uma elipse e é modelada por um modelo de movimento provido de translação e rotação e um modelo de aparência descrito por uma distribuição normal. Assim, a probabilidade de cada pixel pertencer a uma camada, ou seja, a um objecto detectado é calculada com base na posição e forma anteriores do objecto, sendo depois adicionada a probabilidade relativa à aparência modelada por uma distribuição gaussiana. Finalmente são estimados os parâmetros da mistura, através do algoritmo EM, que maximizam a probabilidade do pixel pertencer à camada em questão. Este algoritmo é capaz de lidar com oclusões entre os vários objectos e o *background*. (Zhou, Yuan, & Shi, 2008) propuseram a utilização do algoritmo EM para inferir acerca das características obtidas por SIFT, conjuntamente com um modelo de aparência baseado na cor e obtido por *mean-shift*. Esta abordagem é invariante relativamente à translação e rotação do objecto e é fiável mesmo quando uma das duas medidas se torna instável.

O seguimento de objectos baseado em *multiview appearence models*, ao contrário dos anteriores, utiliza modelos de aparência gerados *offline*, ou seja, necessita de ser treinado de modo a construir o modelo do objecto sobre diferentes pontos de vista. Esta necessidade existe devido às mudanças significativas que podem ocorrer na aparência do objecto, não permitindo que o mesmo possa ser

seguido correctamente ao longo do tempo caso o modelo seja construído à custa das observações mais recentes. Este tipo de algoritmos não lidam com todas as transformações lineares do objecto e não são capazes de implementar o seguimento de vários objectos. Os autores (Black & Jepson, 1998) transformaram a imagem para o espaço próprio (*eigenspace*), comparando depois a mesma com o modelo de aparência do objecto, construído através do algoritmo *Principal Component Analysis* (PCA). A utilização de classificadores baseados em *Support Vector Machine* (SVM) no seguimento de um único objecto foram utilizados por (Avidan, 2001). Durante o treino do classificador os exemplos positivos consistem no objecto a ser seguido e os exemplos negativos consiste na restante imagem que não é relevante para efeitos de seguimento. Ao contrário das abordagens anteriores este algoritmo tem como objectivo maximizar a classificação SVM nas regiões da imagem de modo a encontrar aquela que apresenta maior pontuação na classe dos exemplos positivos. Esta abordagem tem a vantagem de incorporar explicitamente o *background* no seguidor, obtendo um conhecimento acerca dos objectos que não devem ser seguidos.

Os autores (Lu & Tan, 2004) utilizaram uma abordagem mais simples, baseada apenas na previsão e actualização do estado de um objecto através de um filtro de *kalman*. Cada objecto é descrito por uma elipse sendo depois utilizada uma janela de procura cujo tamanho depende do eixo maior e menor dessa elipse associado a uma tolerância vertical e horizontal. A procura é efectuada com base na posição prevista pelo filtro de *kalman* para o objecto. A posição de cada novo objecto encontrado no fotograma actual é comparada com a posição prevista pelo filtro de *kalman* e determinada a distância entre eles. A menor das distâncias encontrada corresponderá à nova posição do objecto na imagem actual. O estado de um objecto é descrito pela sua posição e velocidade medidas em fotogramas consecutivos. Este algoritmo é capaz de lidar com a oclusão parcial, mas durante esse período não efectua a análise comportamental dos objectos, sendo que quando estes se separam são considerados objectos novos na cena. Os autores afirmam que esta abordagem permite a entrada e saída de objectos na imagem.

3.2.3 Silhouette Tracking

Os métodos de seguimento baseados em silhuetas têm relevância sobretudo quando as formas exibidas pelos objectos apresentam elevada complexidade, tais como a não rigidez. O objectivo de um algoritmo deste tipo consiste em identificar em cada novo fotograma o objecto com base no seu modelo de aparência. Este modelo pode ser conseguido através de histogramas ou dos contornos da região definida pelo objecto. Estes métodos dividem-se em dois grupos, designados

de *shape matching* e *contour tracking*. No primeiro, o objecto é normalmente modelado a partir da sua representação baseada em contornos e procurado no fotograma seguinte através de uma métrica de similaridade. Os autores (Huttenlocher, Noh, & Rucklidge, 1993) utilizaram a distância de *Hausdorff* para construir uma superfície de correlação entre os pontos pertencentes ao contorno do objecto no fotograma anterior e no actual de forma a seleccionar o mínimo e efectuar assim a correspondência entre os mesmos. No entanto, existem metodologias que fazem a correspondência de objectos entre fotogramas consecutivos baseados não só no movimento dos mesmos, tal como nos algoritmos do tipo *Point Tracking*, mas também no seu modelo de aparência. Este tipo de modelos consiste normalmente em funções densidade de probabilidade, isto é, histogramas, na região ocupada pelo objecto na imagem. No entanto, (Kang, Cohen, & Medioni, 2004) propuseram a definição de histogramas na área dos círculos centrados nos pontos de controlo pertencentes à circunferência que circunscreve o objecto. Através da utilização deste tipo de histogramas, o modelo passa a ser invariante relativamente à translação, rotação e escala do objecto. Estes autores utilizaram como métrica de similaridade a distância de *Bhattacharya* e a divergência de *Kullback-Leibler*. (Dalal & Triggs, 2005) propuseram histogramas de gradientes orientados para descrever os objectos utilizando um conjunto pesado de vários classificadores fracos, treinados previamente. No entanto, existem abordagens que se baseiam no cálculo do fluxo óptico no interior da silhueta para gerar a direcção e a velocidade do objecto de acordo com o fluxo dominante. (Sato & Aggarwal, 2004) aplicaram a transformada de *Hough* a cada região de pixéis em movimento pertencentes ao objecto numa janela temporal, ao longo de vários fotogramas consecutivos, de modo a calcular as chamadas matrizes de voto. Estas matrizes constituem a imagem 4D designada por *Temporal Spatio-Velocity* (TVS), em cada fotograma. Este tipo de técnica proporciona a correspondência entre silhuetas de uma imagem para a seguinte baseada apenas nos padrões de velocidade detectados na região que compõe o objecto, não sendo baseada num modelo de aparência, como as anteriores. Deste modo é menos sensível às variações de aparência causadas por objectos não rígidos.

Os métodos *contour tracking*, por outro lado, baseiam-se na evolução do contorno do objecto ao longo do tempo em fotogramas consecutivos. Existem autores que utilizam modelos de espaço de estados para prever e actualizar a forma do contorno e o movimento do mesmo. Outros, porém, optaram pela utilização de técnicas como *gradient descent*, um algoritmo de optimização de primeira ordem, de modo a minimizarem a energia do contorno do objecto na sua evolução. (Terzopoulos & Szeliski, 2002) definiram o estado do objecto baseados na dinâmica dos pontos de

controlo do mesmo descrita por um modelo *spring model*. O novo estado do contorno é previsto e actualizado através de um filtro de *Kalman*. Por outro lado, (MacCormick & Blake, 2000) utilizaram as bordas normais ao contorno nos pontos de controlo do mesmo, sendo a previsão e actualização destas efectuada por intermédio de um filtro de partículas. Os autores incluíram ainda um esquema de tratamento de oclusão designado por *exclusion principle* capaz de suportar oclusão entre dois objectos. Contudo, estes métodos de *contour tracking* não permitem mudanças de topologia dos objectos, ou seja, junção e separação dos mesmos. As metodologias baseadas na minimização da energia do contorno do objecto foram inicialmente propostas por (Bertalmio, Sapiro, & Randall, 2000) através da determinação do fluxo óptico. Estes autores utilizaram uma representação para cada ponto pertencente ao contorno do objecto, permitindo mudanças de topologia no mesmo. Uma alternativa à utilização do fluxo óptico foi sugerida por (Yilmaz, Li, & Shah, 2004), consistindo esta na determinação de estatísticas associadas às cores e à textura encontradas na região ocupada pelo objecto. O modelo da forma do objecto era igualmente baseado na sua função *level set*, sendo assim capaz de lidar com oclusões entre múltiplos objectos, junções e separações. Não é necessário parametrizar cada objecto, sendo as formas de todos eles definidas completamente pela fronteira obtida quando a função *level set* toma o valor zero.

Os autores (Eng, Wang, Kam, & Yau, 2004) utilizam os modelos de *foreground* para cada objecto gerados pelo módulo de segmentação. Na ocorrência de oclusões parciais os autores segmentam a região ocupada pelos dois objectos através do algoritmo *k-means*. Cada uma das regiões encontradas pelo algoritmo é depois atribuída a um e outro objecto através de uma rede *Markov Random Field*. No entanto os autores nada referem quanto ao tratamento de oclusões totais e entrada ou saída de objectos da cena.

3.2.4 Discussão

Os desafios do seguimento de objectos centram-se, principalmente, na capacidade dos algoritmos poderem seguir vários objectos, simultaneamente, lidando correctamente com oclusões, entrada e saída dos mesmos na cena, comportando a sua não rigidez. O facto de os objectos mudarem a sua forma ao longo do tempo devido à projecção do mundo 3D para o plano de imagem, provoca sérios problemas ao seguimento. As características do objecto podem ser completamente diferentes do fotograma actual para o seguinte. E isso acontece mesmo em objectos rígidos. Mas

quando os objectos variam a sua forma, mesmo no mundo 3D, a variação provocada pela projecção somada à variação de postura constitui um sério problema para o seguimento.

No contexto do sistema de detecção de afogamento, é fundamental que o módulo de seguimento seja capaz de seguir vários objectos em simultâneo, tratar as oclusões entre os mesmos e a sua entrada e saída de cena, principalmente quando estes estão dentro da área correspondente à piscina. Isto porque, o módulo subsequente, que efectua o reconhecimento de objectos e a análise comportamental, necessita de extrair as características dos objectos ao longo do tempo de forma coerente. Ou seja, é necessário garantir que um determinado objecto seja seguido ao longo de vários fotogramas, de modo a analisar vários parâmetros relativos ao mesmo. Caso existam trocas, devido às oclusões, ou entrada e saída de objectos em cena, a resposta do módulo de inferência de comportamento será erradamente afectada. Este problema é crítico, uma vez que, uma situação de afogamento pode erradamente ser considerada uma situação normal, o que constitui um resultado verdadeiramente catastrófico. É ainda importante destacar que os objectos mais importantes a seguir são não rígidos, variando a sua postura ao longo do tempo. Sendo uma das preocupações no desenvolvimento do sistema a utilização do menor número de câmaras, este indicador permite concluir que os objectos serão seguidos por uma única câmara monocular e fixa. Significa assim que a variação da forma dos objectos, mesmo dos rígidos, existirá, devido à projecção da cena no plano de imagem. Estas variações da forma dos objectos constituem um verdadeiro obstáculo ao seguidor.

3.3 Seguidor de vários objectos com tratamento de oclusões entradas e saídas

O módulo de seguimento de objectos deve ser genérico, no sentido em que, fornecido um mapa binário de *foreground*, o mesmo seja capaz de seguir separadamente cada objecto, ou região, presente no mesmo. Desta forma, não é relevante o tipo de segmentação utilizada, uma vez que o importante é que seja gerado um mapa binário de regiões. Esta abordagem, por camadas, permite desenvolver cada módulo separadamente e sem depender dos demais, desde que o interface seja conhecido. Assim, é possível mudar qualquer módulo, garantindo o funcionamento global do sistema. O módulo de seguimento deve identificar cada região presente no mapa binário de *foreground*, distinguindo-a de todas as outras ao longo de vários fotogramas, de acordo com os requisitos anteriormente descritos. Neste processo de seguimento, o seguidor deve identificar cada

região e fornecer como saída algumas das suas características, tais como a área, a *bounding box* que contém a região, a posição e a velocidade do seu centro de massa.

O algoritmo de seguimento proposto reduz as silhuetas presentes no mapa binário de *foreground*, fornecido como entrada, a pontos, aos quais corresponde o centro de massa da região. Deste modo, a posição de cada ponto ao longo do tempo é estimada por intermédio de um filtro de *Kalman*. A correspondência entre regiões é feita com base na comparação entre as características actuais e as características previstas pelo filtro de *Kalman*. O algoritmo de correspondência entre objectos é baseado numa versão não balanceada do *Hungarian Algorithm* (Bourgeois & Lassalle, 1971). Este foi ainda adaptado para contemplar oclusões, entradas e saídas de objectos. O filtro de *Kalman* utiliza várias medidas, no caso, a posição do centro de massa, a área da região e o tamanho e posição da *bounding box*. Estas medidas são também utilizadas pelo algoritmo de correspondência para medir o grau de similaridade entre os objectos. O esquema do algoritmo de seguimento proposto é apresentado na Figura 3.3.1. Este recebe como entrada o mapa binário de *foreground*, gerado pelo segmentador, e define-se por $FG_n(x, y)$. De seguida, são etiquetadas cada uma das regiões de pixéis adjacentes, com conectividade 8, através do algoritmo de etiquetagem sequencial apresentado em (Horn B. , 1997). Este algoritmo fornece como saída um mapa de regiões, identificadas por números inteiros sequenciais, designada por $LM_n(x, y)$. A partir deste mapa são determinadas as dimensões e a posição do canto superior esquerdo das j *bounding boxes* que delimitam cada região de pixéis adjacentes, $\Phi^b(x, y, h, w) = \{\Phi_1^b, \Phi_2^b, \dots, \Phi_j^b\}$. Muitas destas regiões podem pertencer ao mesmo objecto, mas como estão separadas, devido a erros no processo de segmentação, a sua união é feita com base na sobreposição das *bounding boxes*. Ou seja, admite-se que, sempre que duas ou mais *bounding boxes* estejam sobrepostas, as suas regiões pertencem ao mesmo objecto. O Algoritmo 3.3.1 tem como objectivo a aglomeração das *bounding boxes*. Após esta operação existe um número k, igual ou inferior a j, de *bounding boxes* e, como tal, de objectos. Todos os pixéis dentro da região delimitada pela respectiva *bounding box* fazem parte do objecto, pelo que, a área e a posição do centro de massa de cada região são calculadas através dos momentos cartesianos de primeira e segunda ordem de acordo com a Equação (3.3.1), que se segue:

$$m_{pq} = \sum_{x=1}^{M} \sum_{y=1}^{N} x^p y^q P_{xy} \qquad \textbf{(3.3.1)}$$

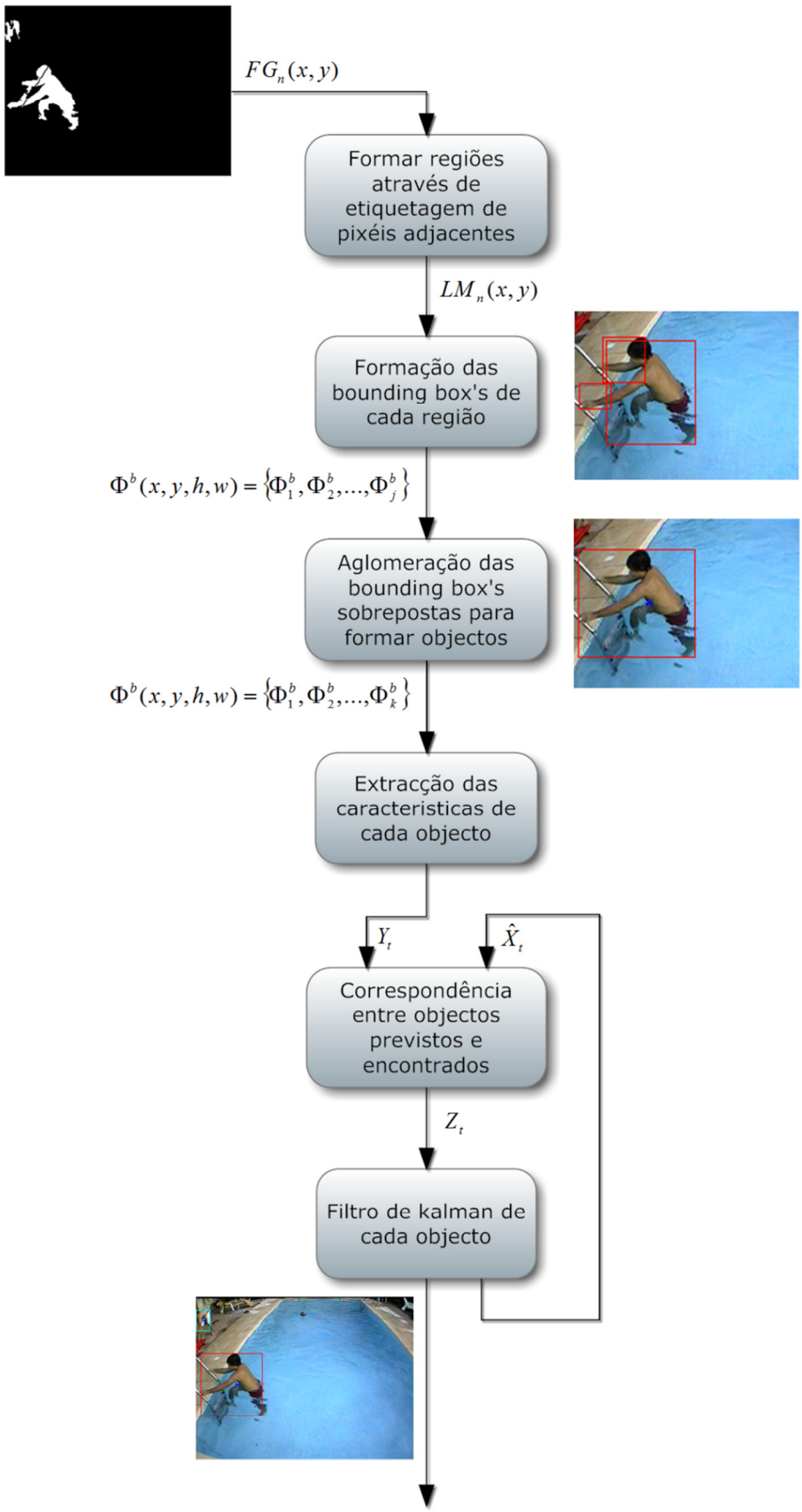


Figura 3.3.1: Esquema do algoritmo de seguimento de vários objectos com tratamento de oclusões, entradas e saídas.

Se $j > 1$ então:
$k = 1$
Enquanto $k < j - 1$ fazer:
De $i = k + 1$ até j fazer:

$$xDim = \begin{bmatrix} x_k^b & 0 \\ x_k^b + w_k^b & 0 \\ x_i^b & 1 \\ x_i^b + w_i^b & 1 \end{bmatrix}$$

$$yDim = \begin{bmatrix} y_i^b + h_i^b & 0 \\ y_i^b & 0 \\ y_i^b + h_i^b & 1 \\ y_i^b & 1 \end{bmatrix}$$

Ordenar as linhas de $xDim$ e $yDim$ por ordem crescente em relação à 1ª coluna
Se $xDim(1,2) \neq xDim(2,2) \wedge yDim\,(1,2) \neq yDim(2,2)$ então:
$h = yDim(4,1) - yDim(1,1)$
$w = xDim(4,1) - xDim(1,1)$
$\Phi_k^b = (xDim(1,1), yDim(1,1), h, w)$
$\Phi_i^b = (0,0,0,0)$
Elimina as *bounding box* nulas
$k = 1$
$j = j - 1$
Fim
$k = k + 1$
Fim

Algoritmo 3.3.1: Aglomeração das *boundind boxes* sobrepostas.

Assim, a área de cada objecto é dada por:

$$\Phi^a = \{A_1, A_2, \dots, A_k\}, com\ \Phi_k^a = m_{00}^k$$

A posição do centro de massa é dada por:

$$\Phi^c = \{\Phi_1^c, \Phi_2^c, \dots, \Phi_k^c\}, com\ \Phi_k^c = (x_k^c, y_k^c) = \left(\frac{m_{10}^k}{m_{00}^k}, \frac{m_{01}^k}{m_{00}^k}\right)$$

As características dos objectos medidas, no instante de tempo t são assim definidas por:

$$Z_t = \{\Phi^c, \Phi^b, \Phi^a\}$$

Como se sabe, as medições das características dos vários objectos contêm erros, devido a todo um conjunto de processos ao qual é submetida a imagem até chegar a esta fase. Isso significa que,

existe um certo grau de incerteza associado a estas variáveis. Sabe-se também que, a evolução destas variáveis ao longo do tempo é descrita por um sistema dinâmico, que descreve o estado físico do objecto em termos de posição, velocidade e aceleração. Mas este sistema dinâmico também contém ruído, designado ruído do processo, impossibilitando a modelação do sistema com recurso a modelos determinísticos. Deste modo, torna-se necessário filtrar os ruídos associados às medidas e ao próprio processo, para estimar os estados do sistema, de acordo com as medidas observadas e as leis associadas à dinâmica do mesmo. O filtro de *Kalman* discreto, permite estimar o estado do sistema $X \in \Re^n$, governado pela seguinte equação:

$$X_t = Ax_{t-1} + w_{t-1}$$

(3.3.2)

Onde os valores das variáveis medidas $Z \in \Re^m$ são descritas pela equação:

$$Z_t = Hx_t + v_t$$

(3.3.3)

A matriz $A_{n\times n}$ designa-se por matriz de transição de estado e a matriz $H_{m\times n}$ relaciona a matriz de estado com as variáveis medidas. As variáveis aleatórias w_t e v_t representam o ruído do processo e das medidas, respectivamente. É assumido que estes ruídos são independentes e descritos por distribuições gaussianas:

$$p(w) \sim N(0, Q)$$

$$p(v) \sim N(0, R)$$

Deste modo, Q representa a matriz de co-variância do ruído associado ao processo e R representa a matriz de co-variância do ruído associado às medições.

As transições de estado são determinadas com recurso a funções lineares, tratando-se assim do filtro de *Kalman* standard, que, para o caso, se adequa perfeitamente, como se poderá verificar, posteriormente, na Secção 3.5. O filtro de *Kalman* é um algoritmo recursivo composto por dois passos. O primeiro passo designa-se de *time update* (ou previsão) e determina o estado esperado e a respectiva matriz de co-variância, em função do estado anterior e da matriz de transição de estado. O segundo passo designa-se de *measurement update* (ou correcção) e utiliza o estado estimado no passo anterior e os valores medidos para corrigir o estado do sistema e a

correspondente matriz de co-variância, P. As equações específicas de cada passo são apresentadas de seguida:

1. Previsão

   i) Prevê o próximo estado

$$\hat{X}_t^- = A\hat{X}_{t-1} \tag{3.3.4}$$

   ii) Prevê a co-variância do erro no próximo estado

$$P_t^- = AP_{t-1}A^T + Q \tag{3.3.5}$$

2. Correcção

   i) Calcular o ganho de *Kalman*

$$K_t = P_t^- H^T (HP_t^-)^{-1} \tag{3.3.6}$$

   ii) Actualizar o valor estimado através da medição actual Y_t

$$\hat{X}_t = \hat{X}_t^- + K_t\left(Z_t - H\hat{X}_t^-\right) \tag{3.3.7}$$

   iii) Actualizar a co-variância do erro

$$P_t = (I - K_t H)P_t^- \tag{3.3.8}$$

A representação do seguidor de cada objecto no espaço de estados implica a definição do próprio estado e da matriz que provoca a transição entre estados. Partindo do princípio de que o movimento dos objectos é governado pelas leis da física correspondentes ao movimento com aceleração constante, a matriz de transição de estado é baseada nas seguintes equações:

$$p(t) = p(t-1) + v(t-1)\Delta t + a(t-1)\frac{\Delta t^2}{2} \tag{3.3.9}$$

$$v(t) = v(t-1) + a(t-1)\Delta t \tag{3.3.10}$$

$$a(t) = a(t-1)$$

(3.3.11)

Em que $p(t)$, $v(t)$ e $a(t)$ representam, respectivamente, a posição, velocidade e aceleração actuais e $p(t-1)$, $v(t-1)$ e $a(t-1)$ representam, respectivamente, a posição, velocidade e aceleração anteriores, ou seja, no fotograma anterior. Δt corresponde ao intervalo de tempo decorrido entre dois fotogramas consecutivos. Deste modo, a posição do centro de massa e do canto superior esquerdo da *bounding box* são actualizadas de acordo com as equações anteriores. O vector de estado do sistema e a matriz de transição de estado são dados por:

$$X_t = \begin{bmatrix} x_t^c \\ y_t^c \\ v_{x_t}^c \\ v_{y_t}^c \\ a_{x_t}^c \\ a_{y_t}^c \\ x_t^b \\ y_t^b \\ h_t^b \\ w_t^b \\ A_t \end{bmatrix}, A = \begin{bmatrix} 1 & 0 & T & 0 & \frac{1}{2}T & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & T & 0 & \frac{1}{2}T & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & T & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & T & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Por exemplo, a uma velocidade de 25 fotogramas por segundo o valor de $T = \frac{1}{25}$. A matriz H, que relaciona o estado com os valores medidos, é dada por:

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

As matrizes de co-variância do ruído do processo e do ruído associado às medidas das variáveis observadas são, respectivamente dadas por:

$$Q = 0.05 I_{11}$$

$$R = I_7$$

Por último, as condições iniciais, no instante $t = 0$, para o estado do sistema e para a matriz de co-variância do erro são, respectivamente, dados por:

$$X_0 = [0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0]^T,\ P_0 = 10 I_{11}$$

Cada um dos objectos identificados terá associado a si um filtro de *Kalman*, tal como o que foi anteriormente descrito. Mas, tal como o esquema da Figura 3.3.1 mostra, é necessário um algoritmo que efectue a correspondência entre os objectos encontrados em cada fotograma consecutivo. É esse bloco o responsável pela associação dos novos objectos, gerados no estágio de segmentação, cujos valores, no instante de tempo t, são dados por Y_t, às posições e áreas previstas pelo filtro de *Kalman* de cada um deles, dados por $\hat{X}_t$. A saída do bloco de correspondência, dada por Z_t, é depois utilizada pelo filtro de *Kalman* para estimar novamente o próximo estado.

Tal como foi mencionado na secção introdutória deste capítulo, o algoritmo de correspondência é baseado no *Hungarian Algorithm*, adaptado para matrizes de custo rectangulares, ou seja, não balanceado (Bourgeois & Lassalle, 1971). Esta abordagem permite determinar a correspondência mesmo quando o número de objectos previstos pelos filtros de *Kalman* e detectados pelo segmentador são diferentes, daí o termo não balanceado. Contudo, este algoritmo não permite averiguar o tipo de ocorrência, ou seja, se aconteceu uma oclusão, uma entrada ou uma saída de um ou mais objectos, mesmo quando a matriz de custo é balanceada. Deste modo, depois de construída a matriz de custo e determinada a correspondência, os resultados obtidos têm de ser verificados, através do seu valor de similaridade, de modo a descobrir as várias possibilidades. Esta adaptação permite lidar com os vários problemas associados aos seguidores de vários objectos, já mencionados anteriormente. A matriz de custo, utilizada pelo algoritmo de correspondência, é construída com base nas características determinadas pelo módulo de segmentação, Z_t, e previstas pelo filtro de *Kalman*, $\hat{X}_t$. Assumindo que o número de objectos previstos é i e que o número de objectos actualmente encontrados é j, a matriz de custo é definida por:

$$CM = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1j} \\ S_{21} & S_{22} & \dots & S_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ S_{i1} & S_{i2} & \cdots & S_{ij} \end{bmatrix}$$

Esta matriz de custo é constituída por coeficientes de similaridade, S_{ij}. Estes coeficientes são determinados através do somatório pesado das diferenças entre as várias características tal como a seguinte equação demonstra:

$$S_{ij} = d_{ij}^{c} + d_{ij}^{b} + \sqrt[2]{\left(h_i^b - h_j^b\right)^2 + \left(w_i^b - w_j^b\right)^2} + \alpha\left|A_i - A_j\right| \tag{3.3.12}$$

Sendo d_{ij}^{c} a distância euclidiana entre o centro de massa dos objectos i e j, definida por:

$$d_{ij}^{c} = \sqrt[2]{\left(x_i^c - x_j^c\right)^2 + \left(y_i^c - y_j^c\right)^2}$$

(3.3.13)

E d_{ij}^{b} a distância euclidiana entre o canto superior esquerdo da *bounding box* i e j, definida por:

$$d_{ij}^{b} = \sqrt[2]{\left(x_i^b - x_j^b\right)^2 + \left(y_i^b - y_j^b\right)^2}$$

(3.3.14)

O peso α serve para adaptar a escala das características, de modo que todas elas tenham uma influência semelhante na contribuição para o coeficiente de similaridade. Tendo a imagem um tamanho de $M \times N$, o peso α é definido por:

$$\alpha = \frac{N}{MN} = \frac{1}{M}$$

A matriz de correspondência, determinada pelo algoritmo de correspondência, é binária, sendo definida por AM_{ij}. Sempre que na posição $AM(i,j)$ o valor seja 1, significa que existe uma probabilidade dos objectos i e j serem o mesmo. No entanto, é necessário efectuar uma verificação e correcção de correspondências através do Algoritmo 3.3.2.

Este algoritmo determina, para cada objecto, um coeficiente de confiança, definido por:

$$C = \{C_1, C_2, \dots, C_k\},\ C_i \in \left[1\ \frac{1}{T}\right]\ \forall\ i \in [1\ k]$$

E fornece um identificador definido por:

$$ID = \{ID_1, ID_2, \dots, ID_k\},\ ID_i \in [-1\ 20]$$

Deste modo, e de acordo com acontecimentos, tais como, oclusões, entradas e saídas, os níveis de confiança de cada objecto variam, sendo estes válidos enquanto o seu valor for superior a 0. Um objecto válido tem associado a si um filtro de *Kalman*, tal como foi descrito anteriormente, enquanto um objecto com nível de confiança nulo considera-se que não existe, podendo o seu lugar ser ocupado por um objecto novo que, entretanto, apareça. Níveis de confiança nulos implicam que o identificador tenha o valor -1. O algoritmo está definido para um máximo de 20 objectos, por ser um número considerado suficiente para a aplicação em questão, mas poderá ser

superior. Deste modo a matriz de custo terá um tamanho de 20x20, no máximo. Caso o número de objectos detectados e previstos seja igual, a correspondência é balanceada, não existindo, no entanto, certeza acerca da verdadeira correspondência encontrada. Deste modo, é efectuado um teste ao valor do coeficiente de similaridade S_{ij}, de modo a verificar se o mesmo poderá fazer sentido. Ou seja, para valores de $S_{ij} > \tau, \tau \cong \frac{4}{T}$, os objectos i e j não podem ser considerados correspondentes.

De $k = 1$até i fazer:
    De $l = 1$ até j fazer:
        Se $AM(i,j) = 1$ então:
            Se $CM(k,l) \leq \tau$ então:
                $\Phi(k) = \Phi(l)$
                Se $C(k) < \frac{1}{T}$ então:
                    $C(k) = C(k) + 1$
            Senão:
                Se $(x_k^c > M \vee x_k^c < 1) \wedge (y_k^c > N \vee y_k^c < 1)$ então:
                    $ID(k) = -1$
                    $C(k) = 0$
                Senão:
                    $C(k) = C(k) - 1$
                    Se $C(k) \leq 0$ então:
                        $ID(k) = -1$
                Encontra um ID vazio entre 1 e 20, $newID$
                $ID(newID) = \Phi(l)$
                $C(newID) = 1$
    Fim
Fim
Se $i > j$
    De $k = 1$ até i fazer:
        $$assignment = \sum_{l=1}^{j} AM(k,l)$$
        Se $assignment = 0$ então:
            $C(k) = C(k) - 1$
    Fim

Senão, se $i < j$
    De $l = 1$ até j fazer:
        $$assignment = \sum_{k=1}^{i} AM(l,k)$$
        Se $assignment = 0$ então:
            Encontra um ID vazio entre 0 e 20, $newID$
            $ID(newID) = \Phi(l)$
            $C(newID) = 1$
    Fim

Algoritmo 3.3.2: Verificação e correcção de correspondência entre objectos.

Um deles desapareceu, aparecendo outro num local diferente, ou então, dois objectos fundiram-se, dando origem a uma oclusão, tendo aparecido um novo objecto noutro local. Neste caso, o nível de confiança do objecto i será diminuído em uma unidade e assume-se que a sua posição real é igual à posição prevista, tal como se as condições se mantivessem inalteráveis. Ou seja, o objecto i continuará com a mesma aceleração que tinha desde a última vez que foi detectado. Quanto ao objecto j, este será considerado um objecto novo, sendo fornecido um coeficiente de confiança igual à unidade e activado um filtro de *Kalman*. Sempre que um objecto abandona a zona correspondente às dimensões da imagem, a sua posição é mantida constante, até que o nível de confiança aumente, por correspondência efectiva, ou então chegue a 0, sendo, portanto, eliminado. No caso da matriz de correspondência ser não balanceada o procedimento é semelhante, com a excepção de que já existem objectos previstos não correspondidos e objectos novos não correspondidos. O procedimento para esses objectos é no entanto igual ao caso balanceado, quando são detectadas situações de não correspondência.

Apenas os objectos que possuam o nível de confiança máximo são tomados em consideração pelo módulo de reconhecimento e análise de comportamento. Esta abordagem permite que, mesmo durante a oclusão, os objectos possam ser seguidos, até que esta cesse, o que acontece, normalmente, logo de seguida. No entanto, caso a oclusão seja muito demorada, o objecto irá ser descartado, sendo considerado novo, logo que volte a aparecer.

3.4 Implementação

A implementação do módulo de seguimento de vários objectos foi efectuada no Simulink, tal como a Figura 3.4.1 pretende demonstrar. Com o propósito de testar correctamente este estágio do sistema, foram geradas trajectórias aleatórias com círculos de áreas e acelerações variáveis, afectadas por ruído branco gaussiano. Esta metodologia permitiu provocar oclusões, entrada e saída de objectos no plano de imagem, fornecendo os resultados que serão apresentados na próxima secção. Na Figura 3.4.1 podem ser observados os vários módulos de processamento que constituem o seguidor de objectos. Existem 4 linhas de entrada, sendo que, a primeira corresponde ao mapa binário de *foreground* e as seguintes à imagem no formato RGB. Tal como se pode verificar, o algoritmo recebe apenas a imagem binária, sendo as restantes linhas de entrada apenas necessárias para marcar a imagem real com as *bounding boxes* e os centros de massa dos objectos.

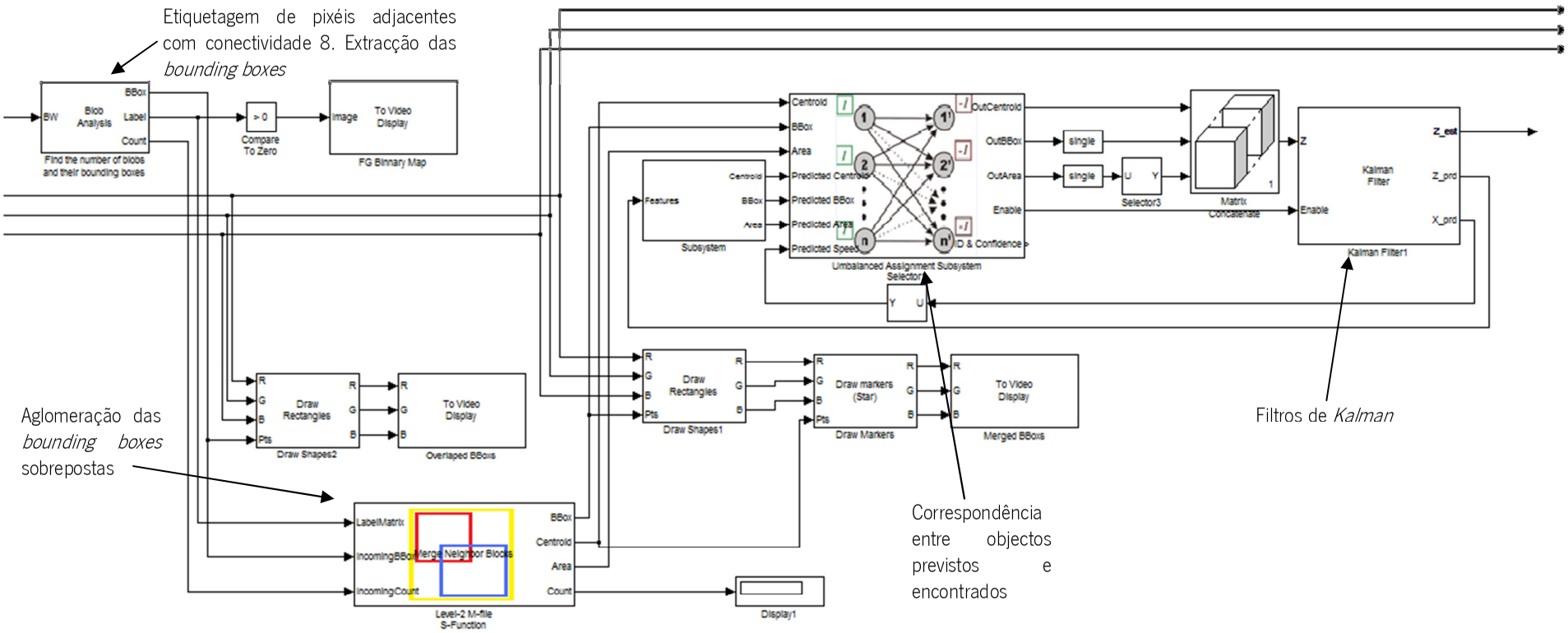


Figura 3.4.1: Diagrama de blocos correspondente ao seguidor de vários objectos com tratamento de oclusões, entradas e saídas, implementado no *Simulink*.

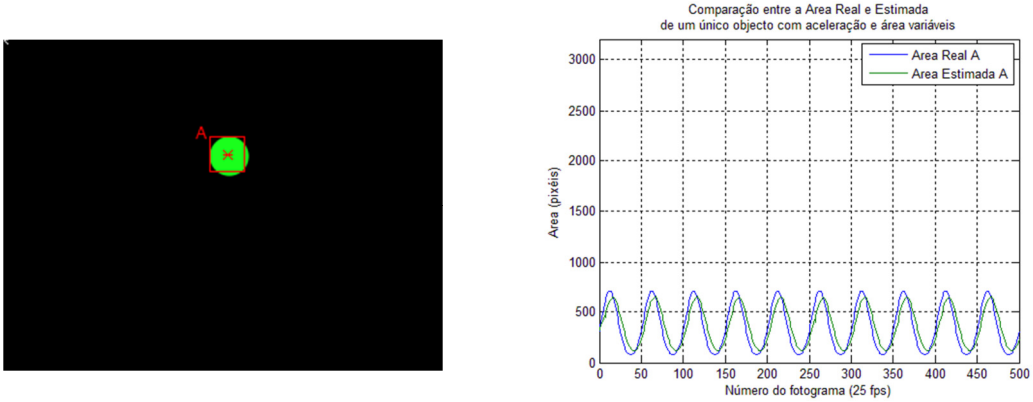
3.5 Resultados Experimentais

3.5.1 Dados sintéticos

Os testes realizados com dados sintéticos foram protagonizados por objectos e trajectórias geradas aleatoriamente, durante 20 segundos, a uma taxa de 25 fotogramas por segundo, sob diferentes condições. Os testes tiveram como objectivo a verificação das diferenças entre as verdadeiras posições dos centros de massa dos objectos e as posições estimadas pelo seguidor:

1. Um único objecto com aceleração e área variáveis, sujeitas a ruído branco gaussiano:

Na Figura 3.5.1 (a) pode observar-se uma amostra do objecto circular durante o seu deslocamento, sendo identificado pela letra "A". Para cada valor do ID, o seguidor atribui letras de forma sequencial, de tal modo que a 1 corresponde a letra "A", a 2 a letra "B" e assim sucessivamente.

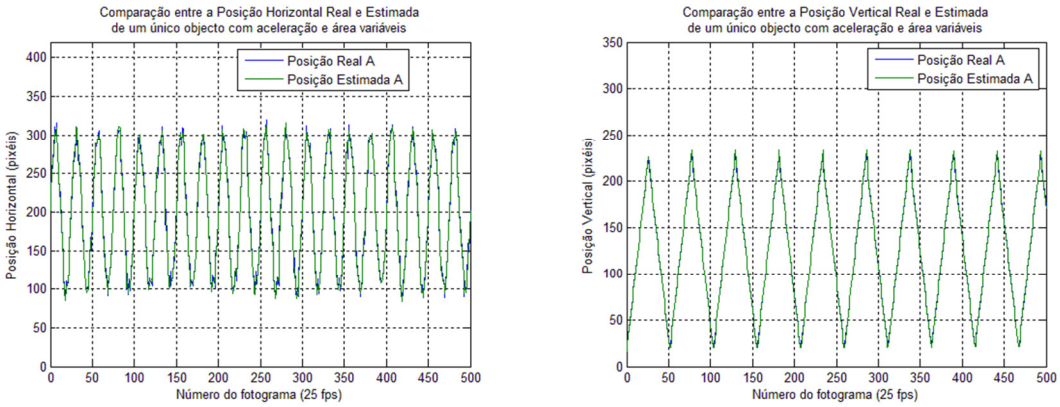


a) b)

Figura 3.5.1: Deslocamento de um único objecto com aceleração e área variável sujeitas a ruído.

a) Imagem gerada;
b) Gráfico da área real e da área estimada do objecto.

No gráfico da Figura 3.5.1 (b) verifica-se que, a área estimada pelo seguidor e a área real do objecto apresentam uma diferença diminuta. Em baixo, os gráficos da Figura 3.5.2 (a) e (b) permitem concluir que, mesmo com acelerações variáveis e sujeitas a ruído, o seguidor estima a posição do objecto com uma elevada precisão. No entanto, este é também o caso mais simples, pois só existe um objecto, sendo o filtro de *Kalman* o único interveniente na estimação da posição do objecto. Os testes mais interessantes, do ponto de vista dum seguidor de vários objectos, são efectuados com recurso a vários objectos, de forma que possam existir oclusões e verificar a verdadeira capacidade do algoritmo em lidar com as mesmas.



a) b)

Figura 3.5.2: Gráficos da posição horizontal e vertical do centro de massa do objecto.

a) **Posição horizontal x^c;**

b) **Posição vertical y^c.**

2. Dois objectos, um deles com aceleração variável e outro com aceleração constante, com existência de oclusão:

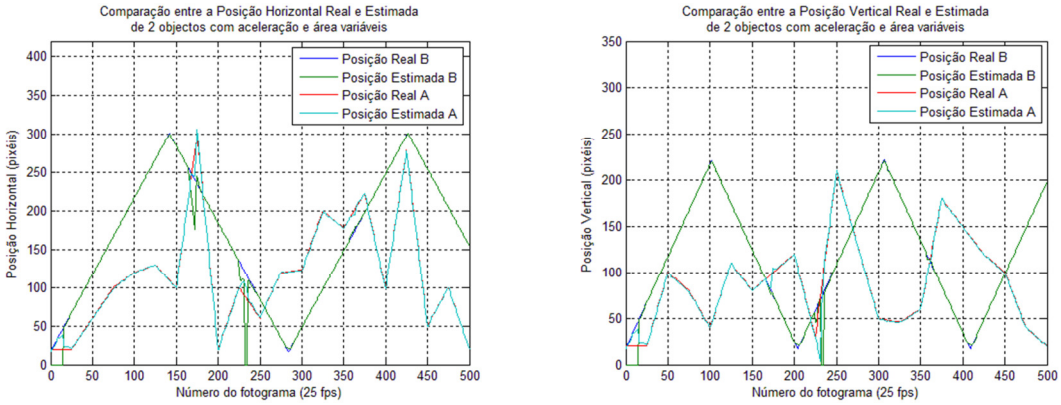
Na Figura 3.5.3 (a) e (b) podem observar-se oclusões parciais entre os objectos "A" e "B". Repare-se que, tal como foi explicado anteriormente, na ocorrência deste tipo de acontecimento, o seguidor assume que o objecto tapado segue a mesma trajectória com a última velocidade estimada, até que o objecto seja de novo visível.

a) b)

Figura 3.5.3: Oclusões entre dois objectos.

Nos gráficos da Figura 3.5.4 encontram-se os resultados obtidos com a simulação de dois objectos com diferentes trajectórias e diferentes variações de velocidade e de área. Neles pode ser visível o surgimento da oclusão entre os fotogramas 150 e 200 e a extinção da mesma entre os fotogramas

200 e 250. Repare-se que, apenas existem diferenças entre as posições estimadas e medidas dos objectos no momento da junção e da separação das regiões.



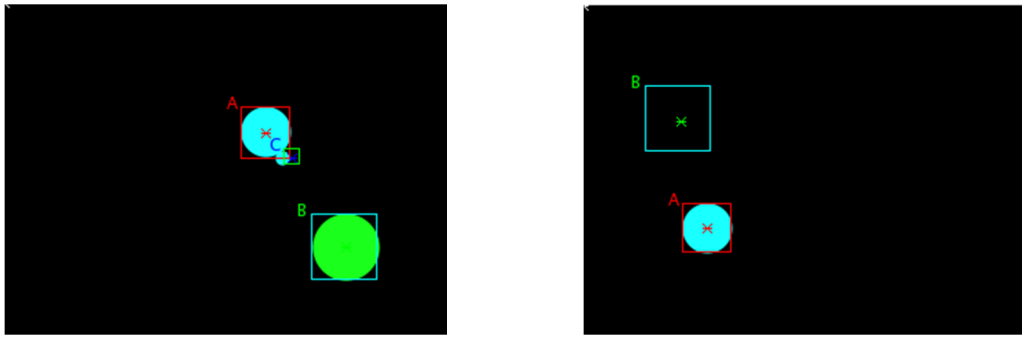
a) b)

Figura 3.5.4: Gráficos da posição do centro de massa de 2 objectos com ocorrência de oclusões.

a) Posições horizontais x^c;
b) Posições verticais y^c.

3. Vários objectos, uns com aceleração constante e outros com aceleração variável, com existência de oclusões, entradas e saídas:

Este terceiro teste é uma extensão do anterior, comportando agora o aparecimento e desaparecimento de objectos na cena. Durante a ocorrência destes acontecimentos é importante que os objectos identificados não percam a sua identificação. Além disso, é também importante garantir que, durante um certo período de tempo, o objecto que desapareceu continue a ser identificado, pois poderá ter desaparecido momentaneamente, causado por erros no módulo de segmentação.

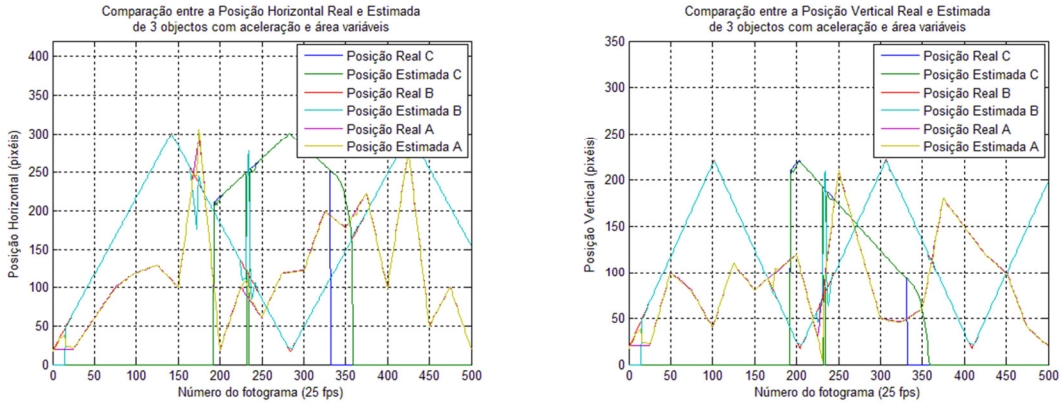


a) b)

Figura 3.5.5: Vários objectos com oclusões, entradas e saídas.

a) Objectos com inicio de uma oclusão;
b) Desaparecimento do objecto "B".

Em baixo, nos gráficos (a) e (b), da Figura 3.5.6, são apresentados os resultados das posições medidas e estimadas, pelo segmentador, nas condições mencionadas. À semelhança do teste anterior, existem também oclusões entre os objectos "A" e "B". Mas além disso, existe também o aparecimento do objecto "C" um pouco antes do fotograma 200, sendo de imediato identificado pelo segmentador e seguido, tal como se pode verificar pela análise dos gráficos, curvas verde e azul.



a) b)

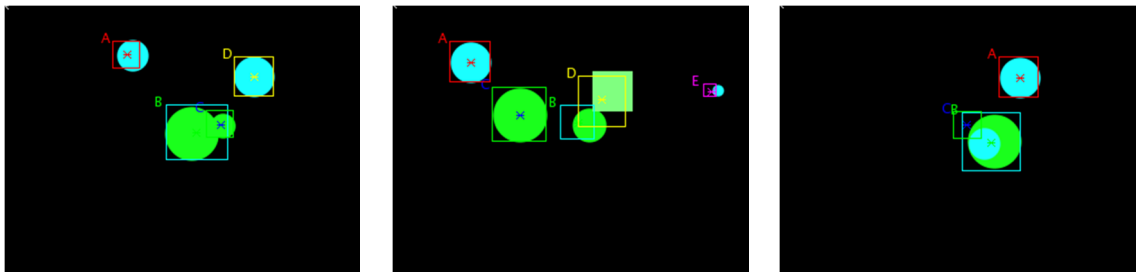
Figura 3.5.6: Gráficos da posição do centro de massa de 3 objectos, com ocorrência de oclusões, entradas e saídas.

a) Posições horizontais x^c;
b) Posições verticais y^c.

No entanto repare-se que, um pouco antes do fotograma 350, o objecto "C" volta a desaparecer. Este desaparecimento, como seria de esperar, não tem quaisquer efeitos nas posições estimadas dos objectos "A" e "B". A trajectória estimada a partir desse momento é que permanece diferente da trajectória real, que não existe mais. Como o objecto não voltou a aparecer, a sua trajectória manteve-se até este ultrapassar os limites da imagem e ser considerado definitivamente perdido.

Foram ainda retiradas amostras, Figura 3.5.7, numa situação ainda mais complexa, com um número de objectos elevado, sendo muito provável a ocorrência de oclusões. Na imagem (a) e (b) observam-se oclusões entre os objectos "B" e "C" e entre os objectos "B" e "D", respectivamente. O caso apresentado na amostra (c) evidência uma oclusão total do objecto "C", provocada pelo objecto "B". Repare-se que não existe informação acerca das cores dos objectos, pelo que, a dificuldade em estimar a trajectória de um objecto totalmente escondido é obviamente muito elevada. Neste caso, verifica-se que, a trajectória estimada é um pouco diferente da trajectória real,

porque o objecto "C" tem aceleração variável com ruído gaussiano, o que faz com que a posição varie consideravelmente, não sendo possível prever, de forma nenhuma, a mesma. No entanto, quando a oclusão terminar, a trajectória será imediatamente corrigida, coincidindo de novo com a trajectória real.



a) b) c)

Figura 3.5.7: Vários objectos com existência de oclusões parciais e totais.

a) **Oclusão entre os objectos "B" e "C";**
b) **Início da oclusão entre os objectos "B" e "D";**
c) **Oclusão total do objecto "C" pelo objecto "B".**

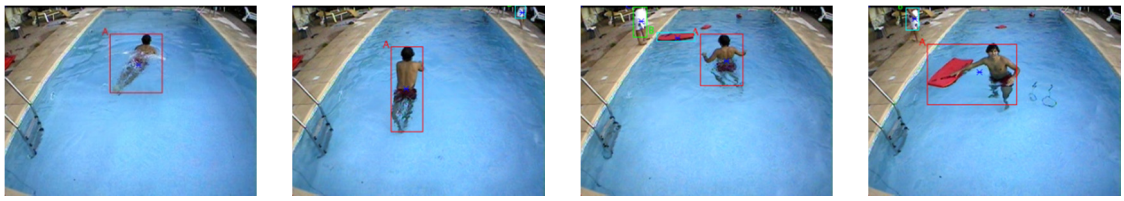
Os testes com as trajectórias geradas aleatoriamente demonstram a elevada capacidade do algoritmo na gestão de oclusões totais e parciais, bem como entrada e saída de objectos em cena. Estes requisitos do seguidor são essenciais para que as características dos vários objectos sejam coerentes ao longo do tempo e possam ser utilizadas pelo módulo seguinte do *pipeline*. Na próxima secção, o algoritmo foi posto à prova em condições reais, já no ambiente da piscina, em cenas segmentadas pelo módulo descrito na secção 2.7.

3.5.2 Dados reais

Os testes reais foram realizados recorrendo às imagens gravadas, utilizadas também na Secção 2.9.2. Na Figura 3.5.8, em baixo, é possível verificar algumas amostras de uma dessas gravações, sendo a mesma muito rica, no que diz respeito a junções e separações entre diferentes objectos. O algoritmo de seguimento de vários objectos foi utilizado para extrair as características dos diferentes objectos encontrados pelo módulo de segmentação nestas imagens.

Os gráficos presentes na Figura 3.5.9 permitem concluir que, apesar do aparecimento de outros objectos na cena, não houve perda de identidade do objecto "A", ou seja, o indivíduo no interior da piscina permaneceu todo o tempo com a identificação "A". Para isso atente-se no fotograma 320, Figura 3.5.8 (a), quando um novo indivíduo aparece na cena. Ao analisar ambos os gráficos, torna-

se claro que, os valores da posição horizontal e vertical continuam coerentes, após o evento mencionado.

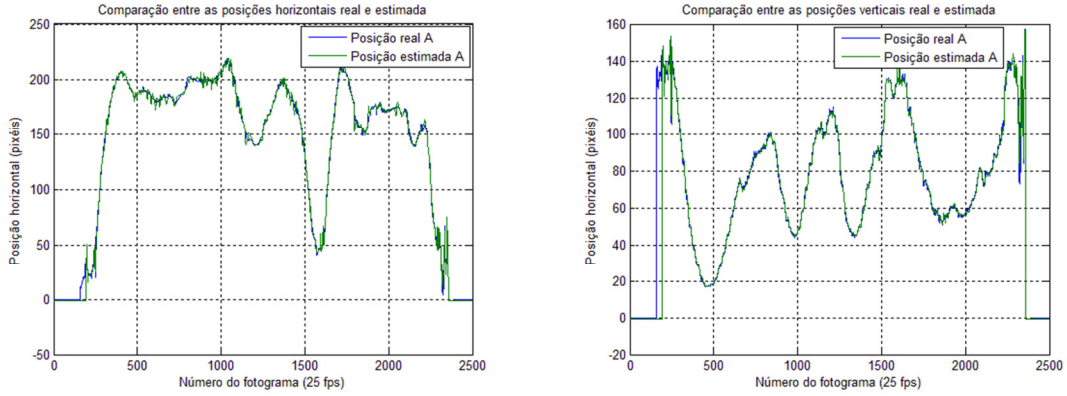


a) b) c) d)

Figura 3.5.8: Testes reais com vários indivíduos e objectos na cena.

a) **Indivíduo sozinho a nadar, (fotograma 320);**
b) **Um indivíduo na piscina e outro no exterior (fotograma 1234);**
c) **Um indivíduo na piscina, outro no exterior e alguns objectos à superfície da água (fotograma 1782);**
d) **O indivíduo na piscina pega num dos objectos, enquanto o indivíduo no exterior observa (fotograma 2165).**

Contudo, no fotograma 2165, Figura 3.5.8 (d), devido à junção prolongada de dois objectos, o objecto "A" demonstrou algumas variações bruscas, tendo como efeito, o desaparecimento do objecto "B". Este desaparecimento ocorre devido à perda de confiança no objecto "B", por ter desaparecido durante um longo período de tempo.



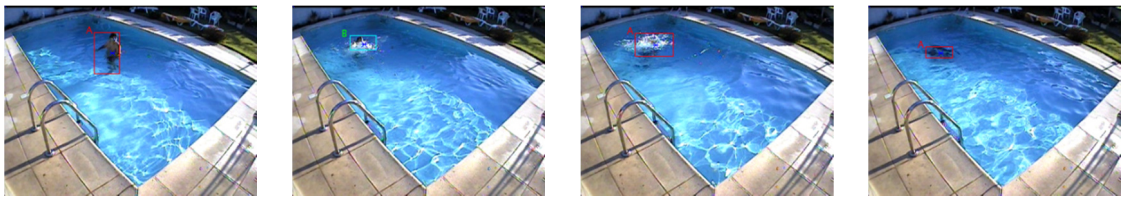
a) b)

Figura 3.5.9: Gráficos das posições verticais e horizontais, estimadas e reais, do objecto "A" presente na gravação, cujas amostras são apresentadas na Figura 3.5.8.

a) **Posição horizontal x^c;**
b) **Posição vertical y^c.**

As amostras da Figura 3.5.10 correspondem a uma gravação de uma cena que simula um afogamento silencioso de um indivíduo. Tal como se pode verificar nas imagens (b) e (c), existe um elevado nível de salpicos de água, devido aos esforços do nadador para se manter à superfície da água. Este tipo de acontecimento provoca variações consideráveis nas características do objecto,

tal como se pode analisar pelos gráficos, da posição do centro de massa do mesmo, presentes na Figura 3.5.11.

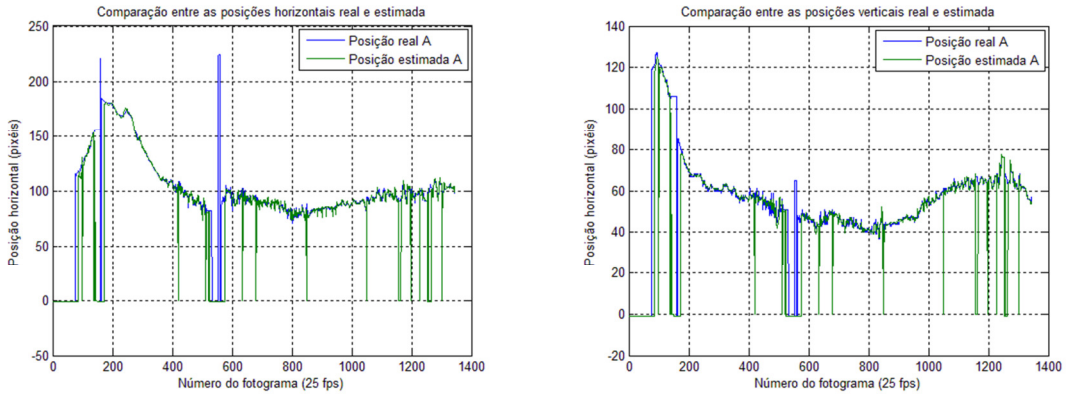


a) b) c) d)

Figura 3.5.10: Testes reais que simulam o afogamento silencioso.

a) Indivíduo sozinho a andar no interior da piscina, (fotograma 305);
b) Troca do objecto "A" pelo "B" (fotograma 521);
c) Nova troca, agora de "B" para "A" (fotograma 578);
d) O indivíduo perdeu a consciência, devido ao afogamento (fotograma 1077).

Estas variações contribuem para o aumento da probabilidade de ocorrência de trocas de objectos, tal como aconteceu nos fotogramas 521 e 578. Apesar de só existir um indivíduo na imagem, o módulo de segmentação, por não ser completamente imune ao ruído, por vezes gera algumas regiões que são falsos positivos. Se estas regiões estiverem suficientemente próximas da região identificada, podem ser confundidas com a mesma, por terem um factor de similaridade elevado.



a) b)

Figura 3.5.11: Gráficos das posições verticais e horizontais, estimadas e reais, do indivíduo presente na gravação, cujas amostras são relativas à Figura 3.5.10.

a) Posição horizontal x^c;
b) Posição vertical y^c.

Deste modo, nas zonas do gráfico onde a posição estimada é nula, significa que a região mudou de identificação. É de extrema importância que estes fenómenos sejam reduzidos ao mínimo, caso

contrário é muito difícil classificar o comportamento dos objectos com base nas suas características ao longo do tempo.

3.6 Discussão

Analisando os resultados obtidos pelo algoritmo, tanto com os dados sintéticos, mas principalmente com os dados reais, torna-se claro que, nem sempre o algoritmo cumpre com os requisitos impostos, mas ainda assim mantém resultados aceitáveis no âmbito global do sistema. Por vezes existem trocas de identidade entre objectos devido a erros de segmentação. O seu funcionamento não é perfeito, uma vez que, em determinadas situações, tais como mudanças de trajectória dos objectos durante a ocorrência de oclusões, podem efectivamente provocar trocas de identidade. Além disso, quando um objecto sofre uma oclusão total durante muito tempo, o algoritmo considera que o mesmo desapareceu, podendo ser-lhe fornecida uma nova identificação quando reaparecer.

Este tipo de problemas não é fácil de resolver com apenas uma câmara monocular e fixa. O deslocamento dos objectos é incerto e além disso, quando existem oclusões totais não é possível prever a localização dos mesmos, pois não existe informação suficiente. A utilização de mais que uma câmara pode efectivamente resolver o problema das oclusões, pois a probabilidade do objecto tapado não ser visível em nenhuma câmara é praticamente nula, partindo do princípio de que as câmaras estão correctamente posicionadas. Os histogramas espaciais, correspondentes às áreas dos objectos na imagem real, poderiam ser utilizados, extraindo-se a partir dos mesmos a distância de *Bhattacharya*, servindo esta como mais uma métrica de similaridade. No entanto deverá existir precaução na sua utilização, uma vez que estaríamos a introduzir de novo ruídos causados pelas variações de luminosidade, um problema já eliminado pelo segmentador. A utilização de uma janela temporal no processo de correspondência entre objectos poderá ser a melhor abordagem, para garantir que as características são coerentes ao longo do tempo, não permitindo também a troca de identidade entre objectos.

Por último, salienta-se a necessidade de um seguidor que cumpra com os requisitos impostos. A troca de identidade não pode acontecer frequentemente, pois trará muitos problemas para o módulo seguinte, que efectua a análise de comportamento. Estes problemas estão relacionados com a mudança repentina nas características dos objectos, fazendo com que a detecção de padrões varie e por conseguinte a classificação dos mesmos varie, variando o comportamento inferido. Como exemplo imagine-se que o objecto A se encontra parado e o objecto B em

movimento. Caso exista uma troca entre A e B, para o módulo de análise de comportamento o objecto A está em movimento e o B parado, o que não é de todo verdade. Caso A fosse uma pessoa em aflição antes da troca, depois da troca existem informações que indicam que se move, não estando mais em aflição, sendo interrompida a detecção de afogamento. Neste âmbito, considera-se fundamental a resolução de alguns dos problemas do seguidor concebido, de modo a aumentar a eficácia e a robustez do sistema completo.

Capítulo 4

4 Reconhecimento de Objectos e Análise de Comportamento

4.1 Introdução

O último módulo do sistema de detecção de afogamento corresponde ao reconhecimento de objectos e à análise de comportamento dos mesmos. Ao ser detectado um risco potencial de afogamento, o módulo deve gerar um alerta com vários níveis de intensidade, de acordo com a evolução do processo de afogamento. O presente capítulo começa por destacar as abordagens utilizadas no âmbito da detecção de afogamento em piscinas públicas, nos sistemas já destacados em capítulos anteriores. O reconhecimento de padrões e a classificação são uma área, no campo da vídeo vigilância, que, actualmente, maior foco tem por parte da comunidade científica. Uma breve discussão acerca dos requisitos do módulo em questão é depois levantada. Seguidamente é apresentada uma revisão da literatura associada ao afogamento, sendo enumerados os vários estágios neste processo. Com base nos comportamentos aferidos, são depois determinados os descritores de cada objecto que serão utilizados nos algoritmos de reconhecimento de objectos e de análise de comportamento. Por último são extraídos os dados de treino necessários para a modelação estatística de padrões necessária para a classificação de comportamento, sendo elaboradas algumas conclusões acerca dos mesmos.

4.2 Estado da arte no reconhecimento e análise de comportamento para detecção de afogamentos

No sistema DEWS, de modo a detectar o afogamento de nadadores em piscinas públicas, os autores (Eng, Toh, Yau, & Wang, 2008) utilizaram uma abordagem baseada numa máquina de estados finita FSM (*Finite State Machine*), cujas transições de estado assentam num HMM (*Hidden Markov Model*). As regras para as transições de estado são introduzidas manualmente no sistema, de acordo com o conhecimento de peritos em relação à forma como o afogamento ocorre. Deste modo, o módulo de análise de comportamento é composto por um descritor de características dos objectos, um módulo de fusão de dados e por último, um HMM.

Para definir as características de cada indivíduo seguido pelo sistema, os autores utilizaram seis variáveis capazes de, no seu conjunto, permitirem a distinção de uma situação de afogamento ou situação de aflição, de uma situação normal. Assim, foram utilizadas as seguintes variáveis: Amplitude de movimento, que mede a velocidade do centro de massa do objecto; Produto da velocidade, que descreve o padrão de movimento ao comparar a direcção da velocidade anterior e actual; Variação da postura, que se refere ao ângulo entre o eixo maior da elipse, que contém o objecto, e a horizontal; Variação de actividade, cujo objectivo consiste em detectar a forma da elipse, de acordo com a relação entre o eixo maior e menor; Variação de tamanho, que relaciona a diferença entre a área mínima e a máxima com a média, numa janela temporal; Índice de submersão, que mede a saturação na área correspondente à silhueta.

Para fundir os vários dados, provenientes destas variáveis, os autores utilizaram um classificador RM (*Reduced Model*), (Toh, Tran, & Srivivasan, 2004). Este classificador assenta, basicamente, no polinómio das entradas correspondentes aos descritores, afectadas de coeficientes de peso. O cálculo dos pesos do polinómio é depois efectuado com recurso à minimização de uma função LSE (*Least-Squares Error*), através de um conjunto de dados de treino, previamente definidos. No caso deste sistema, o classificador deve descobrir três classes distintas nos dados de entrada: aflição e afogamento; andar; e nadar na piscina. Cada uma destas classes corresponde a um estado, dos três possíveis para cada indivíduo detectado.

O modelo de *Markov* é constituído por três nós completamente interligados, representando cada um deles um dos possíveis estados do indivíduo. O problema consiste em determinar a sequência de estados, através da sequência de observações. A distribuição de probabilidade das observações é gaussiana multivariada, sendo utilizada no HMM de modo a inferir a sequência de estados através do algoritmo *Viterbi* (Forney, 1973).

Em (Lu & Tan, 2004) os autores conceberam um módulo de inferência baseado numa máquina de estados finita, regulada por regras baseadas nas $\boldsymbol{M}$ amostras de três variáveis de cada objecto detectado, durante uma janela temporal. Esta janela temporal tem uma dimensão igual a um segundo, para uma taxa de 8 fotogramas por segundo. As três características que descrevem os indivíduos são: Forma, que mede a relação entre o eixo maior e o eixo menor da elipse que contém o objecto; Velocidade de deslocamento, que corresponde à velocidade vertical e horizontal do centro de massa da elipse entre cada fotograma; Variação de tamanho, que corresponde à variância da área medida no período de tempo correspondente à janela temporal.

Os autores definiram três regras que lhes permite detectar um afogamento, baseados nas características dos objectos, medidas ao longo do tempo. A primeira regra consiste na detecção da diminuição da velocidade do indivíduo. A segunda regra testa a verticalidade do corpo, através da medida da relação entre o eixo maior e menor da elipse que contém o objecto. Por último, a terceira regra avalia a existência de movimentos rápidos do corpo do indivíduo. Este conjunto de regras permite a transição entre os diferentes estados. O estado inicial de um nadador é sempre o "Normal", passando para o estado "Possível afogamento/Aflição" caso as regras um e dois sejam satisfeitas. Estando nesse estado, a satisfação da terceira regra faz com que o estado comute para "Afogamento", sendo lançado um alarme, caso o tempo de permanência neste estado seja superior a um valor predefinido.

Mas para que as características possam ser classificadas têm que ser gerados os modelos que definem cada classe, através de dados de treino. Assim, existem dois conjuntos de características, isto é, conjuntos de $\boldsymbol{M}$ valores, correspondentes às $\boldsymbol{M}$ amostras extraídas no espaço da janela temporal. Um deles corresponde à velocidade de deslocamento e à medida da relação entre o eixo maior e menor da elipse que contém o objecto e o outro conjunto de valores corresponde à variação do tamanho do objecto. As distribuições destes conjuntos são geradas, e extraídas as suas componentes, através do algoritmo EM, de modo a representar as mesmas como uma mistura gaussiana. Cada conjunto pode conter duas distribuições distintas, correspondendo aos diferentes estados existentes. Ou seja, quando um nadador se move rapidamente e a sua elipse apresenta uma diferença elevada entre o eixo maior e o menor, o nadador está a nadar normalmente, hipótese 1. Caso contrário, significa que o nadador está com dificuldades, hipótese 2. Os autores elaboraram um esquema que permite detectar o instante de tempo $\boldsymbol{t_c}$ em que esta transição ocorre, através do cálculo do argumento $\boldsymbol{t_c}$ que permite maximizar o *log-likelihood* da razão entre a hipótese 2 e a hipótese 1. Com as classes definidas, o teste efectua-se através da medição da divergência de Kullback-Leibler (Kullback & Leibler, 1951). A descoberta de cada classe permite comutar de estado para um determinado tempo $\boldsymbol{t_c}$.

4.2.1 Discussão

Os autores dos algoritmos de detecção de afogamento para piscinas públicas optaram por uma abordagem baseada numa máquina de estados finita, controlada pela descoberta de padrões específicos nos descritores seleccionados. As abordagens diferem na forma de classificar os padrões, sendo contudo semelhantes, no sentido em que ambas são supervisionadas. Ou seja, é

necessário em ambos os casos treinar o sistema com dados relativos a afogamentos para que o mesmo os possa detectar. A realidade é que um computador não é capaz de detectar um afogamento sozinho, a menos que seja instruído para tal. Ou seja, uma abordagem não supervisionada até é possível, pois é possível detectar padrões não comuns nos dados. No entanto, enveredar por esta metodologia implica que vários dados normais, isto é, sem ocorrer afogamento, sejam armazenados ao longo do tempo, para que, numa situação anormal, de afogamento, o sistema a possa detectar automaticamente. O reconhecimento de padrões anormais efectua-se recorrendo a técnicas de aglomeração de dados, capazes de resumir os mesmos, podendo assim diferenciá-los. Não parece no entanto um caminho a seguir, pelo menos para já, no contexto de um sistema de detecção de afogamento. Pode, porém, ser utilizado num sistema de detecção de comportamentos anormais, que resultem em ameaça.

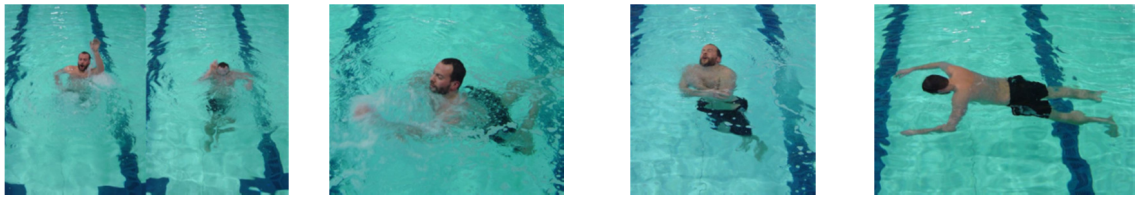
Ambos os autores utilizaram descritores baseados na elipse que engloba o objecto, podendo estes fornecer informações muito importantes, principalmente devido à localização das câmaras de vídeo vigilância. A vista de cima, faz com que seja muito fácil saber se os nadadores se encontram numa posição vertical ou horizontal, pela forma da elipse e reduz de forma significativa o número de oclusões, sendo que oclusões totais são praticamente inexistentes. Além disso, o facto de o contexto ser uma piscina pública, limita o tipo de comportamento dos indivíduos. Ou seja, normalmente nestas piscinas, o tipo de comportamento normal é mais previsível, visto que as pessoas nadam continuamente de uma ponta para a outra, quase sempre entre as linhas. No caso de uma piscina particular os utilizadores podem exibir uma diversidade enorme de comportamentos, tornando difícil distinguir acontecimentos comuns de acontecimentos não comuns. Como exemplo, os utilizadores interagem uns com os outros e com objectos da piscina, podendo mergulhar, chapinhar, atirar água uns aos outros, sem que isso implique que se estejam a afogar. Além disso podem existir outros objectos na superfície da água, tais como bóias, bolas ou colchões, que têm de ser distinguidos das pessoas. Por último, a localização da câmara de vídeo deve ser discreta, não podendo estar no topo da piscina e tendo por isso um ponto de vista mais favorável à ocorrência de oclusões. A situação mais grave verifica-se quando um indivíduo permanece sozinho na piscina, tal como uma criança, à semelhança do que foi mencionado no capítulo introdutório. É neste caso que o sistema tem maior importância, uma vez que é ele o responsável pelo indivíduo.

Em suma, o ambiente associado a uma piscina doméstica é mais agressivo no que respeita ao reconhecimento de padrões de modo a detectar o afogamento. É necessário lidar com uma enorme variedade de comportamentos das pessoas e de diferentes tipos de objectos. A localização das câmaras de vídeo deve ser discreta, não sendo possível, na maioria dos casos, a colocação das mesmas directamente sobre a piscina. O módulo de análise de comportamento deve assim distinguir pessoas de outros objectos presentes na piscina. Deve analisar o comportamento das pessoas, inferindo acerca do seu deslocamento na piscina, e detectar se as mesmas se encontram numa situação de aflição ou se encontram imóveis e afundadas. Nestes dois últimos casos o sistema deve gerar um alerta com diferentes níveis de intensidade correspondentes ao grau de risco associado ao comportamento.

4.3 O Afogamento

O afogamento define-se como sendo "o processo que implica a falta de respiração devido à submersão/imersão num qualquer líquido" (Beeck, Christine, Szpilman, Modell, & Bierens, 2006). O autor (Pia, 1974) refere a existência de dois tipos de problemas associados à interacção de um indivíduo com a água, a aflição e o afogamento. Na maioria dos casos, uma situação de aflição, que pode ser causada por vários factores, é, normalmente, seguida de afogamento. Este tipo de comportamento num indivíduo, quando descoberto, corresponde a um sinal de risco potencial de afogamento, sendo de extrema importância para um salvamento atempado. Num ambiente associado à piscina existem vários tipos de situações que podem causar aflição a um indivíduo, podendo estas culminar com o afogamento do mesmo. Na Figura 4.3.1 mostram-se quatro situações distintas que podem levar ao afogamento. A detecção deste tipo de comportamento é essencial para salvar as vítimas antes mesmo do afogamento acontecer, garantindo assim que, nenhuma lesão grave ocorrerá, devido ao tempo decorrido no estado inconsciente.

O afogamento pode ser causado por várias razões. Assim, de acordo com (Osinski, 2006), destacam-se, a incapacidade do indivíduo flutuar ou nadar, o cansaço excessivo, o pânico ou um trauma na espinal medula, que impossibilite os movimentos. Problemas cardíacos, perda repentina de consciência, hipotermia, laringoespasmo ou sufocamento e intoxicações por álcool ou drogas são também situações que podem terminar em afogamento. Seguidamente são apresentadas as várias fases que constituem o afogamento, associadas a diferentes tipos de afogamento, causados pelos factores anteriormente expostos.



a) b) c) d)

Figura 4.3.1: Vários acontecimentos associados ao risco potencial de afogamento.

a) O nadador "perdeu o pé" e não consegue flutuar;
b) Um nadador fraco, que, devido a algum problema, normalmente cansaço, não consegue nadar de forma eficiente;
c) Indivíduo com alguma lesão, não consegue movimentar-se na água, ou fá-lo exibindo muitas dificuldades;
d) Indivíduo que perdeu a consciência, devido a vários factores, tais como, uma paragem cardio-respiratória.

Existem afogamentos conscientes, nos quais a vítima luta para se manter à superfície da água, até que acaba por aspirar a mesma pelas vias respiratórias e ficar inconsciente. E existem afogamentos nos quais a vítima fica repentinamente inconsciente ou perde os movimentos, devido a alguma lesão. No primeiro caso, a primeira fase do afogamento designa-se por surpresa (Ellis & White, 2000), e corresponde ao indivíduo em dificuldades na água com consciência de que corre grande perigo, num estado de pânico. O comportamento da vítima está associado a uma posição vertical ou muito próxima da vertical, com a face voltada para cima, mexendo os braços de uma forma desordenada, chapinhando, não tendo quaisquer efeitos em termos de deslocamento. O indivíduo afunda-se frequentemente, sendo difícil manter a cabeça à superfície da água. Esta fase é muita curta e tem uma duração que se situa entre os 10 e os 20 segundos (Osinski, 2006). Durante esta fase a vítima não pede socorro nem faz qualquer barulho, visto que a principal preocupação é não permitir a entrada de água para as vias respiratórias. A fase seguinte do afogamento designa-se por suspensão involuntária da respiração e tem uma duração de 30 a 90 segundos. Neste estágio, devido ao afundamento da cabeça, pouco abaixo da superfície da água, a epiglote fecha, em consequência da entrada de água na boca. Sem oxigénio a vítima perde a consciência rapidamente. Está-se assim perante a terceira fase do afogamento, com uma duração de aproximadamente 60 segundos, a vítima permanece imóvel, normalmente, com a face voltada para baixo. A partir desta fase, inclusive, os acontecimentos estão associados aos dois tipos de afogamento mencionados inicialmente. Numa quarta fase, a mais curta de todas, cerca de 5 a 10 segundos, a vítima poderá eventualmente apresentar convulsões hipóxicas, devido à falta de oxigénio no cérebro. A partir deste momento, se a respiração e a circulação sanguínea não forem

restabelecidas dentro de um período de 4 minutos, poderão ocorrer sérias lesões neurológicas e até mesmo a morte.

4.4 Análise de comportamento para detecção de afogamento

O algoritmo concebido para detectar afogamentos é baseado nas descrições de comportamento destacadas na anterior e é composto por quatro fases. Na primeira fase é determinada a posição do objecto detectado relativamente à localização da piscina, de forma a saber se este se encontra ou não na piscina. Objectos que não se encontrem na piscina são ignorados, uma vez que, a probabilidade de se afogarem é praticamente nula. Os objectos restantes são tomados em consideração para efeitos de análise de comportamento. A segunda fase determina se os objectos detectados na piscina apresentam algum deslocamento ou estão parados. Os objectos que apresentem uma deslocação clara na piscina são ignorados, pois esse comportamento não é característico de um afogamento. No entanto, os objectos que se movem muito lentamente ou que estão parados são alvo de uma análise mais refinada, uma vez que este tipo de comportamento pode indicar o início da ocorrência de um afogamento. A terceira fase vai examinar algumas características dos objectos nesta situação e inferir acerca do seu tipo. O objectivo nesta fase consiste em determinar quais dos objectos detectados são pessoas de modo a descartar todos os outros. Por último, a quarta fase analisa o comportamento das pessoas dividindo-o em três tipos, normal, aflição e perda de consciência. Um comportamento de aflição ocorre quando a vítima se encontra na fase de surpresa do afogamento, descrita na Secção 4.3. A perda de consciência é caracterizada pela ausência total de movimento da pessoa, podendo esta permanecer afundada a diferentes profundidades. A situação normal corresponde ao caso contrário, ou seja, às duas últimas situações descritas.

Toda esta análise assenta numa triagem proporcionada por uma árvore de decisão baseada nos padrões detectados nas características de cada objecto. As decisões assentam em vários classificadores *naive Bayes* (Mitchell, 1997) dispersos ao longo da árvore, construídos através da modelação estatística de dados de treino referentes aos vários comportamentos encontrados numa piscina doméstica por diferentes tipos de objectos. Os modelos estatísticos baseiam-se em misturas gaussianas multivariadas com diferentes dimensões nos diferentes ramos da árvore de decisão, de acordo com as características que se pretendem avaliar. Em cada momento são assim inferidos os comportamentos actuais, segundo o mecanismo de decisão mencionado, sendo o estado do objecto mantido por uma máquina de estados finita, que engloba o factor tempo e uma relação

entre o estado anterior e o estado actual do objecto. Cada objecto terá assim uma máquina de estados finita associada.

4.4.1 Descritores de comportamento

Para analisar o comportamento de um objecto é necessário medir determinadas características do mesmo que sejam uma imagem do comportamento que se pretende detectar. Ou seja, as características devem ser distintas para os diferentes tipos de comportamento para que se possa construir um modelo estatístico que permita diferenciar as várias classes. Neste contexto, foram avaliados vários descritores de comportamento, i.e., características dos objectos que permitem concluir acerca dos comportamentos que se pretendem detectar, descritos na parte introdutória da Secção 4.4. As características dos objectos devem ser normalizadas, para que estejam todas na mesma gama e para que, diferenças provocadas pela projecção do mundo real num plano 2D não influenciem o valor das mesmas. Para exemplo considere-se a área de um objecto perto da câmara e do mesmo objecto muito afastado da câmara. Apesar de terem as mesmas dimensões, o objecto longínquo apresenta uma área menor. Contudo, se se medirem os coeficientes de dispersão da distribuição de probabilidade normalizada da área ao invés da própria área essas diferenças provocadas pela projecção são diminuídas ao ponto de poderem ser consideradas irrelevantes. Se todas as características forem normalizadas desta forma, deixam de existir diferenças de escala entre diferentes características. Deste modo, salvo pequenas excepções consideradas posteriormente, são os coeficientes de dispersão das características que são medidos, ao invés dos valores das próprias características.

O índice de permanência na zona da piscina, designado por ϕ_1, mede a razão entre a quantidade de pixéis simultaneamente pertencentes ao mapa binário de *foreground*, $FG(x,y)$, e à máscara binária da localização da piscina, $PM(x,y)$, e a quantidade de pixéis pertencentes ao mapa binário de *foreground*. A Equação 4.4.1 permite determinar este índice e baseia-se nos momentos cartesianos de primeira ordem na área da *bounding box* do objecto de dimensões $M \times N$.

$$\phi_1 = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N}\left(PM(x,y) \wedge FG(x,y)\right)}{\sum_{x=1}^{M}\sum_{y=1}^{N} FG(x,y)}$$

(4.4.1)

Quando este índice for superior a um determinado limite τ entre os 70 e os 85 % é possível admitir que esse objecto se encontra no interior da piscina. Devido à projecção do mundo real num plano,

nem sempre esta medida é válida, uma vez que o *foreground* pode efectivamente ocupar a área da piscina, sem que o objecto esteja no interior da mesma. Contudo, com uma só câmara e não tendo informação acerca das dimensões reais da cena esta é a abordagem possível. Os valores do limite τ foram alcançados experimentalmente.

A velocidade de cada objecto ϕ_2 é calculada a partir da média da posição do centro de massa no fotograma actual e anterior sendo dada pela Equação 4.4.2.

$$\phi_2 = \sqrt{(\mu_x^{t-1} - \mu_x^t)^2 + \left(\mu_y^{t-1} - \mu_y^t\right)^2} \tag{4.4.2}$$

A média da posição do centro de massa corresponde a uma suavização do sinal de modo a absorver variações elevadas resultantes de erros nos processos de segmentação e seguimento. Deste modo, os valores de μ_x e μ_y, num dado instante de tempo t são calculados de acordo com a Equação 4.4.3 e a Equação 4.4.4.

$$\mu_x = \frac{\sum_{n=t-T}^{t} x_n}{T} \tag{4.4.3}$$

$$\mu_y = \frac{\sum_{n=t-T}^{t} y_n}{T} \tag{4.4.4}$$

Os valores de x e y correspondem à posição do centro de massa do objecto, calculados a partir dos momentos de segunda ordem, tal como foi apresentado na Secção 3.3. O valor de T depende da velocidade da captura, devendo-se encontrar entre os 2 e os 4 segundos de vídeo. Para uma velocidade de 25 fotogramas por segundo o seu valor estará entre os 50 e os 100 fotogramas.

O índice de dispersão, ϕ_3, é baseado na variação da medida de circularidade que relaciona o quadrado do perímetro com a área do objecto (Costa & Cesar Jr., 2001). Esta característica está relacionada com a complexidade da forma do objecto. Uma pessoa apresenta por norma um índice de dispersão maior que um objecto rígido, uma vez que a sua forma varia ao longo do tempo. A Equação 4.4.5 permite determinar este índice sabendo a área e perímetro de um objecto ao longo de T fotogramas.

$$\phi_3 = \frac{\sqrt{\frac{1}{T}\sum_{n=t-T}^{t}\left(\frac{P_n^2}{A_n} - \mu_{PA}\right)^2}}{\mu_{PA}}, com\ \mu_{PA} = \frac{1}{T}\sum_{n=t-T}^{t}\frac{P_n^2}{A_n}$$

(4.4.5)

O valor da área, A, é determinado através do momento cartesiano de primeira ordem, presente também no denominador da Equação 4.4.1. O valor do perímetro, P, é determinado a partir da contagem dos pixéis resultantes da aplicação do algoritmo de *Canny* (Canny, 1986) no mapa binário de *foreground* $FG(x, y)$, no interior da *bounding box* de dimensões $M \times N$.

A variação da postura aparente, ϕ_4, é uma medida que permite concluir acerca da variação da relação comprimento largura da *bounding box* que envolve o objecto. Em pessoas esta variação tende a ser elevada, ao passo que em objectos rígidos esta variação tende a ser próxima de zero. Sendo o comprimento da *bounding box* dado por w e a largura por h, a Equação 4.4.6 permite determinar esta característica.

$$\phi_4 = \frac{\sqrt{\frac{1}{T}\sum_{n=t-T}^{t}\left(\frac{w_n}{h_n} - \mu_{wh}\right)^2}}{\mu_{wh}}, com\ \mu_{wh} = \frac{1}{T}\sum_{n=t-T}^{t}\frac{w_n}{h_n}$$

(4.4.6)

A variação da deformação da forma do objecto, ϕ_5, é uma característica que permite analisar a média da distância da periferia do objecto ao seu centro de massa, Figura 4.4.1. Esta característica é determinada com base na distância entre cada pixel pertencente ao contorno do objecto e o seu centro de massa efectuando posteriormente a média das diferentes distâncias. Quando a média desta distância varia estamos perante um objecto cuja forma também varia ao longo do tempo. Esta medida fornece uma imagem acerca do movimento do objecto relativamente a ele próprio, podendo também ser considerada uma medida da complexidade da forma. Em pessoas esta característica apresenta valores substancialmente maiores relativamente a objectos rígidos. A Equação 4.4.7 permite determinar o valor desta característica.

$$\phi_5 = \frac{\sqrt{\frac{1}{T}\sum_{n=t-T}^{t}\left(\mu_\rho^n - \mu_{SD}\right)^2}}{\mu_{SD}}, com\ \mu_{SD} = \frac{1}{T}\sum_{n=t-T}^{t}\mu_\rho^n$$

(4.4.7)

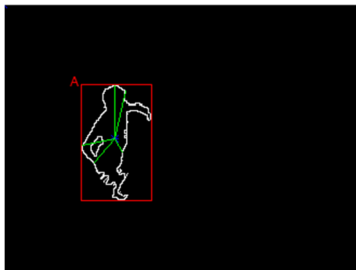


Figura 4.4.1: Medição da média da deformação da forma. O comprimento das linhas de cor verde corresponde à distância do centro de massa do objecto ao respectivo ponto de periferia.

A média da distância de cada pixel pertencente ao contorno do objecto é obtida através da Equação 4.4.8. Considere-se que $P(x,y)$ corresponde aos contornos do mapa binário de *foreground*, que $M \times N$ são as dimensões da *bounding box* que contém o objecto e que x_c e y_c são as coordenadas do seu centro de massa.

$$\mu_\rho = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N}\sqrt{(x-x_c)^2+(y-y_c)^2}P(x,y)}{\sum_{x=1}^{M}\sum_{y=1}^{N}P(x,y)}$$

(4.4.8)

A variação da área, ϕ_6, é mais um indicador da variação da forma de um objecto. Em pessoas este valor é consideravelmente mais elevado, principalmente quando estas estão numa situação de aflição. Em objectos rígidos este índice atinge valores próximos de zero. A Equação 4.4.9 permite determinar o seu valor, sendo A o valor da área do objecto.

$$\phi_6 = \frac{\sqrt{\frac{1}{T}\sum_{n=t-T}^{t}(A_n-\mu_A)^2}}{\mu_A}, com\ \mu_A = \frac{1}{T}\sum_{n=t-T}^{t}A_n$$

(4.4.9)

Por último, a média da componente saturação, ϕ_7, na zona correspondente ao objecto fornecida pelo mapa binário de *foreground* mede o nível de profundidade do mesmo e a quantidade de bolhas presentes na sua proximidade. Esta medida é determinada pela Equação 4.4.10.

$$\phi_7 = \frac{1}{T}\sum_{n=t-T}^{t}\mu_S^n$$

(4.4.10)

Sendo μ_S dado pela Equação 4.4.11, onde $S(x,y)$ corresponde à componente saturação da imagem actual.

$$\mu_S = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N} S(x,y)\cdot FG(x,y)}{\sum_{x=1}^{M}\sum_{y=1}^{N} FG(x,y)}$$

(4.4.11)

Foram ainda exploradas outras características, tais como a variação do fluxo residual, determinada através da velocidade de cada ponto da periferia ou perímetro relativamente à velocidade do centro de massa, ou a média da intensidade da imagem actual na zona pertencente ao objecto. No entanto, estas medidas não apresentavam diferenças significativas, nas diferentes situações de comportamento, pelo que não foram utilizadas.

4.4.2 Inferência Bayesiana e o classificador *Naive Bayes*

Tal como foi anteriormente mencionado, o algoritmo de análise de comportamento baseia-se numa árvore de decisão, onde as várias decisões efectuadas ao longo desta árvore derivam de classificadores *naive Bayes*. A utilização destes classificadores deve-se à necessidade de aprendizagem dos padrões das características mencionadas na Secção 4.4.1. Ou seja, a análise de comportamento trata-se efectivamente de uma classificação supervisionada, a qual, a partir de dados de treino, permite gerar modelos estatísticos representativos dos padrões a detectar. A teoria de decisão *Bayesiana* está na base do classificador *naive Bayes*, sendo uma abordagem estatística fundamental para a solução do problema da classificação de padrões, (Duda, Hart, & Stork, 2000). A escolha do classificador *naive Bayes* está relacionada com a sua simplicidade e bons resultados obtidos na prática para problemas com grandes dimensões de dados, i.e., várias características, mesmo quando comparado com outros algoritmos de aprendizagem, tais como as redes neuronais, (Mitchell, 1997).

O classificador *naive Bayes* deriva da conhecida regra de *Bayes*, Equação 4.4.12, que permite determinar a probabilidade a *posteriori* sabendo a probabilidade a *priori* e a probabilidade condicionada, do inglês, *likelihood*.

$$P(\omega_j|x) = \frac{p(x|\omega_j)P(\omega_j)}{p(x)}$$

(4.4.12)

Onde,

$$p(x) = \sum_{j=1}^{K} p(x|\omega_j)P(\omega_j)$$

(4.4.13)

A densidade de probabilidade $p(x)$ apenas serve de factor de normalização para que o valor da probabilidade a *posteriori* esteja compreendido entre 0 e 1. K corresponde ao número de classes e a probabilidade a *priori* serve para indicar a probabilidade do valor x pertencer à classe j sem qualquer conhecimento extra. A densidade de probabilidade condicionada $p(x|\omega_j)$ define o valor da probabilidade de x caso a classe a que este pertença seja j. No caso em que não exista conhecimento da probabilidade a *priori* para cada classe, como é o caso, são atribuídas probabilidades iguais a $\frac{1}{K}$. O objectivo desta regra consiste em poder determinar de que modo é que o conhecimento proporcionado pelos dados de treino, que dão origem a $p(x|\omega_j)$, influenciam a probabilidade a *priori*, ou seja, a probabilidade do valor x pertencer à classe j.

Este tipo de classificador aplica-se a tarefas de aprendizagem onde o vector de características x é descrito por um conjunto de variáveis ou atributos $\langle \phi_1, \phi_2, \dots, \phi_n \rangle$ e a função objectivo pode tomar um conjunto finito Ω de valores discretos ω_j, sendo o tamanho desse conjunto k, (Mitchell, 1997). Esta abordagem permite classificar novos conjuntos de características através da procura da classe j que maximiza a probabilidade a *posteriori*, dada pela regra de *Bayes*, tal como a Equação 4.4.14 pretende evidenciar.

$$\omega_{MAP} = \arg\max_{\omega_j \in \Omega} p(\omega_j|\phi_1, \phi_2, \dots, \phi_n) =$$

$$= \arg\max_{\omega_j \in \Omega} \frac{p(\phi_1, \phi_2, \dots, \phi_n|\omega_j)P(\omega_j)}{p(\phi_1, \phi_2, \dots, \phi_n)} =$$

$$= \arg\max_{\omega_j \in \Omega} p(\phi_1, \phi_2, \dots, \phi_n|\omega_j)P(\omega_j)$$

(4.4.14)

No entanto, a estimação de $P(\phi_1, \phi_2, \dots, \phi_n|\omega_j)$ é complexa, pois exige um enorme conjunto de dados de treino para contemplar todas as hipóteses em que pode ser desdobrada a probabilidade condicionada pelos vários termos $\langle \phi_1, \phi_2, \dots, \phi_n \rangle$. Deste modo, o classificador *naive Bayes*

assume que os diferentes atributos ou características são independentes, simplificando o problema e reduzindo o termo $p(\phi_1, \phi_2, \ldots, \phi_n|\omega_j)$ ao produto da probabilidade condicionada de cada atributo. Ou seja $p(\phi_1, \phi_2, \ldots, \phi_n|\omega_j) = \prod_{i=1}^{n} p(\phi_i|\omega_j)$, o que permite reescrever a Equação 4.4.14 na forma apresentada pela Equação 4.4.15.

$$\omega_{NB} = \arg\max_{\omega_j \in \Omega} P(\omega_j) \prod_{i=1}^{n} p(\phi_i|\omega_j)$$

(4.4.15)

Onde ω_{NB} corresponde à classe atribuída ao conjunto de características $\langle \phi_1, \phi_2, \ldots, \phi_n \rangle$ pelo classificador *naive Bayes*.

A estimação da densidade de probabilidade $p(\phi_i|\omega_j)$ é efectuada com base na classificação manual dos vários valores do atributo ϕ_i em diferentes situações fornecidas no conjunto de dados de treino $\mathcal{D}_i$, as quais se pretendem diferenciar ou classificar. A partir dos dados de treino são criados histogramas que são depois modelados por uma mistura de gaussianas uni variadas, via algoritmo EM, dando origem à Equação 4.4.16.

$$p(\phi_i|\mathcal{D}_i) = \sum_{j=1}^{K} \alpha_j \frac{1}{\sigma_j \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\phi_i - \mu_j}{\sigma_j}\right)^2}$$

(4.4.16)

Assim sendo, a densidade de probabilidade condicionada para cada atributo ϕ_i e cada classe ω_j é dada pela Equação 4.4.17.

$$p(\phi_i|\omega_j) = \alpha_j \frac{1}{\sigma_j \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\phi_i - \mu_j}{\sigma_j}\right)^2}$$

(4.4.17)

Com os modelos estatísticos gerados a partir dos dados de treino fornecidos pela Equação 4.4.17 é possível determinar a classe à qual pertence uma nova instância de características $\langle \phi_1, \phi_2, \ldots, \phi_n \rangle$ a partir da Equação 4.4.15.

4.4.3 Algoritmo de reconhecimento de objectos e análise de comportamento

Na Figura 4.4.2 apresenta-se a árvore de decisão que permite classificar o tipo de objectos e o seu comportamento, baseando-se nos descritores apresentados na Secção 4.4.1 e nos classificadores apresentados na Secção 4.4.2.

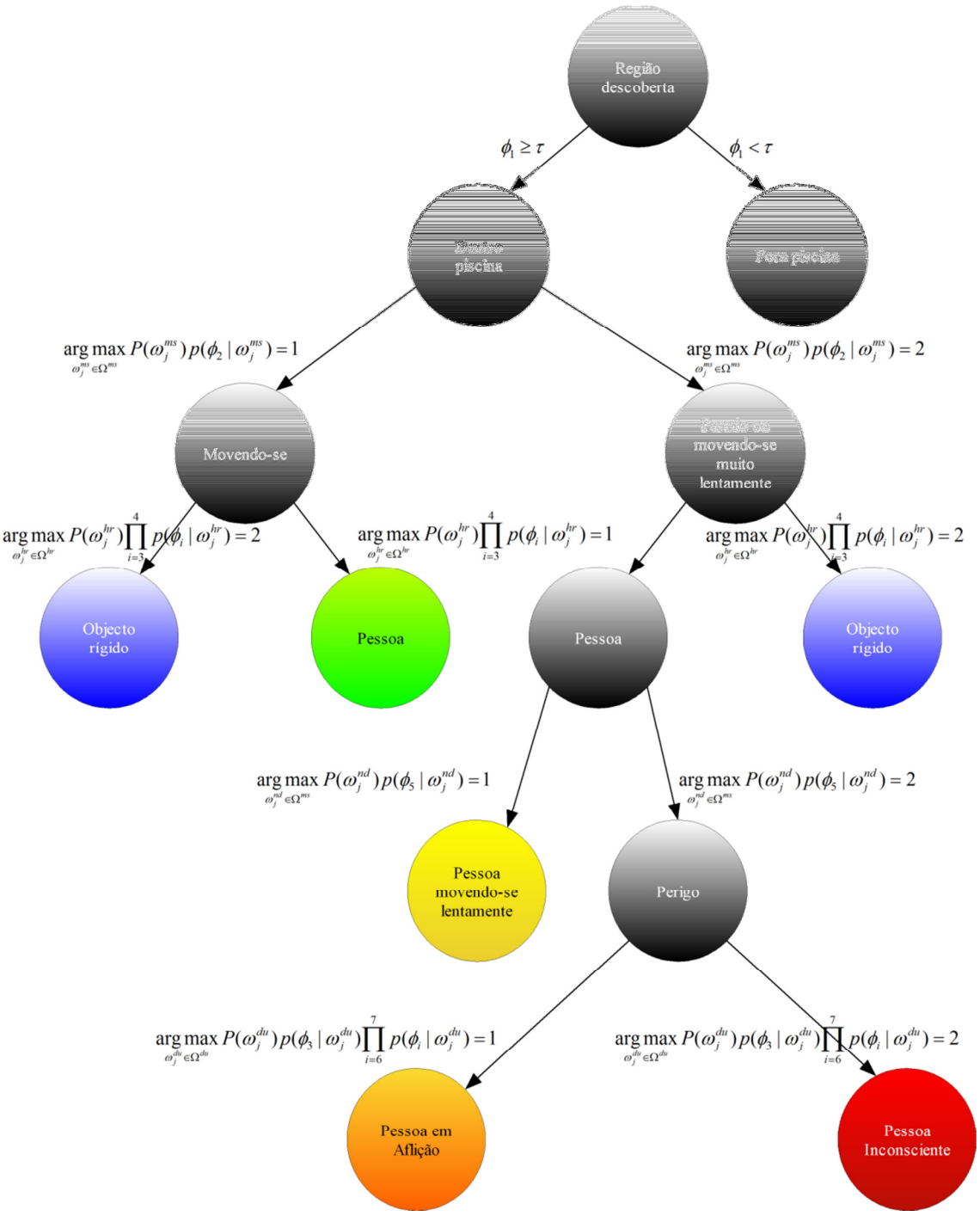


Figura 4.4.2: Árvore de decisão capaz de inferir acerca do tipo de objecto e do seu comportamento, com base nas suas características no instante de tempo t, tendo em conta as várias classes provenientes dos modelos estatísticos criados por intermédio dos dados de treino.

A árvore binária inicia-se com a descoberta de um objecto efectuada pelo módulo de seguimento e a consequente decisão acerca da localização deste. Utilizando o descritor $\boldsymbol{\phi_1}$ e submetendo este a um limite $\boldsymbol{\tau}$, tal como foi mencionado na Secção 4.4.1, é possível saber se o objecto se encontra dentro ou fora da piscina. Objectos que se encontrem fora da piscina são descartados, uma vez que a possibilidade de ocorrer um afogamento fora desta é praticamente nula. Caso o objecto esteja dentro da piscina é efectuada a classificação do mesmo no que diz respeito à sua velocidade de deslocamento. A classificação deriva directamente da teoria de decisão *bayesiana*, sendo calculada a probabilidade a *posteriori* para cada uma das duas classes do conjunto $\boldsymbol{\Omega^{ms}}$ representarem a variável $\boldsymbol{\phi_1}$. A classe $\boldsymbol{\omega_1^{ms}}$ representa velocidades de deslocamento maiores e a classe $\boldsymbol{\omega_2^{ms}}$ representa velocidades de deslocamento muito baixas ou próximas de zero. A Equação 4.4.18 permite efectuar esta classificação, sendo que, caso o valor das probabilidades seja igual decide-se pela classe que representa velocidades de deslocamento baixas. Esta estratégia permite garantir que, em caso de dúvida, se opte por considerar o pior caso, ou seja, o caso que pode levar mais rapidamente ao afogamento.

$$\omega_{NB}^{ms} = \underset{\omega_j^{ms} \in \Omega^{ms}}{\arg\max}\, P\left(\omega_j^{ms}\right) p(\phi_2 | \omega_j^{ms})$$

(4.4.18)

Estando o objecto parado ou em movimento procede-se à análise da sua forma de modo a verificar a variabilidade desta ao longo do tempo. Se a forma variar muito trata-se de uma pessoa, pelo contrário, se a forma variar pouco, considera-se que o objecto não é uma pessoa. Para efectuar esta análise recorre-se à classificação a partir de um classificador *naive Bayes* dado pela Equação 4.4.19.

$$\omega_{NB}^{hr} = \underset{\omega_j^{hr} \in \Omega^{\mathrm{hr}}}{\arg\max}\, P(\omega_j^{hr}) \prod_{i=3}^{4} p(\phi_i | \omega_j^{hr})$$

(4.4.19)

Esta equação permite classificar o conjunto de características $\langle \boldsymbol{\phi_3}, \boldsymbol{\phi_4} \rangle$ em duas classes possíveis pertencentes ao conjunto $\boldsymbol{\Omega^{hr}}$. A classe $\boldsymbol{\omega_1^{hr}}$ representa os objectos do tipo pessoa e a classe $\boldsymbol{\omega_2^{hr}}$ representa os restantes tipos de objectos, ou seja, aqueles que apresentam maior rigidez. Tal como se pode verificar pela árvore esta operação permite descartar os objectos rígidos das pessoas, uma vez que estes não se afogam. Se o objecto for classificado como pessoa e além disso se estiver a

movimentar-se na piscina considera-se uma situação normal, que não exige nenhuma atenção especial. À semelhança do que foi mencionado na Secção 4.3 acerca dos vários tipos de afogamento, quando um indivíduo se encontra numa situação de aflição ou inconsciente não existe deslocamento efectivo. Quando o objecto se trata de uma pessoa e se move lentamente ou está mesmo parado é necessário efectuar uma análise mais aprofundada de modo a verificar o comportamento do indivíduo. Para isso é classificada a variação da deformação da forma $\boldsymbol{\phi_5}$ em função da Equação 4.4.20.

$$\omega_{NB}^{nd} = \underset{\omega_j^{nd} \in \Omega^{nd}}{\arg\max}\, P\left(\omega_j^{nd}\right) p(\phi_5 | \omega_j^{nd}) \tag{4.4.20}$$

Com um conjunto $\boldsymbol{\Omega^{nd}}$ composto por duas classes, designadamente, $\boldsymbol{\omega_1^{nd}}$ e $\boldsymbol{\omega_2^{nd}}$, o comportamento do indivíduo pode ser considerado de risco médio, no primeiro caso, e de risco elevado, no segundo. Risco médio significa que a pessoa não apresenta comportamento de aflição ou inconsciência, apesar de estar parada, ao passo que risco elevado significa que uma destas duas situações pode estar a ocorrer. De modo a avaliar o comportamento de um indivíduo numa situação de risco levado são tomadas em consideração três características $\boldsymbol{\langle \phi_3, \phi_6, \phi_7 \rangle}$. No caso, o classificador *naive Bayes* dado pela Equação 4.4.21 distingue duas classes $\boldsymbol{\omega_1^{du}}$ e $\boldsymbol{\omega_2^{du}}$ pertencentes ao conjunto $\boldsymbol{\Omega^{du}}$. A primeira representa um comportamento de aflição, no qual o indivíduo esbraceja na água tentando não se afundar, sem deslocamento efectivo e causando o aparecimento de bolhas de ar na água à sua volta. A segunda classe caracteriza o comportamento de um indivíduo inconsciente, que não apresenta qualquer movimento, nem de deslocamento nem dos membros do corpo e que está normalmente afundado.

$$\omega_{NB}^{du} = \underset{\omega_j^{du} \in \Omega^{du}}{\arg\max}\, P\left(\omega_j^{du}\right) p(\phi_3 | \omega_j^{du}) \prod_{i=6}^{7} p(\phi_i | \omega_j^{du}) \tag{4.4.21}$$

Todas as distribuições de probabilidade foram geradas a partir de dados de treino retirados de situações simuladas dos vários comportamentos exibidos pela árvore. O processo de concepção destes modelos estatísticos encontra-se descrito na próxima secção.

A árvore de decisão permite, num dado instante de tempo $\boldsymbol{t}$, saber qual o estado de um objecto. No entanto, não existe um histórico acerca dos vários comportamentos que um objecto teve ao

longo do tempo. De forma a inferir o seu comportamento optou-se por uma máquina de estados finita, Figura 4.4.3, que rege as transições de comportamento com base na sequência de decisões gerada pela árvore. Esta máquina de estados apenas existe para objectos do tipo pessoa que se encontrem na piscina, sendo os restantes casos descartados, uma vez que nesses não ocorre afogamento.

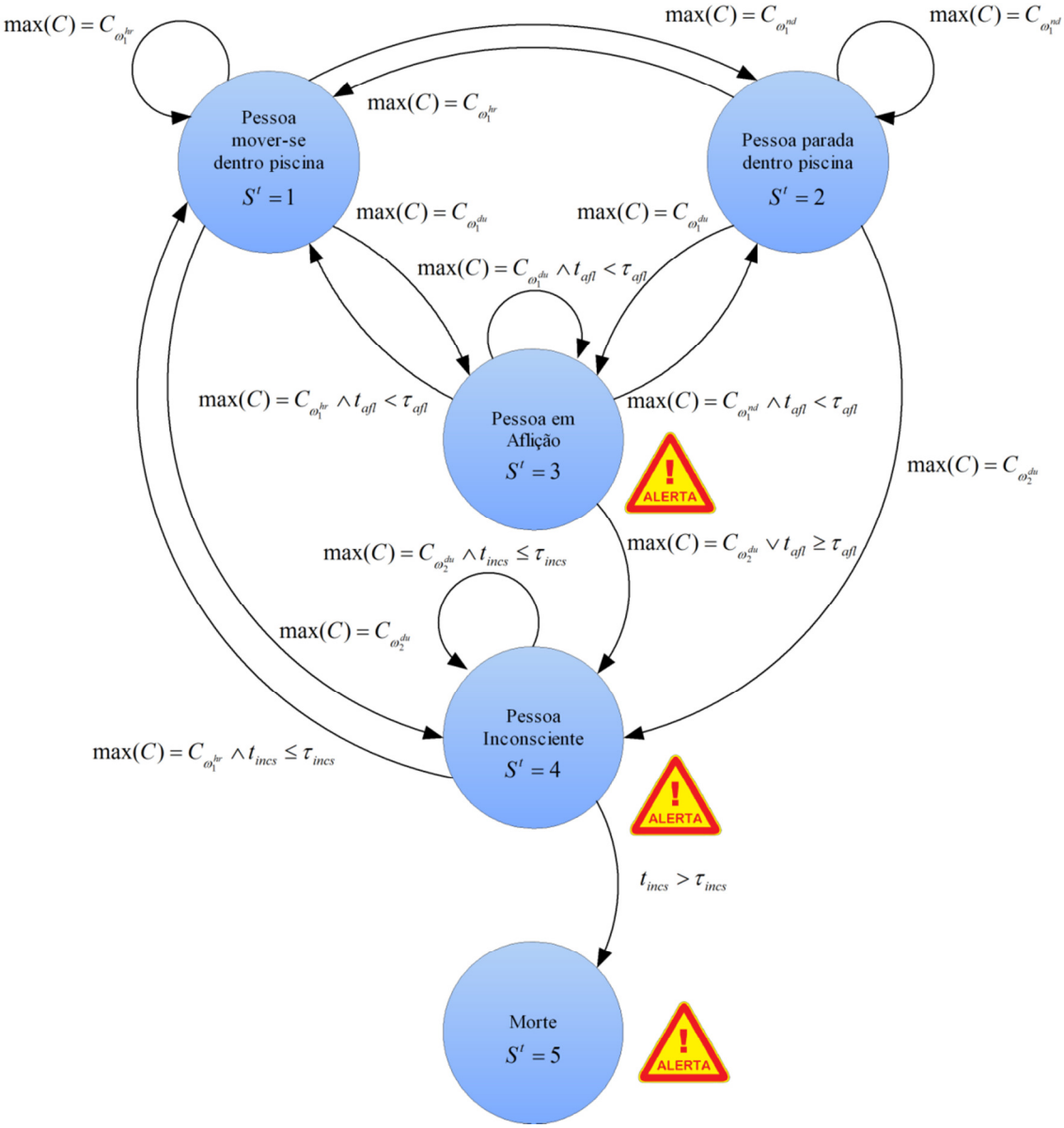


Figura 4.4.3: Máquina de estados finita que define o comportamento de um indivíduo quando este se encontra dentro da piscina.

As saídas relevantes da árvore de decisão são compostas pelo conjunto $\Omega = \{\omega_1^{hr}, \omega_1^{nd}, \omega_1^{du}, \omega_2^{du}, \omega\}$. Cada uma destas saídas descreve, respectivamente, um indivíduo a mover-se na piscina, uma pessoa parada ou movendo-se muito lentamente na piscina,

uma pessoa em aflição, um indivíduo inconsciente e qualquer um dos restantes casos. Para inferir o comportamento de uma pessoa que se encontre na piscina é necessário observar a sequência de saídas proporcionada pela árvore de decisão ao longo do tempo. Sendo T o tamanho do *buffer* que guarda a sequência de decisões da árvore e considerando que a árvore efectua uma decisão a cada fotograma, a uma velocidade de 25 fotogramas por segundo, fazendo $T = 50$, tem-se um período temporal de 2 segundos. Neste período temporal são contadas as frequências de cada um dos 5 tipos de saídas presentes no conjunto $\boldsymbol{\Omega}$, sendo os resultados da contagem de cada tipo guardadas no vector $\mathcal{C} = \left[C_{\omega_1^{hr}}, C_{\omega_1^{nd}}, C_{\omega_1^{du}}, C_{\omega_2^{du}}, C_{\omega}\right]$. A saída mais frequente, ou seja, o valor máximo no vector $\mathcal{C}$ dita a comutação de estado da máquina no instante de tempo t dado por S^t. O conjunto de estados possíveis da máquina é dado por $S = \{1,2,3,4,5\}$ aos quais correspondem os seguintes comportamentos de alto nível: {"*Pessoa a mover-se dentro da piscina*", "*Pessoa parada dentro da piscina*", "*Pessoa em aflição*", "*Pessoa inconsciente*", "*Morte*"}. Com esta topologia, um indivíduo pode apresentar vários comportamentos ao longo do tempo. Assim sendo se uma pessoa se estiver a movimentar na piscina ela pode parar sem que exista perigo, pode ficar imediatamente inconsciente, por exemplo devido a um ataque cardíaco, ou entrar em aflição, marcando este estado o início de um afogamento. Quando uma pessoa se encontra parada na piscina tanto pode retomar o deslocamento, como pode entrar em aflição ou ficar inconsciente. É possível a uma pessoa em aflição voltar para o estado de movimentação normal ou parada na piscina. Este tipo de transição tem como objectivo evitar falsos alarmes quando as pessoas chapinham na água e param dentro de um tempo inferior a τ_{afl}. Este tempo, correspondente a 20 segundos, é aliás o tempo máximo de permanência no estado de aflição, de acordo com os factos apresentados sobre o afogamento na Secção 4.3. Tempos iguais ou superiores a τ_{afl} no estado de aflição implicam a transição do estado da pessoa para inconsciente, sendo a situação considerada um afogamento e accionado um alerta. Uma vez no estado de inconsciência apenas é possível voltar para o estado de deslocamento na piscina, pois indica que pode ter sido um falso alarme. Se o indivíduo se move significativamente tudo indica que o afogamento não ocorreu. No entanto, esta transição só é possível dentro de um período de tempo que não exceda o valor de τ_{incs} que corresponde a 4 minutos. Se a pessoa se mantiver no estado inconsciente por um período igual ou superior a τ_{incs} a comutação para o estado "Morte" acontece não podendo comutar para qualquer outro. Quando um indivíduo entra no estado de aflição é despoletado um alerta de nível 2 que representa perigo de ocorrência de um possível afogamento. Quando um

indivíduo entra num estado de inconsciência é despoletado um alerta de nível 1 que indica claramente a ocorrência de um afogamento, tal como o esquema da Figura 4.4.3 demonstra.

4.4.4 Concepção dos modelos estatísticos

Para construir os modelos estatísticos, isto é, gerar as funções densidade de probabilidade condicionada para cada conjunto de características sabendo a sua classe foi necessário recolher amostras de dados para cada uma dessas características nas diferentes classes. Desta forma, o treino dos classificadores baseou-se na recolha de características de vários objectos presentes em amostras de vídeo que continham os comportamentos que se pretendiam identificar. Recolhidos esses valores, é feita uma classificação manual dos mesmos, agrupando-os em função dos comportamentos identificados nas amostras de vídeo. A partir dessa classificação manual são calculados os parâmetros das misturas gaussianas univariadas que melhor traduzem as distribuições dos dados recolhidos, pois assume-se que as características são independentes.

Na Figura 4.4.4 são apresentadas as amostras do descritor ϕ_2 ao longo de 3000 fotogramas. Esta recolha foi efectuada em objectos rígidos e pessoas a diferentes velocidades de deslocamento. Os valores já estão agrupados no gráfico, sendo cada conjunto recolhido separado pela linha vermelha vertical.

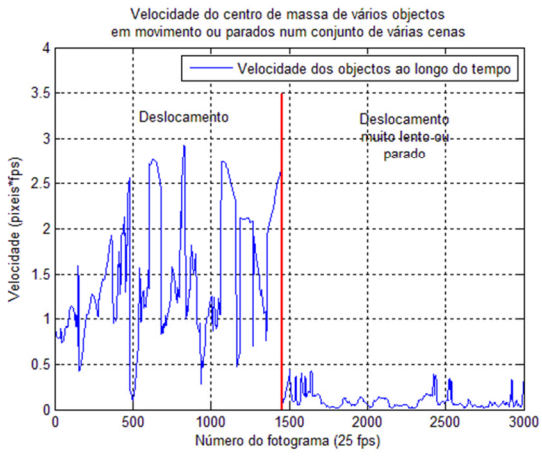


Figura 4.4.4: Velocidade do centro de massa de vários objectos deslocando-se a diferentes velocidades. Valor do descritor ϕ_2 ao longo de 3000 fotogramas. Cada grupo tem aproximadamente o mesmo número de amostras.

A partir destes valores são determinados os parâmetros das gaussianas univariadas que representam cada classe. Assim, para cada grupo de dados apresentado no gráfico da Figura 4.4.4 são calculados a média e desvio padrão com um peso de 0.5 para cada distribuição. Os valores encontrados apresentam-se de seguida:

Classe 1: $\mu_{21}^{ms} = 1.4754, \sigma_{21}^{ms} = 0.6826$

Classe 2: $\mu_{22}^{ms} = 0.1028, \sigma_{22}^{ms} = 0.1391$

Determinados os parâmetros do modelo estatístico é possível definir completamente as densidades de probabilidade utilizadas no classificador *naive Bayes* apresentado na Equação 4.4.18:

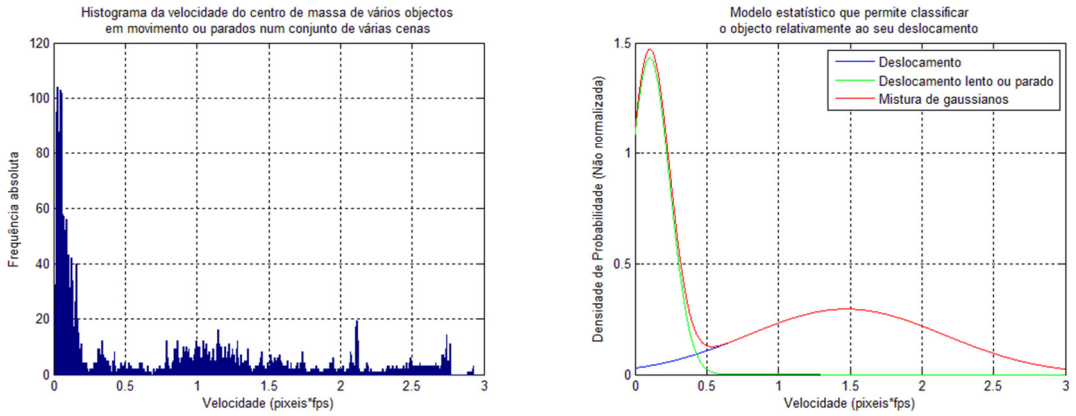
$$p(\phi_2|\omega_1^{ms}) = \mathcal{N}(\mu_{21}^{ms}, \sigma_{21}^{ms})$$

$$p(\phi_2|\omega_2^{ms}) = \mathcal{N}(\mu_{22}^{ms}, \sigma_{22}^{ms})$$

Considera-se igual a probabilidade a *priori* de cada uma das classes uma vez que ambas as classes têm a mesma probabilidade de aparecer, de tal modo que:

$$P(\omega_1^{ms}) = P(\omega_2^{ms}) = 0.5$$

No gráfico (a) da Figura 4.4.5 pode ser visualizado o histograma gerado a partir dos dados recolhidos das amostras que dão origem ao modelo estatístico apresentado no gráfico (b).



a) b)

Figura 4.4.5: Distribuição dos dados da amostra relativa à velocidade de deslocação dos vários objectos e respectivo modelo estatístico.

a) **Histograma da distribuição dos valores da amostra;**
b) **Modelo estatístico composto por duas distribuições gaussianas uni variadas.**

Como se pode verificar, existe uma similaridade elevada entre o histograma e a mistura de gaussianos que o representa indicando a validade do modelo concebido.

O gráfico da Figura 4.4.6 apresenta cerca de 6000 valores de cinco características de vários objectos de diferentes tipos e com diferentes comportamentos. Assim, pode-se distinguir um primeiro grupo de dados relativos a comportamentos normais, de diferentes pessoas, que não

envolvem risco de afogamento. Estes dados vão desde o fotograma 0 ao fotograma 3136. Seguidamente encontra-se um grupo de dados correspondentes a objectos rígidos, onde é notória uma variação baixa dos descritores relativamente às pessoas. Foram ainda capturados dados referentes a cenas simuladas de aflição e inconsciência de pessoas, as quais se encontram entre os fotogramas 4980 - 5531 e 5532 - 5882, respectivamente. Analisando o gráfico notam-se diferenças nos valores das características que evidenciam certos tipos de objectos e os diferentes comportamentos destes. Deste modo, os diferentes modelos estatísticos foram concebidos a partir dos dados apresentados no gráfico da Figura 4.4.6.

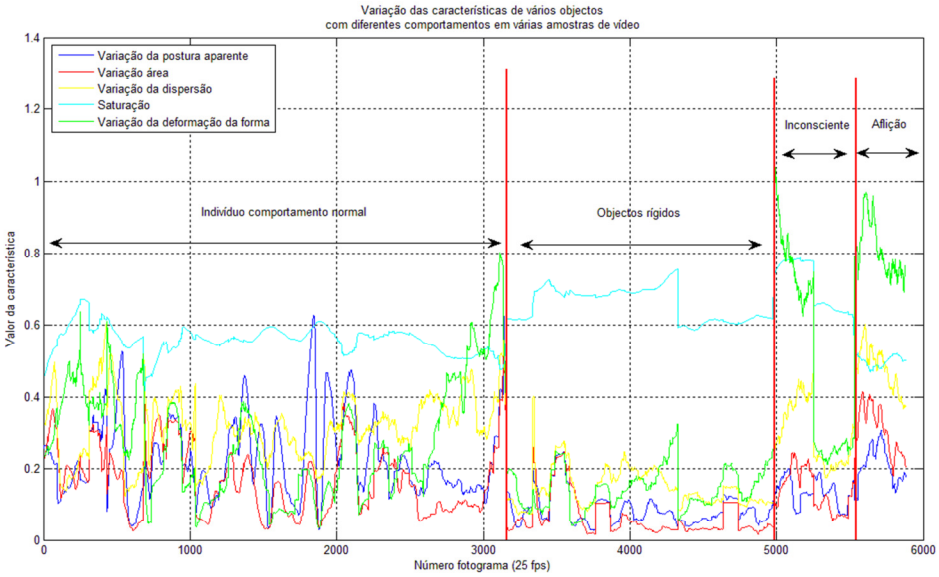


Figura 4.4.6: Amostras de quatro grupos de valores de características ou descritores que representam os comportamentos que se pretendem detectar. Estas amostras correspondem na sua totalidade a cerca de 6000 fotogramas e foram utilizadas como dados de treino dos vários classificadores *naive Bayes* utilizados na árvore de decisão.

A distinção entre objecto rígido e pessoa é baseada nos descritores ϕ_3 e ϕ_4, os quais representam, respectivamente, a variação da dispersão e a variação da postura aparente. Analisando estes valores no gráfico da Figura 4.4.6 é possível concluir que as diferenças existentes são suficientes para formar um modelo estatístico composto por duas gaussianas univariadas para cada característica. À semelhança do que foi feito para distinguir objectos parados de objectos com deslocamento efectivo, os dados foram divididos em dois grupos. Um dos grupos representa as pessoas com comportamento normal ou em situação de afogamento e o outro grupo representa os objectos rígidos. De seguida foram calculados os parâmetros de cada gaussiana representativa de cada grupo para cada uma das características. Os parâmetros do modelo estatístico que

representa as duas classes relativamente à característica variação da postura aparente são dados por:

$$\text{Classe 1: } \mu_{13}^{hr} = 0.2127, \sigma_{13}^{hr} = 0.1047$$

$$\text{Classe 2: } \mu_{23}^{hr} = 0.0889, \sigma_{23}^{hr} = 0.0621$$

Os parâmetros relativos ao modelo estatístico que representa a variação da dispersão são os que se seguem:

$$\text{Classe 1: } \mu_{14}^{hr} = 0.3282, \sigma_{14}^{hr} = 0.0989$$

$$\text{Classe 2: } \mu_{24}^{hr} = 0.1490, \sigma_{24}^{hr} = 0.0613$$

Determinados os parâmetros do modelo estatístico é possível definir completamente as densidades de probabilidade utilizadas no classificador *naive Bayes* apresentado na Equação 4.4.19:

$$p(\phi_3|\omega_1^{hr}) = \mathcal{N}(\mu_{13}^{hr}, \sigma_{13}^{hr})$$

$$p(\phi_3|\omega_2^{hr}) = \mathcal{N}(\mu_{23}^{hr}, \sigma_{23}^{hr})$$

$$p(\phi_4|\omega_1^{hr}) = \mathcal{N}(\mu_{14}^{hr}, \sigma_{14}^{hr})$$

$$p(\phi_4|\omega_2^{hr}) = \mathcal{N}(\mu_{24}^{hr}, \sigma_{24}^{hr})$$

Considera-se igual a probabilidade a *priori* de cada uma das classes uma vez que ambas as classes têm a mesma probabilidade de aparecer, de tal modo que:

$$P(\omega_1^{hr}) = P(\omega_2^{hr}) = 0.5$$

Na Figura 4.4.7 podem ser visualizados na primeira coluna os histogramas relativos aos dados de treino recolhidos para as duas características utilizadas na classificação de objectos rígidos ou pessoas. Na segunda coluna encontram-se os respectivos modelos estatísticos. Repare-se que o erro neste classificador é substancialmente maior que no classificador anterior, que distingue o deslocamento dos objectos. Isto acontece porque ambas as características não permitem uma diferenciação absoluta relativamente aos dois tipos de objectos. No entanto, existe um ponto de separação das classes que é claro, quando se observam os modelos. Note-se ainda que a mistura de gaussianos, que corresponde efectivamente ao modelo estatístico dos dados resumidos no histograma se assemelha à forma deste.

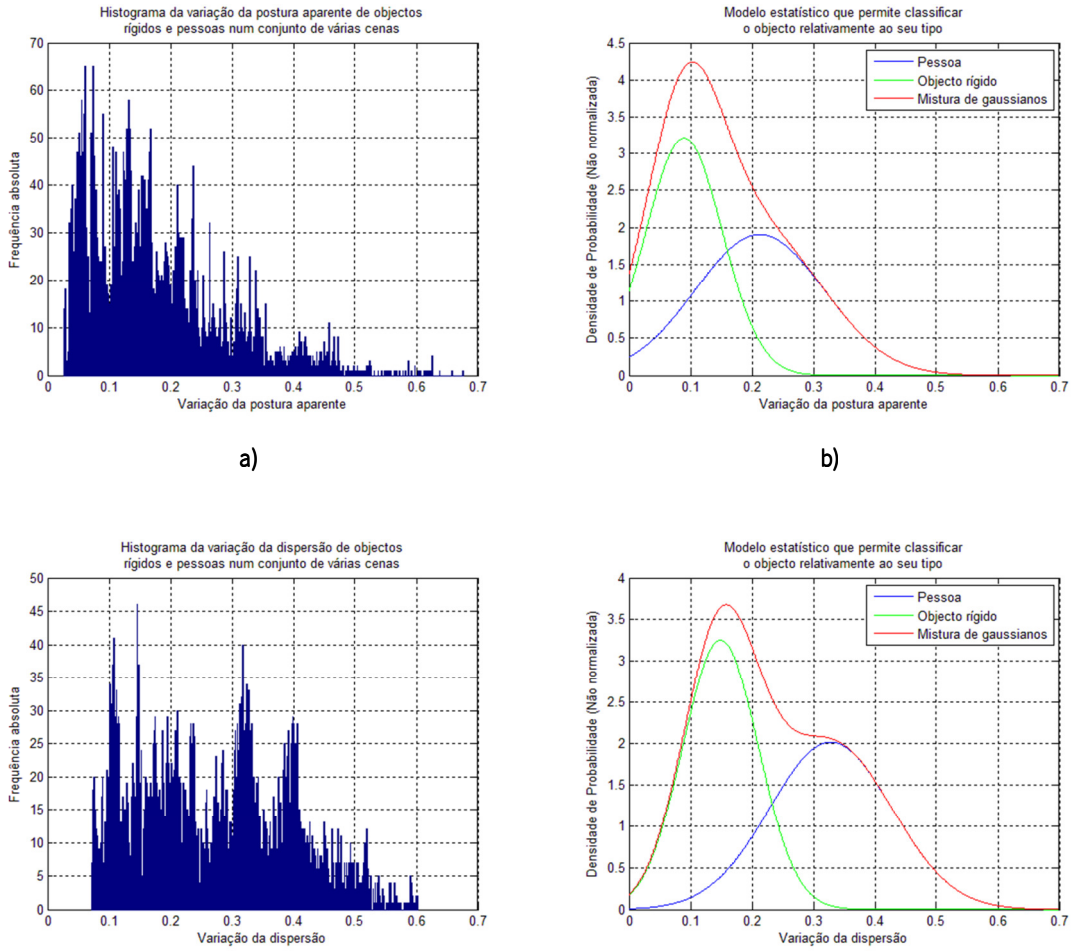


Figura 4.4.7: Histogramas dos dados de treino recolhidos e respectivos modelos estatísticos para as características variação da postura aparente e variação da dispersão.

a) Histograma da variação da postura aparente;
b) Modelo estatístico da variação da postura aparente;
c) Histograma da variação da dispersão;
d) Modelo estatístico da variação da dispersão.

Seguidamente apresenta-se o processo que permite gerar os modelos estatísticos utilizados na detecção de um comportamento de risco quando uma pessoa se move lentamente na piscina. A característica utilizada para efectuar esta classificação é a variação da deformação da forma ϕ_5. Foi utilizada a mesma metodologia descrita anteriormente, ou seja, os dados recolhidos são classificados manualmente, de acordo com o conteúdo das imagens onde têm origem os dados e geradas distribuições gaussianas uni variadas para cada classe. Os valores obtidos apresentam-se a seguir:

$$\text{Classe 1: } \mu_{15}^{nd} = 0.2782, \sigma_{15}^{nd} = 0.1574$$

Classe 2: $\mu_{25}^{nd} = 0.7545, \sigma_{25}^{nd} = 0.1321$

As duas funções densidade de probabilidade utilizadas no classificador dado pela Equação 4.4.20 são assim definidas de acordo com:

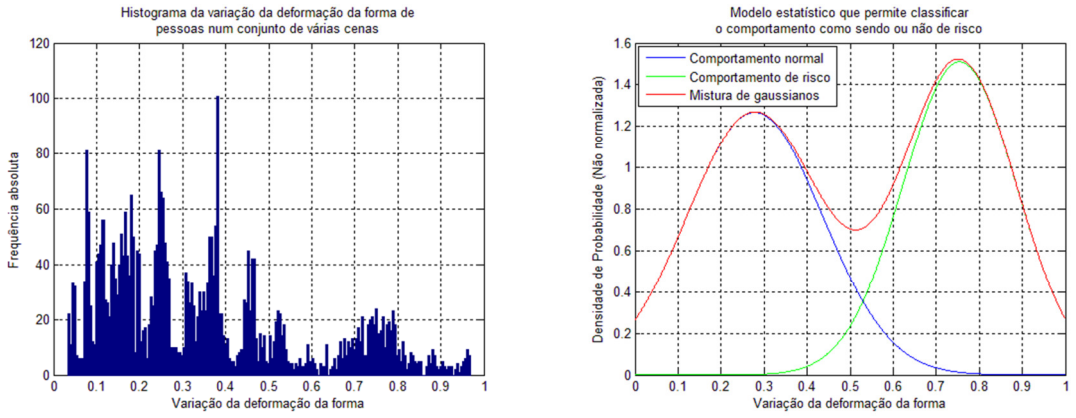
$$p(\phi_5|\omega_1^{nd}) = \mathcal{N}(\mu_{15}^{nd}, \sigma_{15}^{nd})$$

$$p(\phi_5|\omega_2^{nd}) = \mathcal{N}(\mu_{25}^{nd}, \sigma_{25}^{nd})$$

A probabilidade a *priori* de cada uma das classes considera-se igual, uma vez que ambas as classes têm a mesma probabilidade de aparecer, de tal modo que:

$$P(\omega_1^{nd}) = P(\omega_2^{nd}) = 0.5$$

Em baixo, na Figura 4.4.8, é apresentado o histograma e correspondente modelo estatístico dos dados de treino recolhidos para classificar o comportamento de pessoas relativamente ao risco de afogamento. No modelo estatístico apresentado é mais notória a diferença entre as duas classes, contribuindo este facto para uma classificação mais correcta. Na próxima secção são apresentados os erros associados aos classificadores na classificação dos dados de treino que lhes deram origem.



a) b)

Figura 4.4.8: Histograma dos dados relativos à variação da deformação da forma e correspondente modelo estatístico utilizado na classificação de comportamentos de risco.

a) **Histograma da distribuição dos valores da amostra;**
b) **Modelo estatístico composto por duas distribuições gaussianas uni variadas.**

A última distinção da árvore de decisão apresentada na Secção 4.4.3 tem como objectivo detectar o tipo de comportamento de risco de um indivíduo, caso exista algum risco. Uma pessoa pode

estar em aflição ou inconsciente. A abordagem utilizada para conceber os modelos estatísticos utilizados no classificador dado pela Equação 4.4.21 segue os padrões apresentados anteriormente. As características escolhidas para efectuar esta distinção foram a variação da dispersão, a variação da área e a média da componente saturação na área ocupada pelo objecto. Os valores dos parâmetros da mistura de gaussianas univariadas da característica variação da dispersão apresenta-se a seguir:

$$\text{Classe 1: } \mu_{13}^{du} = 0.2942, \sigma_{13}^{du} = 0.0872$$

$$\text{Classe 2: } \mu_{23}^{du} = 0.4780, \sigma_{23}^{du} = 0.0587$$

Em baixo encontram-se os valores dos parâmetros das gaussianas univariadas que compõe o modelo que representa as classes relativas à variação da área de um indivíduo:

$$\text{Classe 1: } \mu_{16}^{du} = 0.1348, \sigma_{16}^{du} = 0.0558$$

$$\text{Classe 2: } \mu_{26}^{du} = 0.3067, \sigma_{26}^{du} = 0.0651$$

Por último apresentam-se os parâmetros do modelo da saturação média de uma pessoa composto por uma mistura de gaussianas uni variadas:

$$\text{Classe 1: } \mu_{17}^{du} = 0.7098, \sigma_{17}^{du} = 0.0666$$

$$\text{Classe 2: } \mu_{27}^{du} = 0.4985, \sigma_{27}^{du} = 0.0160$$

Sabendo os valores dos parâmetros dos modelos estatísticos que descrevem o comportamento de risco da pessoa em aflição ou inconsciente, é possível escrever as funções densidade de probabilidade utilizadas no classificador definido pela Equação 4.4.21:

$$p(\phi_3|\omega_1^{du}) = \mathcal{N}(\mu_{13}^{du}, \sigma_{13}^{du})$$

$$p(\phi_3|\omega_2^{du}) = \mathcal{N}(\mu_{23}^{du}, \sigma_{23}^{du})$$

$$p(\phi_6|\omega_1^{du}) = \mathcal{N}(\mu_{16}^{du}, \sigma_{16}^{du})$$

$$p(\phi_6|\omega_2^{du}) = \mathcal{N}(\mu_{26}^{du}, \sigma_{26}^{du})$$

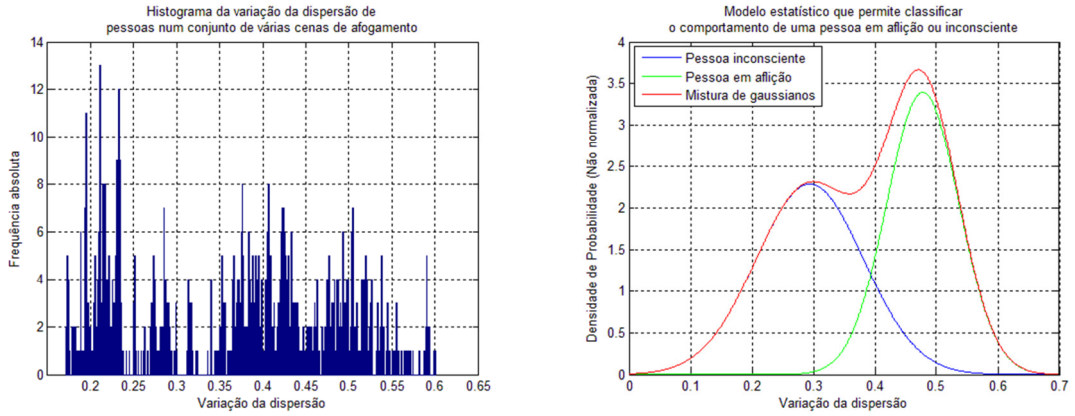
$$p(\phi_7|\omega_1^{du}) = \mathcal{N}(\mu_{17}^{du}, \sigma_{17}^{du})$$

$$p(\phi_7|\omega_2^{du}) = \mathcal{N}(\mu_{27}^{du}, \sigma_{27}^{du})$$

A probabilidade a *priori* de cada uma das classes considera-se igual, uma vez que ambas as classes têm a mesma probabilidade de aparecer, de tal modo que:

$$P(\omega_1^{du}) = P(\omega_2^{du}) = 0.5$$

Nos gráficos da Figura 4.4.9 pode observar-se à esquerda o histograma da distribuição dos dados recolhidos relativamente à variação da dispersão em pessoas a simularem comportamentos de aflição e inconsciência na piscina. As pessoas em aflição apresentam, tipicamente, maiores dispersões devido ao esbracejar característico deste tipo de comportamento, como se teve oportunidade de verificar anteriormente na Secção 4.3. Nesta mesma figura, à direita, pode ser visto o modelo estatístico gerado para os dados de treino recolhidos.

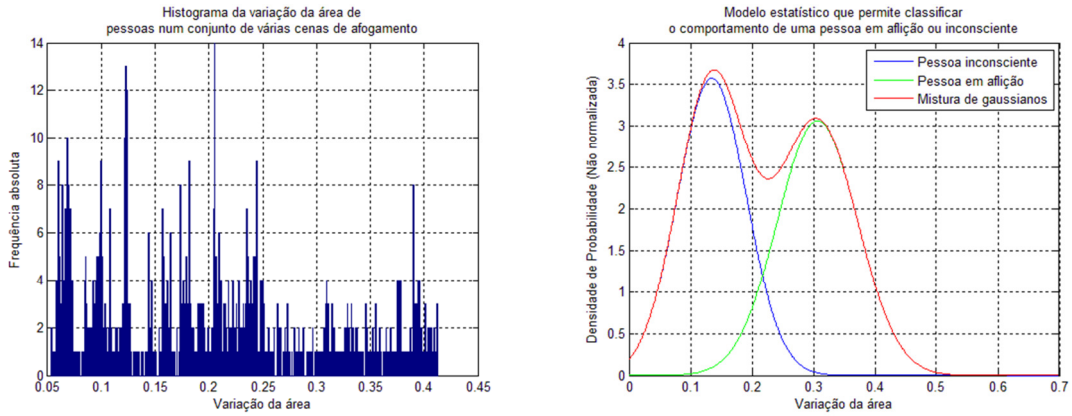


a) b)

Figura 4.4.9: Histograma dos dados relativos à variação da dispersão e respectivo modelo estatístico para as classes de comportamento aflição e inconsciência de um indivíduo.

a) Histograma da distribuição dos valores da amostra;
b) Modelo estatístico composto por duas distribuições gaussianas uni variadas.

O gráfico (a) da Figura 4.4.10 mostra o histograma relativo aos dados recolhidos da característica variação da área de uma pessoa em comportamentos já referidos no caso anterior. À semelhança da característica anterior, também aqui um indivíduo em aflição apresenta variações da área superiores a um indivíduo inconsciente, o que vai de encontro à realidade. Apesar de existir uma fronteira de decisão bastante distinta, a semelhança entre a mistura de gaussianas e o histograma da distribuição de valores não é muito notória.

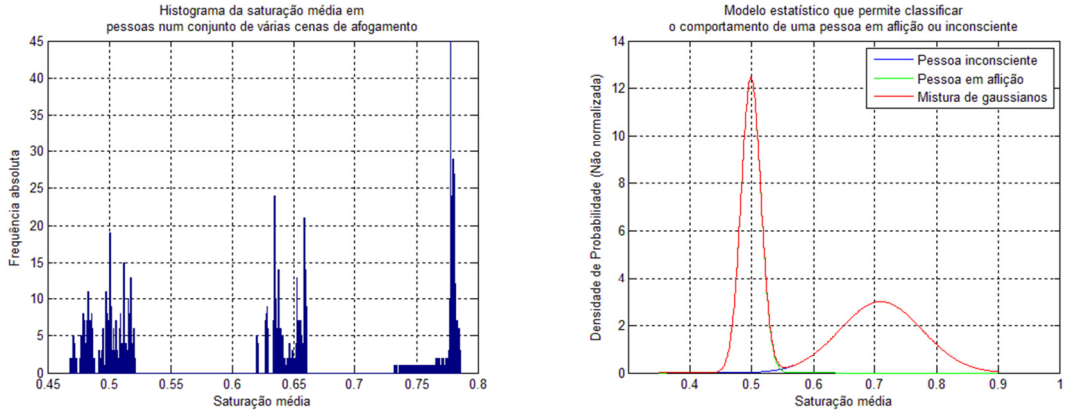


a) b)

Figura 4.4.10: Histograma dos dados relativos à variação da área e respectivo modelo estatístico para as classes de comportamento aflição e inconsciência de um indivíduo.

a) **Histograma da distribuição dos valores da amostra;**
b) **Modelo estatístico composto por duas distribuições gaussianas uni variadas.**

No que diz respeito ao nível médio da saturação na área correspondente a uma pessoa, a distinção é bastante clara, tal como se pode verificar tanto pelo histograma (a) da Figura 4.4.11 como pelo modelo estatístico presente em (b). Um indivíduo inconsciente apresenta níveis de saturação média bastante superiores, devido principalmente, ao facto do corpo afundar. Quanto mais profundo estiver o corpo maior é a saturação média. Pelo contrário, quando uma pessoa apresenta um comportamento de aflição, os níveis de saturação média são muito baixos, indicando que esta, ao esbracejar, gera uma enorme quantidade de bolhas de ar na água à sua volta.



a) b)

Figura 4.4.11: Histograma dos dados relativos à saturação média e respectivo modelo estatístico para as classes de comportamento aflição e inconsciência de um indivíduo.

a) **Histograma da distribuição dos valores da amostra;**
b) **Modelo estatístico composto por duas distribuições gaussianas uni variadas.**

Na secção seguinte são evidenciados os erros dos vários classificadores descritos relativamente aos dados de treinos que lhes deram origem.

4.4.5 Resultados experimentais da classificação dos dados de treino

Para examinar a eficácia dos classificadores concebidos para cada ramo da árvore binária de decisão, foram classificados os dados de treino, que deram origem aos modelos estatísticos que estão na base do processo de inferência dos comportamentos. É importante obter uma taxa de erro baixa, uma vez que este indicador permite medir a qualidade do classificador. No entanto a taxa de erro não deve ser nula, pois problemas de *overfitting* devem ser evitados, caso contrário, novos dados não serão correctamente classificados. Na Tabela 4.4.1 são mostradas as taxas de erro, numa escala percentual, relativa à classificação de comportamento dos dados de treino. As taxas de erro são determinadas através da contagem de classificações automáticas que diferem das classificações manuais dos dados utilizados para treino.

Classificador	Taxa de erro nos dados de treino
Deslocamento do objecto (ms)	1.8327 %
Tipo de objecto (hr)	14.8079 %
Detecção de risco (nd)	3.8919 %
Tipo de comportamento de risco (du)	0.1107 %

Tabela 4.4.1: Taxas de erros dos diferentes classificadores obtidas na classificação dos dados de treino.

Tal como seria de esperar, é o tipo de comportamento de risco que apresenta melhores resultados. Isto deve-se ao facto das duas classes serem bastante distintas, o que se pode verificar pelo modelo apresentado na Figura 4.4.11. A saturação é um descritor que distingue claramente pessoas afundadas de pessoas em aflição, devido às diferenças nos valores desta característica nos dois casos distintos. É exactamente este tipo de descritores que devem ser escolhidos, aqueles que maximizam a distância entre as classes que se pretendem identificar. Já a classificação do tipo de objecto apresenta uma taxa de erro bastante elevada. De facto, os descritores utilizados, a variação da dispersão e da postura aparente não diferenciam claramente estes dois tipos de objectos. É um problema complexo distinguir objectos rígidos de pessoas, uma vez que os objectos rígidos neste ambiente se movem devido à oscilação da superfície da água. Desta forma, devido à projecção do mundo real num plano 2D, a forma dos objectos sofre variações, não na realidade, mas na imagem projectada. Estas variações provocam variações de postura e de dispersão por vezes semelhantes às das pessoas. Além disso, as pessoas que se encontram inconscientes na

piscina apresentam características semelhantes aos objectos rígidos, uma vez que não se movem e dessa forma variam pouco a sua postura aparente e dispersão.

4.5 Discussão

A abordagem levada a cabo no módulo de análise de comportamento centrou-se na descoberta de características dos objectos encontrados pelo módulo de seguimento que diferenciassem os comportamentos que se pretendiam detectar. Para isso, e tendo em atenção os dados existentes relativamente aos tipos de comportamento aliados ao afogamento foram gerados descritores baseados na velocidade, forma e cores dos objectos. Esses descritores foram utilizados para recolher dados de amostras de vídeo que contivessem os comportamentos desejados de forma a construir classificadores. Optou-se por construir uma árvore de decisão que vai eliminando objectos pelas suas características de modo a atingir em último caso comportamentos característicos de afogamento. Esta árvore aplica um conjunto de condicionantes às características de um objecto, de tal modo que o afogamento só é considerado se existirem um conjunto de requisitos específicos e que caracterizam esse comportamento. Estas condicionantes dependem da classificação atribuída a cada vector de características provenientes do objecto em cada ramo da árvore. Os classificadores escolhidos assentam na inferência *Bayesiana*, que obtém bons resultados na prática, sem que seja necessária uma grande complexidade e elevados níveis de processamento. O facto de considerar as características independentes proporciona a utilização de modelos unidimensionais mais simples de tratar, facilitando a tarefa de classificação sem degradar esta de forma muito significativa.

A classificação em cada fotograma tem no entanto que ser credível, não podendo existir mudanças repentinas de comportamentos que não fariam sentido na realidade. Ou seja, é necessário incluir o factor tempo e alguma lógica à sequência de comportamentos detectada. Deste modo foi concebida uma máquina de estados que impõe uma lógica à detecção de afogamento baseada nos tempos médios acoplados a cada fase do afogamento. Com esta máquina de estados, além de se estabilizar o comportamento inferido, é também possível comutar de estado comportamental de forma lógica. Por exemplo, um indivíduo que se encontra inconsciente a um tempo superior a 4 minutos não pode voltar a mover-se. Deste modo, o despoletar de um alerta só deve ocorrer quando de facto existir uma situação de afogamento clara.

A diferenciação entre objectos rígidos e pessoas é um passo necessário, porque objectos não se afogam. Mas por outro lado pode ser um problema, pois pessoas paradas, com comportamentos

típicos de afogamento parecem objectos rígidos. Este é um problema intrínseco da aplicação, pois em piscinas domésticas podem existir vários objectos a flutuar à superfície da água ou mesmo no fundo da piscina. Não se pretendem falsos alertas de afogamento destes objectos. No entanto esse problema não seria crítico, pois ninguém estaria em perigo. Por outro lado, confundir uma pessoa com um objecto pode ser um caso muito grave se a mesma se estiver a afogar. Esta é uma situação crítica que não pode de modo algum acontecer. Contudo trata-se do ponto mais frágil do processo de análise de comportamento, pois não pode ser evitado e para ser resolvido poderá implicar a recolha de características mais diferenciadoras ou a utilização de dados provenientes de câmaras noutras localizações.

Relativamente aos sistemas de detecção de afogamento para piscinas públicas, anteriormente mencionados, existem vários desafios, começando desde logo pela colocação das câmaras. O facto de utilizarem uma vista de cima proporciona a normalização intrínseca dos dados, uma vez que, independentemente da posição dos nadadores na piscina, a sua área sofre poucas variações, o mesmo acontecendo com a velocidade e a forma do indivíduo. Uma vista de cima é uma grande vantagem pois permite também utilizar descritores como a melhor elipse que envolve o objecto, sendo mais fácil detectar quando este apresenta uma postura vertical. Num sistema de detecção de afogamento para piscinas domésticas a localização das câmaras deve ser discreta, não podendo de forma alguma ser colocada directamente por cima da piscina, pois deixa imediatamente de cumprir esse requisito. Isso torna a detecção mais difícil pois a área do objecto diminui com a distância à câmara, caso se desloque com velocidade constante, à medida que se afasta da câmara a velocidade diminui. Determinados descritores de comportamento que são óptimos na vista de cima, como a melhor elipse que envolve o objecto, numa vista como a que se utiliza no sistema descrito neste trabalho esse descritor não proporciona diferenciação entre uma postura vertical e horizontal. É portanto necessário encontrar outros descritores que forneçam uma imagem do comportamento observado. Este é um ponto fundamental que distingue este sistema dos anteriormente descritos. A localização das câmaras assume um papel fundamental na escolha das características dos objectos a medir para inferir o seu comportamento. Outra diferença fundamental tem a ver com a possibilidade de existirem outros objectos a flutuar à superfície da água. Nas piscinas públicas este caso raramente acontece, mas numa piscina doméstica é absolutamente normal. Isso implica uma distinção entre objectos e pessoas, algo que não é necessário na detecção de afogamento em piscinas públicas, pois todos os objectos encontrados são classificados como pessoas. Nos trabalhos apresentados por (Eng, Toh, Yau, & Wang, 2008) e

(Lu & Tan, 2004) não são diferenciados os tipos de objectos encontrados, o que torna mais fácil a detecção de afogamento.

Em suma, a abordagem descrita neste trabalho diferencia-se das demais relativamente aos descritores utilizados, sendo apropriados para qualquer vista e sendo possível distinguir objectos de pessoas, o que para piscinas domésticas é fundamental. Diferencia-se também na estratégia utilizada para detectar o comportamento, a utilização de uma árvore binária de decisão suportada por classificadores *naive Bayes*, cuja complexidade é inferior mas obtendo na prática bons resultados. A concepção dos classificadores é em tudo semelhante à dos autores mencionados, uma vez que se apoia em dados de treino, assentando assim numa filosofia de aprendizagem supervisionada efectuada *off-line*. A utilização da máquina de estados também partilha semelhanças com as outras abordagens, uma vez que em (Eng, Toh, Yau, & Wang, 2008) os HMMs são baseados numa máquina de estados e em (Lu & Tan, 2004) o comportamento de cada pessoa corresponde a um estado baseado nos padrões encontrados em cada momento, remetendo também para a noção de máquina de estados finita.

Capítulo 5

5 Teste do Sistema em Ambiente Real

Neste capítulo são apresentados os vários testes realizados ao sistema final, isto é, composto pelos vários módulos descritos nos capítulos anteriores, tal como foi evidenciado na Figura 1.4.1 do capítulo introdutório. O protótipo do sistema é também abordado relativamente à sua implementação em *Simulink* e respectivo interface gráfico capaz de mostrar os resultados obtidos. As amostras de vídeo utilizadas nos testes consistiram em gravações de simulações de vários tipos de afogamento tendo em conta as características apresentadas na Secção 4.3. Nenhuma destas amostras foi utilizada no treino dos classificadores descritos no capítulo anterior. Esta fase tem como objectivo apurar a eficiência global do sistema, tendo em conta os desafios indicados ao longo das descrições dos vários módulos.

5.1 Ambiente e condições de teste

Para validação e teste dos algoritmos foram gravadas várias imagens, numa cena contendo uma piscina, em diferentes horas do dia e em diferentes situações de utilização e oscilação da superfície da água. Foram também utilizadas várias câmaras de vigilância no processo, incluindo câmaras subaquáticas, em diferentes posições, sendo a captura e a gravação de cada câmara efectuadas de forma sincronizada. A utilização de todas estas câmaras teve como objectivo a escolha do melhor ângulo para captação da imagem. Foi implementado um software em C, utilizando a API (Application *Programming Interface*) fornecida pelo *OpenCV* (*Open Source Computer Vision*), capaz de capturar as várias câmaras através de uma placa de captura analógica, tal como se pode observar na imagem da Figura 5.1.1. Foi também utilizada uma câmara de gravação de vídeo amadora com elevada qualidade e resolução de modo a verificar as diferenças relativamente à imagem fornecida pelas câmaras comuns de vídeo vigilância. Neste caso a gravação foi efectuada no próprio dispositivo de captura, sendo depois convertida para uma resolução menor, mantendo no entanto a qualidade. A disposição dos vários dispositivos de captura na cena pode ser vista nos esquemas das figuras Figura 5.1.2 e Figura 5.1.3.

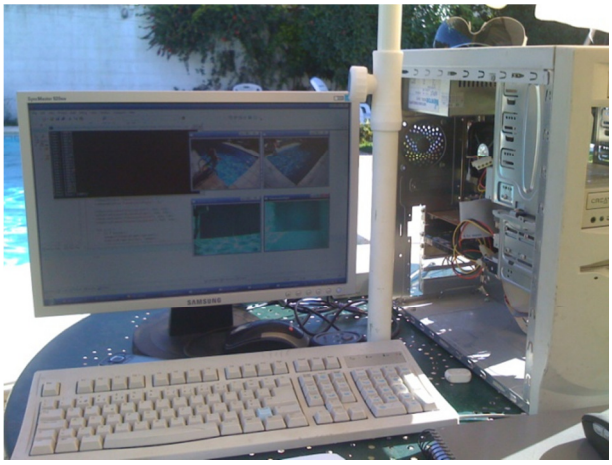


Figura 5.1.1: Máquina utilizada na gravação sincronizada das quatro câmaras analógicas. Na imagem pode ser vista a execução do software responsável pela captura e armazenamento do vídeo.

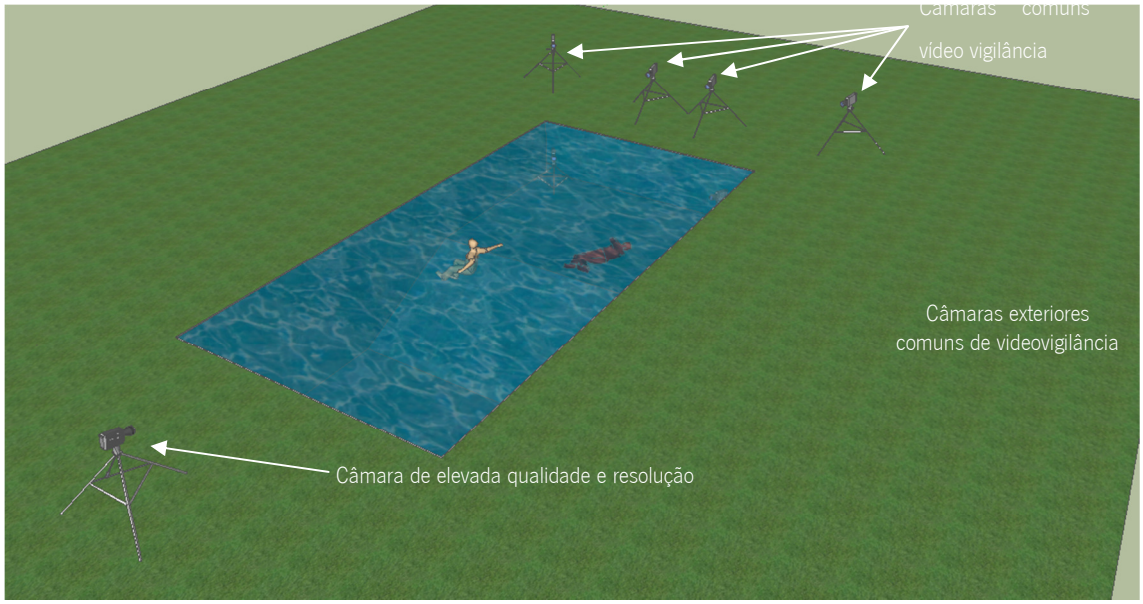


Figura 5.1.2: Disposição das várias câmaras utilizadas na gravação das imagens de referência para efectuar a prova de conceito. Uma das câmaras corresponde a um aparelho doméstico de gravação de vídeo com elevada resolução e qualidade. As restantes são câmaras comuns utilizadas em aplicações típicas de videovigilância, tendo portanto, menor qualidade de imagem e mais baixa resolução. Gerado com o *Google SketchUp*.

Na Figura 5.1.4 podem observar-se as câmaras subaquáticas, utilizadas neste contexto para se retirarem conclusões acerca da sua necessidade no que respeita à influência na fiabilidade e robustez do sistema. Já na Figura 5.1.5 é possível ter uma perspectiva da vista proporcionada pelas câmaras centrais.

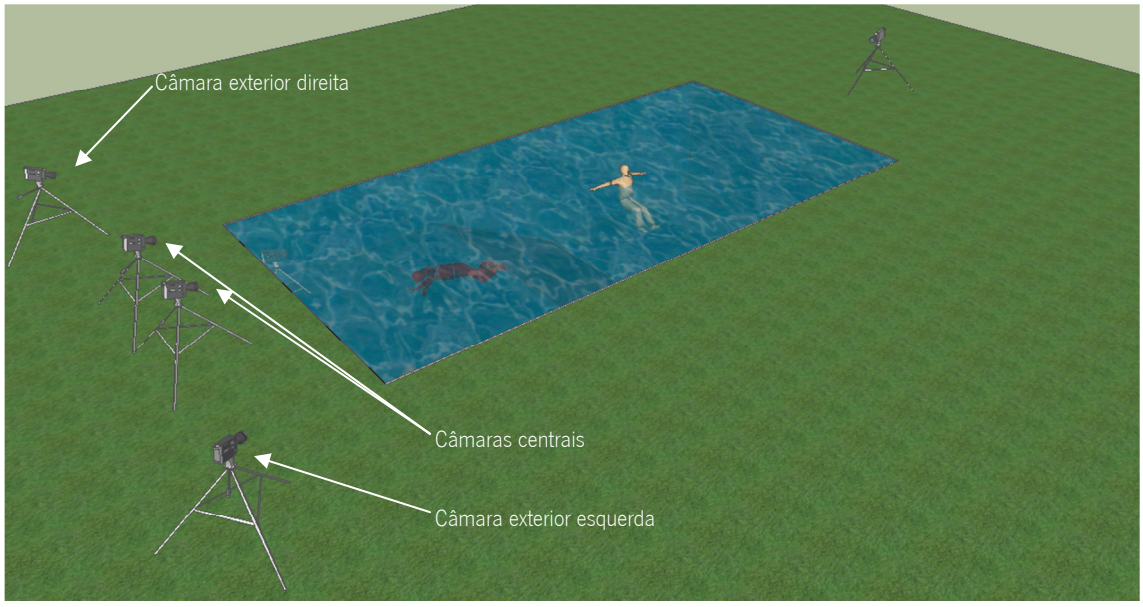


Figura 5.1.3: Vista da piscina do lado das câmaras de vídeo vigilância. Gerado com o *Google SketchUp*.

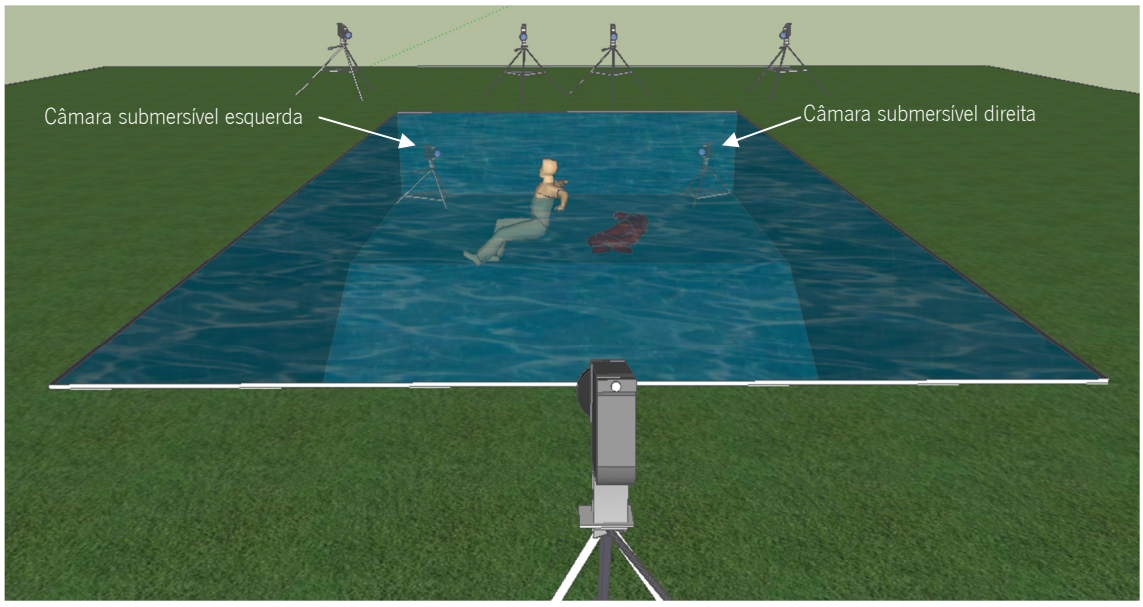


Figura 5.1.4: Vista da piscina pela câmara central de elevada resolução e qualidade. Aqui são visíveis as câmaras submersíveis de vídeo vigilância utilizadas nas gravações. Gerado com o *Google SketchUp*.

As imagens de teste e validação foram cuidadosamente preparadas, sendo capturadas com o objectivo de abarcar um conjunto de situações representativo do ambiente real da piscina. A captura ocorreu sempre com as quatro câmaras em simultâneo e devidamente sincronizadas em diferentes situações. Foram assim capturadas imagens da piscina vazia, com oscilação elevada e reduzida da superfície da água, com diferentes níveis de luminosidade, a diferentes horas do dia e em diferentes condições atmosféricas. Além disso foram utilizadas imagens de teste com indivíduos na piscina a executar diferentes acções. Entre estas acções encontram-se a utilização normal, tal como nadar e mergulhar, a interacção entre os intervenientes e vários objectos passíveis de serem

encontrados numa piscina doméstica e a simulação de situações de afogamento variadas, com base na literatura existente sobre o assunto. Na imagem da Figura 5.1.6 pode também ser vista uma das câmaras exteriores utilizada na captura das imagens de teste e validação, no caso, a câmara direita. Na Tabela 5.1.1 encontram-se as características das câmaras utilizadas nas gravações. Todas as gravações foram efectuadas a uma taxa de 25 fotogramas por segundo com uma resolução de 320 por 240 pixéis, em formato *raw*. As características de hardware da máquina utilizada nos testes do sistema de detecção de afogamento, o sistema operativo utilizado e a versão do *matlab* encontram-se na Tabela 5.1.2.

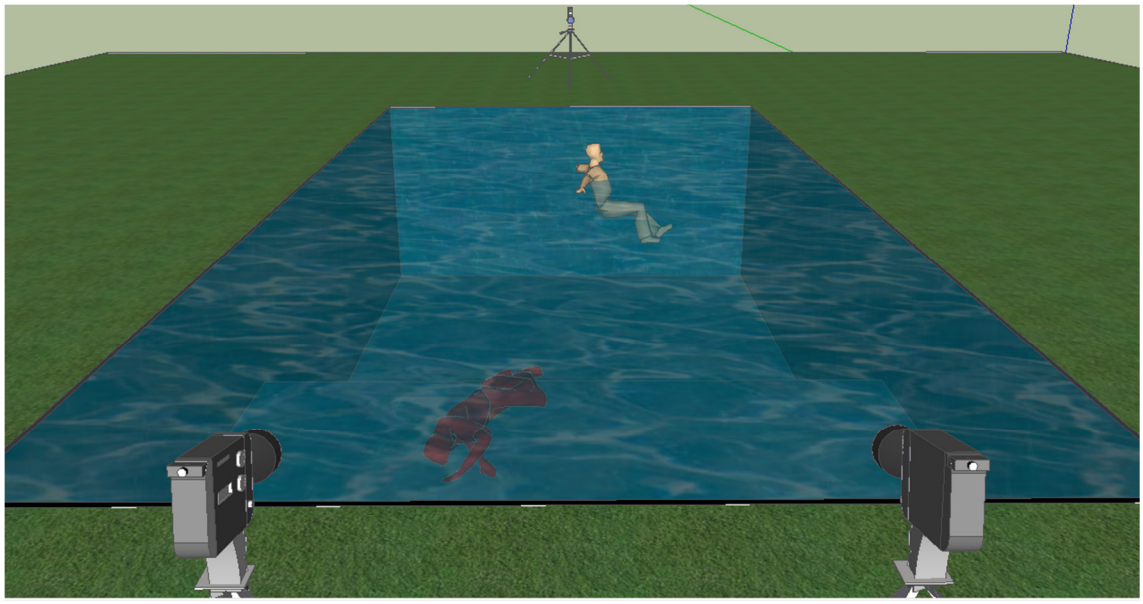


Figura 5.1.5: Vista da piscina pelas câmaras centrais. Gerado com o *Google SketchUp*.

Figura 5.1.6: Câmara exterior direita posicionada num dos cantos da piscina.

Características	Câmaras exteriores	Câmaras subaquáticas
Marca - modelo	CNB – B2000P	Genie CCTV – GCB6020
Saída vídeo	1.0 Vpp (75 Ω, composto)	1.0 Vpp (75 Ω, composto)
Tamanho imagem	1/3''	1/3''
Sensor imagem	Sony Super HAD CCD Exview	Sony Exview HAD Ultra High Sensitivity CCD Sensor
Varrimento	PAL	PAL
Resolução horizontal	480 TV Lines	480 TV Lines
Lente (distância focal)	6 mm	3.6 mm
Iluminação mínima	0.00 Lux (com infravermelhos)	0.05 Lux

Tabela 5.1.1: Características das câmaras utilizadas na gravação dos vídeos.

Características	Máquina
Marca - Modelo	Toshiba - A300
Processador	Intel Core 2 Duo T9300 2.50 GHz
Memória (RAM)	4.00 GB
Sistema Operativo	Windows 7 Ultimate 64 bits
Versão do *matlab*	7.10.0 (R2010a) (64 bits)

Tabela 5.1.2: Características da máquina utilizada nos testes do sistema de detecção de afogamento.

5.2 Metodologia de implementação

A implementação dos algoritmos descritos nos capítulos anteriores foi efectuada recorrendo à ferramenta de modelação *Simulink*, a qual é parte integrante do *Matlab*. Largamente utilizada em investigação, foi escolhida para simplificar e, deste modo, acelerar a tarefa de desenvolvimento. O *Matlab* contém alguns dos algoritmos mais importantes nos domínios da probabilidade, estatística, processamento de sinal e de imagem. Muitas das operações tradicionais de processamento de imagem presentes na literatura encontram-se desenvolvidas, sendo apenas necessário conhecer os parâmetros e os formatos das entradas e saídas para poderem ser utilizados. O *Simulink* permite executar os algoritmos de forma iterativa com passos de tempo discretos, baseado numa lógica de estados, onde as saídas são geradas por intermédio do seu estado actual e passado, à semelhança de um sistema de controlo discreto. Neste ambiente, cada função é representada graficamente por um bloco com entradas, saídas e parâmetros. A ligação

entre os diferentes módulos permite, de uma forma extremamente rápida e eficaz, provar os conceitos idealizados e verificar até que ponto determinado algoritmo é adequado ou não para determinada operação. É ainda possível desenvolver os próprios blocos em linguagem m, do *Matlab*, ou então em linguagens como o C/C++, quando é necessária optimização, existindo assim uma flexibilidade absoluta na utilização desta ferramenta. Como última nota fica a possibilidade do *Matlab* ser capaz de gerar o código C a partir do sistema desenvolvido, para várias plataformas de hardware, e diferentes sistemas operativos, sendo este, mais uma vez, um ponto fundamental na contribuição para a aceleração do processo de desenvolvimento do protótipo.

5.3 Protótipo do sistema

O sistema de detecção de afogamento é composto pelos módulos de segmentação de objectos em movimento, seguimento dos mesmos e análise de comportamento dos que são reconhecidos como pessoas. À semelhança do que foi mencionado na secção anterior, foi implementado um protótipo do sistema em *Simulink*, tal como o esquema da Figura 5.3.1 pretende demonstrar. Neste esquema podemos constatar claramente três módulos de processamento, os quais foram anteriormente definidos e um outro responsável pela gravação e visualização dos vários dados resultantes da execução em diferentes partes do sistema. No modelo apresentado é possível verificar que o sistema recebe o vídeo directamente a partir de um ficheiro não comprimido e uma imagem correspondente à máscara da localização da piscina gerada automaticamente *off-line*. Na Figura 5.3.2 pode ser visto o interface gráfico da aplicação, que permite configurar os parâmetros associados aos vários módulos e a visualização dos resultados obtidos no processo. A imagem à esquerda corresponde ao vídeo escolhido para processar, sendo os objectos encontrados identificados e marcados com o comportamento inferido pelo sistema. Na imagem central é apresentado o mapa binário de *foreground* com a cor do objecto correspondente ao seu comportamento e tipo. Por baixo dessa imagem existe uma legenda que mostra os vários comportamentos e tipos de objectos possíveis de detectar. A imagem à direita mostra o *background* estimado pelo módulo de segmentação de objectos em movimento. Imediatamente por baixo da imagem etiquetada aparece o menu de selecção das cenas gravadas e as respectivas vistas, i.e., as câmaras. Em baixo encontram-se vários botões, sendo os da esquerda responsáveis pela iniciação e paragem da execução do sistema e os da direita responsáveis pela configuração dos parâmetros subjacentes aos vários módulos.

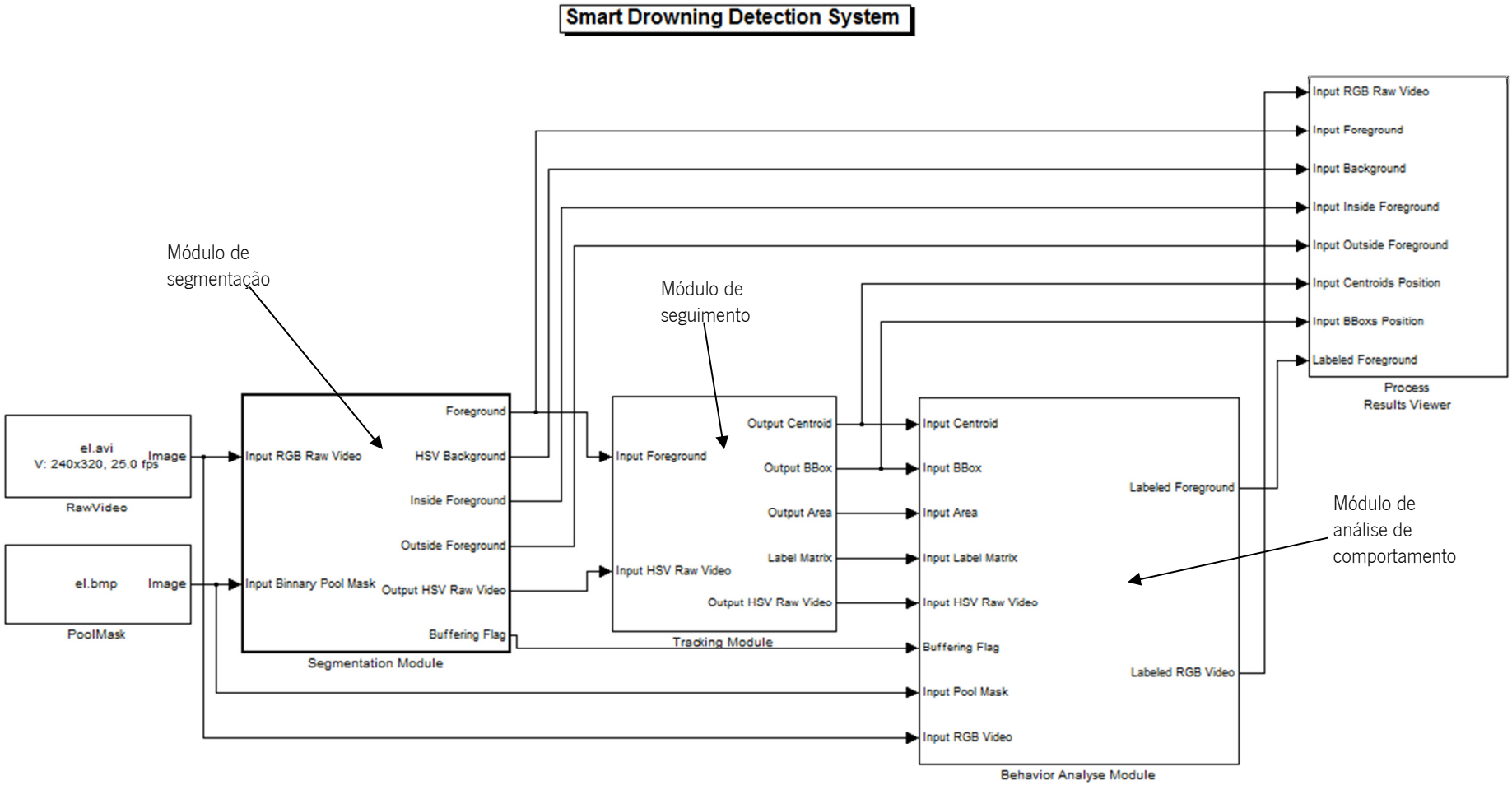


Figura 5.3.1: Protótipo do sistema de detecção de afogamento implementado no *Simulink*. O sistema recebe como entradas o vídeo a processar e a imagem da máscara da localização da piscina. O primeiro módulo efectua a segmentação de movimento, fornecendo como saída o mapa binário de *foreground* que é passado como entrada do módulo de seguimento de objectos. Este por sua vez enumera os objectos e extrai algumas características utilizadas no módulo de análise de comportamento. A saída deste último módulo marca o vídeo real identificando o comportamento das pessoas.

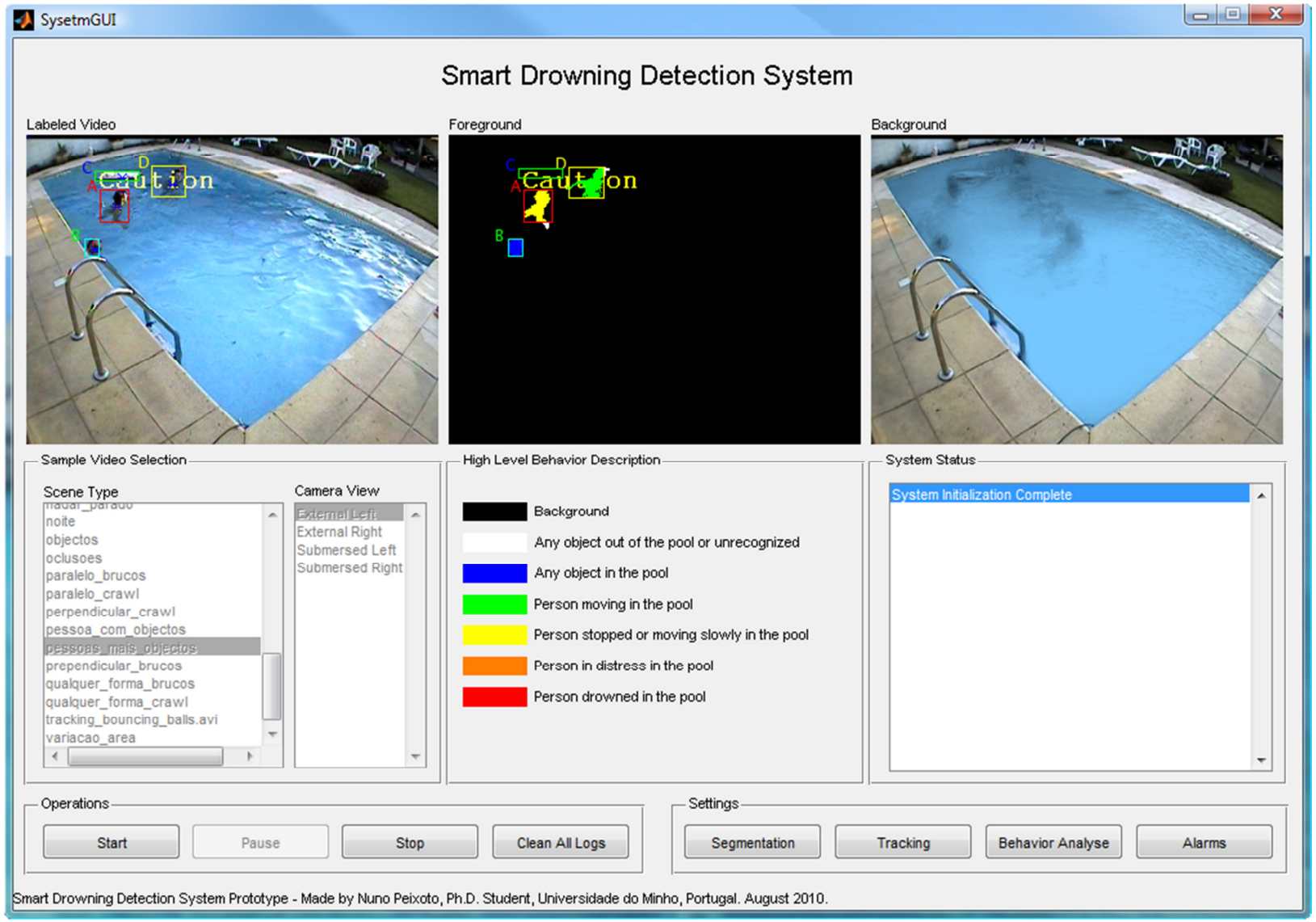


Figura 5.3.2: Interface gráfico do protótipo do sistema de detecção de afogamento para piscinas domésticas.

5.4 Resultados experimentais

Foram efectuados quatro testes fundamentais a partir dos vídeos gravados que envolvem diferentes tipos de afogamento e utilizações normais de uma piscina doméstica. Estes testes têm como objectivo proporcionar um suporte para uma análise qualitativa de robustez e fiabilidade do sistema para as situações que se pretendem detectar num ambiente associado a uma piscina doméstica. Destaca-se aqui a situação mais importante, à qual corresponde a objectivo fundamental do sistema que é a detecção da queda acidental de uma criança à piscina. Além disso, outras situações, tais como a piscina vazia ou utilizada por várias pessoas em simultâneo com objectos a flutuar são também destacadas para que se verifique que o sistema está continuamente a actuar sem que seja necessária qualquer intervenção humana. Os vídeos gravados não foram utilizados para o treino dos classificadores, embora o treino tenha sido efectuado no mesmo ambiente onde foram realizados os testes, ou seja, na mesma piscina e com os mesmos intervenientes.

Para cada um dos testes efectuados foi gerado um gráfico com os estados de cada objecto detectado pelo sistema e com os estados definidos manualmente por um ser humano. A comparação entre os estados detectados automaticamente e o verdadeiro comportamento são comparados de modo a retirar conclusões quanto aos falsos positivos e falsos negativos do sistema relativamente à detecção de afogamento. A partir destas comparações são depois retiradas conclusões acerca da eficácia e robustez do sistema.

5.4.1 Afogamento silencioso

O afogamento silencioso é simulado por um único indivíduo na piscina não existindo objectos a flutuar à superfície da água. O vídeo inicia-se com um indivíduo a entrar na piscina, Figura 5.4.1, deslocando-se posteriormente para a parte mais profunda desta até ficar "sem pé", Figura 5.4.2. Uma vez que o indivíduo não sabe nadar inicia-se a fase de aflição, onde a pessoa tenta desesperadamente manter-se à superfície esbracejando e mantendo uma posição próxima da vertical não existindo qualquer deslocamento, Figura 5.4.3. Esta fase tem uma duração de cerca de 20 segundos, momento a partir do qual a pessoa se afunda ficando inconsciente até ao fim do vídeo, Figura 5.4.4.

Figura 5.4.1: Indivíduo a entrar na piscina.

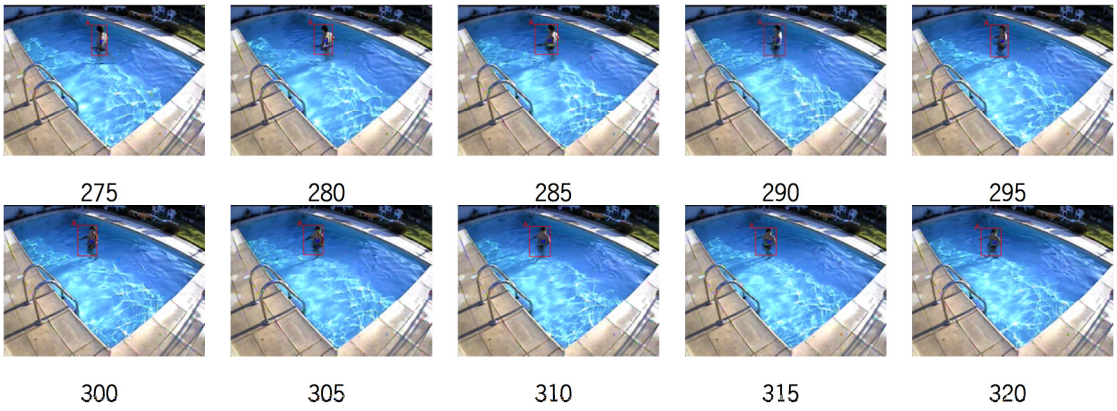


Figura 5.4.2: Indivíduo desloca-se de pé para a parte mais profunda da piscina.

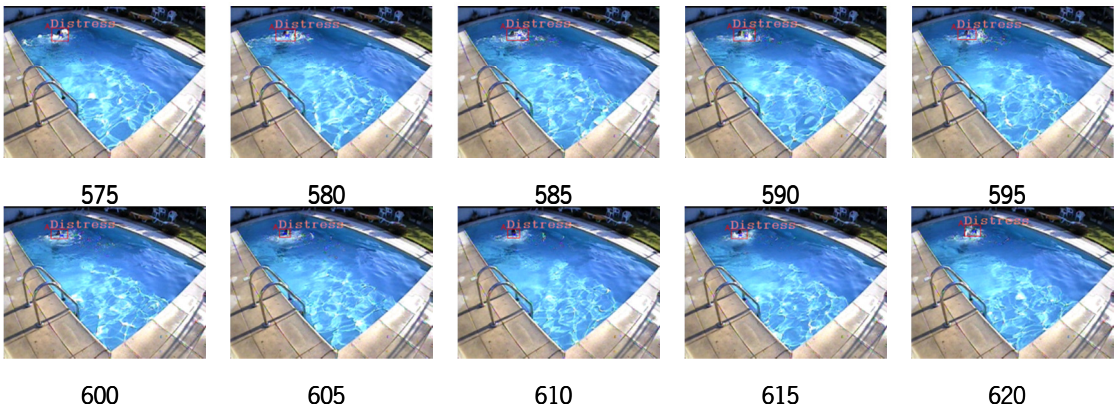


Figura 5.4.3: Indivíduo entra num estado de aflição tentando não se afundar. Esta fase corresponde ao início do afogamento.

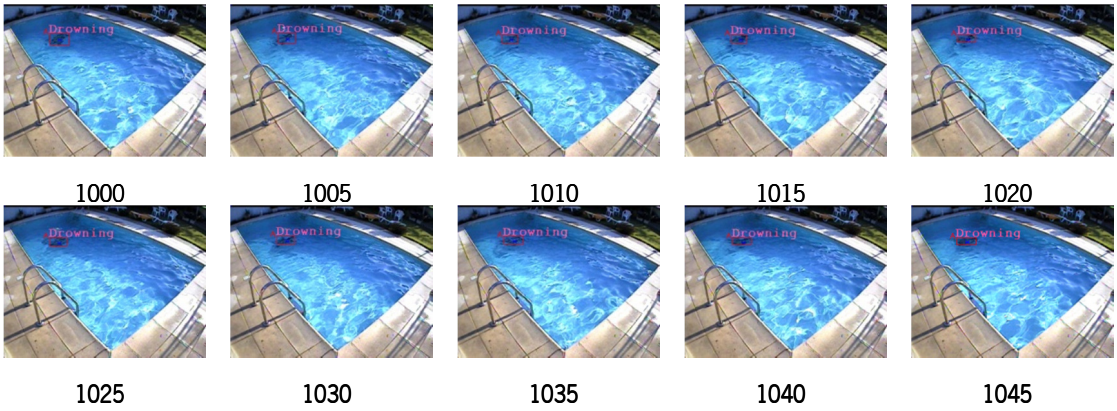


Figura 5.4.4: Indivíduo afunda-se e fica insconsciente. Neste momento considera-se que o indivíduo se afogou.

Na Figura 5.4.5 encontra-se o gráfico que mostra o comportamento inferido pelo sistema de detecção de afogamento e o verdadeiro comportamento, definido por um ser humano, no vídeo apresentado. Este gráfico permite obter uma descrição quantitativa da fiabilidade do sistema. Existem poucos momentos em que o sistema apresenta um comportamento inferido diferente do comportamento real, sendo no entanto pouco relevante, uma vez que a tarefa de detecção de afogamento é comprida. Obviamente haverá sempre um pequeno atraso até um máximo de 4 segundos, devido à filtragem aplicada às características dos objectos detectados, daí o desfasamento entre o início dos comportamentos inferidos e o início dos comportamentos reais.

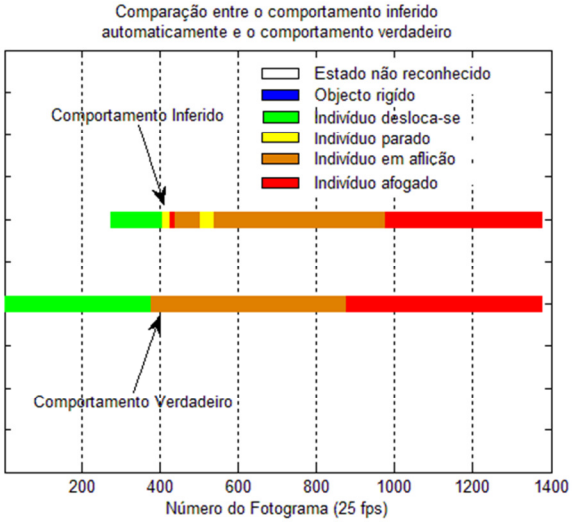


Figura 5.4.5: Gráfico comparativo entre o comportamento inferido pelo sistema de detecção de afogamento e o comportamento real definido por um ser humano num afogamento silencioso.

5.4.2 Afogamento causado por desmaio

O afogamento causado por desmaio corresponde a uma situação em que o indivíduo fica repentinamente inconsciente e imobilizado na piscina, podendo afundar, tal como acontece neste vídeo de teste. Este estado de inconsciência pode dever-se a um ataque cardíaco, uma grave lesão ou simplesmente a um desmaio. O indivíduo pode estar a deslocar-se na piscina ou pode estar parado quando esta situação ocorre. No caso de teste, o indivíduo entra na piscina, Figura 5.4.6, sendo marcado pelo sistema e considerado um objecto fora da piscina ou desconhecido. O indivíduo pára, Figura 5.4.7, e nada deslocando-se para outro ponto da piscina, Figura 5.4.8. Nesse momento desmaia e afunda-se logo de seguida, Figura 5.4.9. Repare-se na inferência de comportamento do sistema de detecção automática de afogamento. Em todas as situações a sua saída corresponde á realidade. Analisando agora o gráfico da Figura 5.4.10 é possível verificar que o sistema detecta uma situação de perigo quando na realidade o indivíduo se encontra a nadar. No entanto o indivíduo está a nadar muito devagar, e note-se que este o faz bastante longe da câmara o que acaba por ter alguma influência na velocidade do mesmo. A detecção de afogamento inferida pelo sistema acontece um pouco antes do afogamento real do indivíduo. Este problema deve-se ao facto do indivíduo estar bastante longe da câmara, o que faz com que os seus padrões de características remetam efectivamente para o afogamento. Estas distâncias foram ainda assim propositadas de modo a verificar a eficácia do sistema nos piores casos. Apesar do afogamento ser detectado antes do tempo não é crítico, uma vez que o principal objectivo, a detecção, acontece.

Figura 5.4.6: Indivíduo entra na piscina.

Figura 5.4.7: Indivíduo parado na piscina.

Figura 5.4.8: Indivíduo a nadar momentos antes do desmaio acontecer.

Figura 5.4.9: Indivíduo desmaiado e afundado. Nesta situação considera-se que está a ocorrer um afogamento.

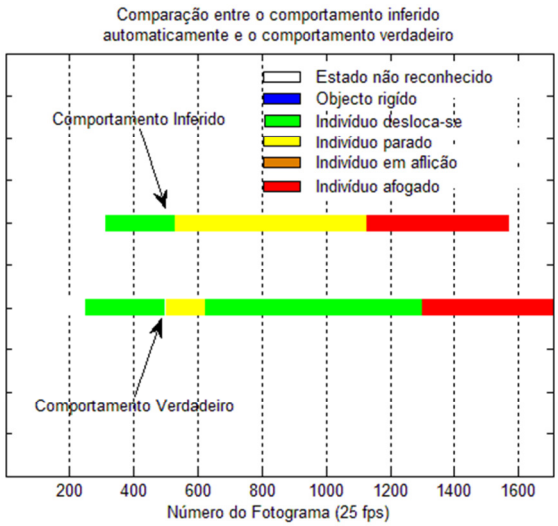


Figura 5.4.10: Comparação entre o comportamento real e o comportamento inferido pelo sistema num afogamento causado por demaio.

5.4.3 Queda de criança à piscina seguida de afogamento

Um dos objectivos fundamentais do sistema corresponde à detecção da queda de uma criança à piscina. Normalmente a criança encontra-se sozinha e aproxima-se da borda caindo à piscina e afogando-se de seguida. Neste contexto foi simulada a queda de uma criança que se chega à borda da piscina, Figura 5.4.11 e Figura 5.4.12. Depois de cair à água o corpo permanece inanimado a flutuar, Figura 5.4.13. Analisando o gráfico da Figura 5.4.14 verifica-se que os comportamentos inferidos pelo sistema são muito próximos da realidade, salvo um pequeno atraso na detecção do afogamento. No entanto o sistema não consegue seguir a criança no momento em que esta entra na água devido à enorme quantidade de bolhas de ar que se soltam, não possibilitando uma medida estável, tal como se pode verificar pela análise dos vários fotogramas da Figura 5.4.12. Mas em poucos segundos após a bolhas cessarem o sistema marca de imediato a criança como afogada, podendo assim ser dado o alerta, tal como é pretendido.

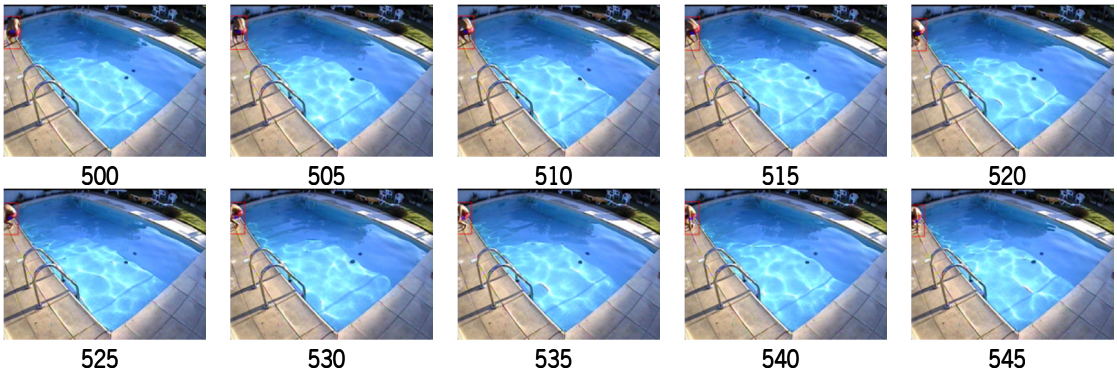


Figura 5.4.11: Simulação de uma criança na borda da piscina.

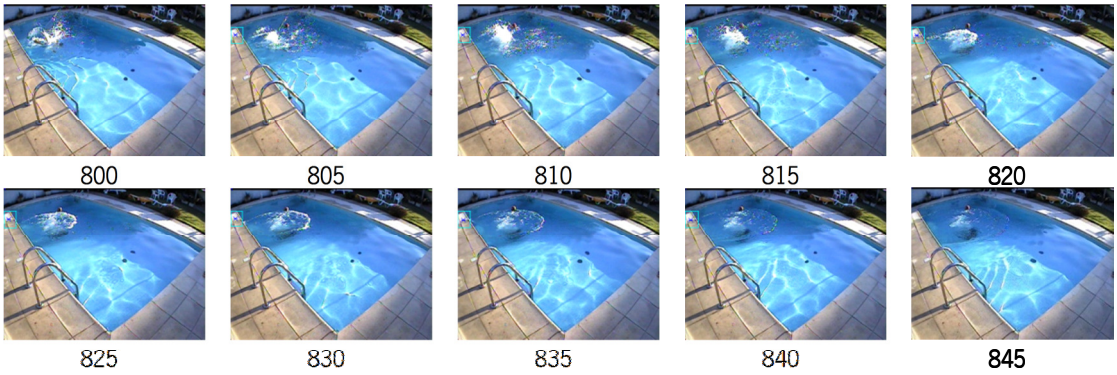


Figura 5.4.12: Simulação de uma criança no momento em que esta cai na piscina.

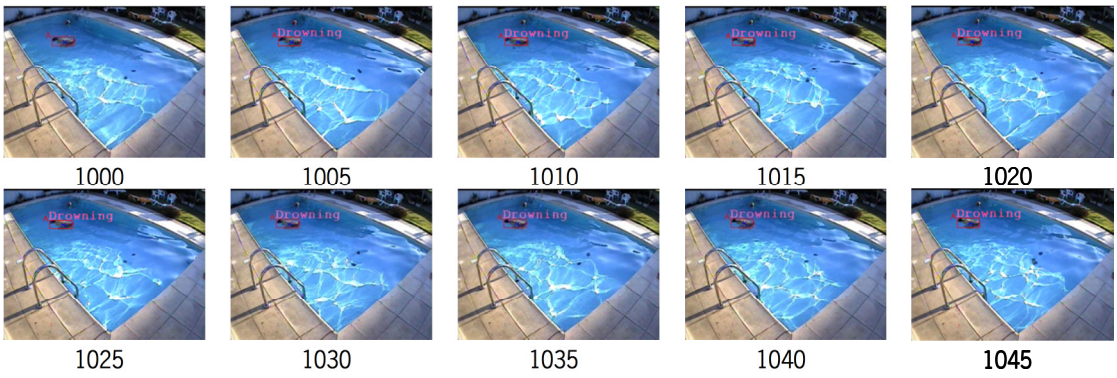


Figura 5.4.13: Simulação do corpo de uma criança a flutuar inanimada na superfície da água.

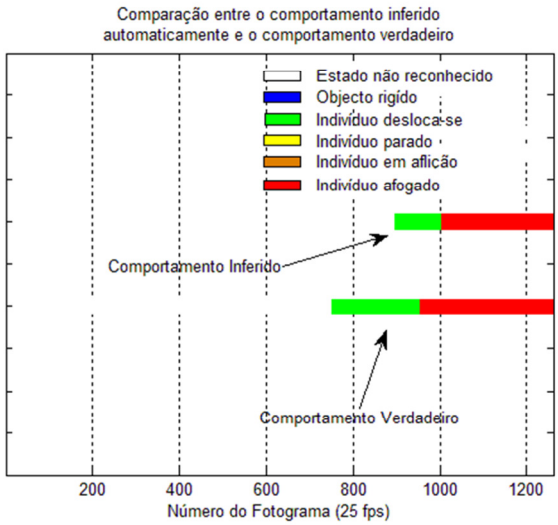


Figura 5.4.14: Comparação entre o comportamento inferido pelo sistema e o comportamento real exibido pela criança.

5.4.4 Duas pessoas com vários objectos e um afogamento

Apesar do caso mais importante na detecção de afogamento ser o de uma criança que possa cair à piscina sem que ninguém veja, a detecção de afogamentos com a presença de vários objectos e pessoas também é uma valência do sistema, embora menos necessária e também menos robusta. Nas próximas figuras são apresentadas amostras de vários fotogramas de um vídeo que tenta ilustrar a utilização normal de uma piscina doméstica. A questão fundamental neste teste é evitar os falsos positivos, continuando a ser fiável na detecção de um eventual afogamento, embora neste caso isso não seja crítico, pois encontra-se mais que uma pessoa na piscina para dar o alerta. Analisando os fotogramas das Figura 5.4.15 e Figura 5.4.16 é possível observar vários objectos a flutuar na superfície da água e dois indivíduos, um deles no exterior da piscina. Repare-se que um dos objectos não está identificado pois é demasiado pequeno e por isso é considerado ruído. O indivíduo na piscina encontra-se parado e o sistema marca-o como estando numa situação de perigo, pois esta fase pode corresponder ao início de um afogamento. Entre os fotogramas 1775 e 1795 o comportamento do indivíduo é erradamente inferido como sendo de aflição. Nos fotogramas da Figura 5.4.17 o comportamento do indivíduo é classificado como de perigo, o que está correcto, uma vez que se encontra parado. Já nos fotogramas da Figura 5.4.18 o indivíduo A é classificado como afogado. Este indivíduo não se está a afogar, mas encontra-se bastante longe da câmara e apresenta características muito semelhantes a um afogamento, uma vez que se encontra a boiar na prancha. Na mesma figura encontra-se o indivíduo B a nadar por baixo de água. Note-se neste caso a dificuldade do sistema em segui-lo devido sobretudo à cor branca dos seus calções que facilmente se confunde com o *background*. Tal como havia mencionado, este vídeo foi escolhido por conter algumas das situações mais complexas para o sistema, de modo a verificar a sua fiabilidade e robustez. Nos fotogramas da Figura 5.4.20 assiste-se ao início da oclusão entre os indivíduos A e B. O módulo de seguimento consegue efectivamente diferenciar os dois por alguns momentos, mas o facto de serem uma região única no mapa binário de *foreground* resultante da segmentação, não é possível prever durante muito tempo as suas posições reais. Desta forma, o seguidor acaba por fazer a junção dos dois objectos num só como forma de resolver o problema. Na Figura 5.4.20 os indivíduos atiram água um ao outro ao mesmo tempo que provocam oclusão. Neste caso bastante complexo, devido à oclusão prolongada o seguidor trata os dois indivíduos como um único. Repare-se que por vezes a marcação falha pois a prancha também provoca oclusão da pessoa que se encontra mais próxima dela.

Figura 5.4.15: Um indivíduo parado dentro da piscina, outro no exterior a pegar num objecto e outros objectos a flutuar à superfície da água.

Figura 5.4.16: Dois indivíduos parados dentro da piscina e vários objectos a flutuar e na borda da picina.

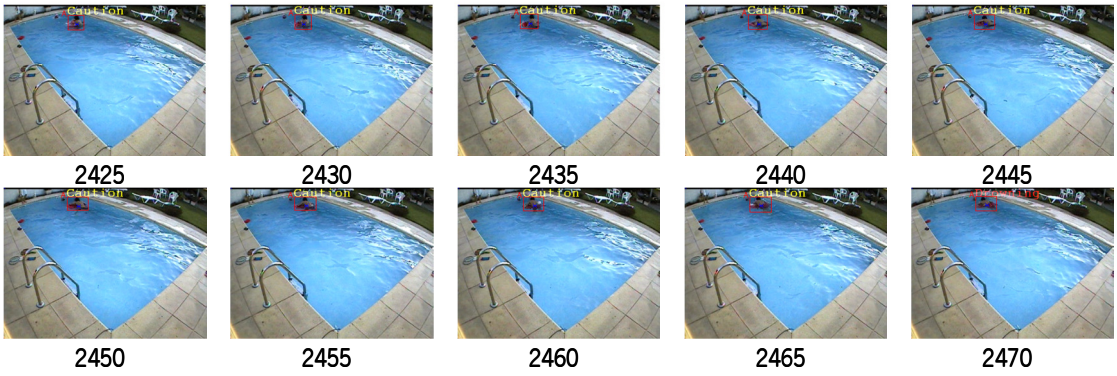


Figura 5.4.17: Um único indivíduo parado na piscina.

Por último, na Figura 5.4.21 são mostrados fotogramas do momento em que o indivíduo B entra num estado de aflição. Repare-se na existência da marcação de falsos objectos devido a erros prolongados no módulo de seguimento. Apesar de breves, estes erros podem por vezes ocorrer, embora por norma desapareçam rapidamente devido às consecutivas filtragens que ocorrem ao longo dos processos de segmentação e seguimento.

Figura 5.4.18: Apenas dois indivíduos na piscina. O indivíduo C nada por baixo de água e o A encontra-se parado.

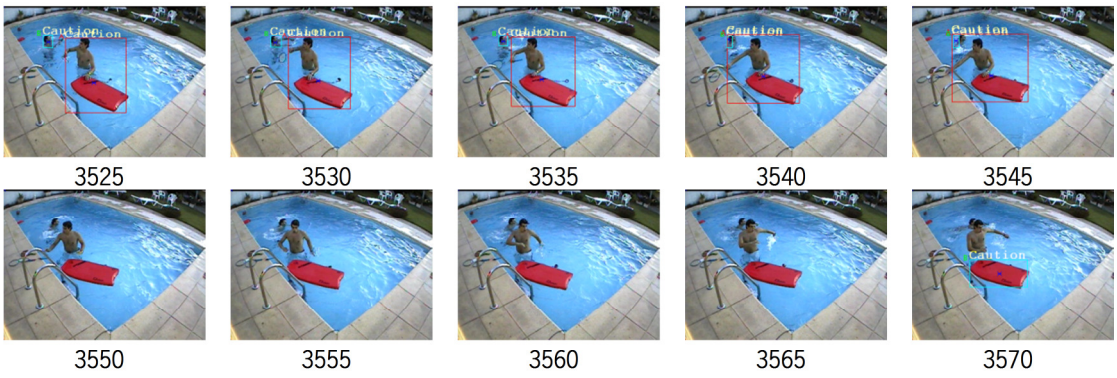


Figura 5.4.19: Um indivíduo com uma prancha e outro atrás dele. Uma situação clara de oclusão.

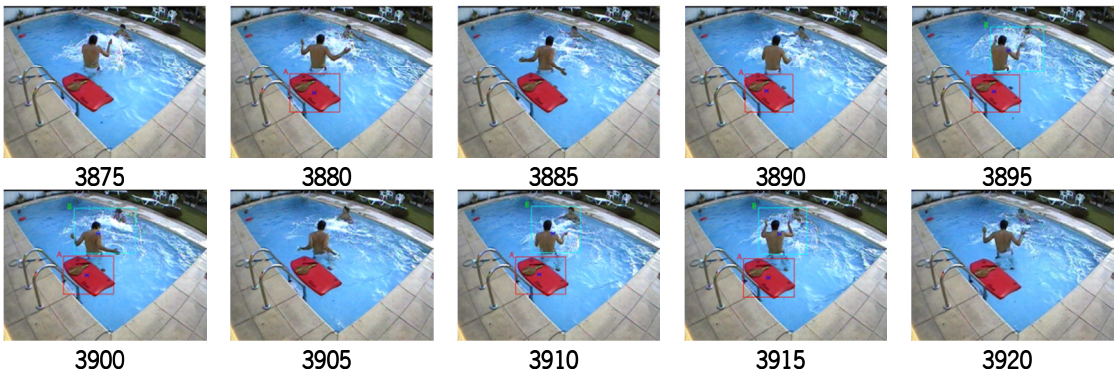


Figura 5.4.20: Um objecto isolado e dois indivíduos atirando água um ao outro.

Este último teste fornece uma imagem fiel do ambiente caótico associado à utilização de uma piscina doméstica. Estas situações são raras numa piscina pública onde os nadadores nadam certinhos dentro dos seus trilhos, tal como os autores dos sistemas já mencionados destacam. A detecção de afogamento em piscinas domésticas apresenta outros desafios, nomeadamente a

existência de objectos e comportamentos completamente imprevisíveis por parte dos seus utilizadores.

Figura 5.4.21: O indivíduo B está em aflição e o A encontra-se parado. Verifica-se a existencia de falsos objectos, C e D, devido à reflexão do mundo exterior na superfície da água e os consequentes erros gerados no módulo de segmentação.

Capítulo 6

6 Conclusões

6.1 Sumário

Este trabalho consistiu no desenvolvimento de um sistema de detecção de afogamento para piscinas domésticas baseado em vídeo. Foram desenvolvidos vários módulos de processamento numa abordagem em camadas de modo a diminuir a complexidade do sistema. O sistema descrito nesta tese é composto pelos módulos de segmentação, seguimento de objectos e reconhecimento e análise de comportamento dos objectos detectados. O primeiro módulo é responsável pela segmentação de objectos em movimento, ou seja, marca todas as regiões da imagem proveniente da câmara onde exista movimento, à excepção da água da piscina. Para tratar os problemas associados a um ambiente aquático complexo e sujeito a todas as condições de luminosidade próprias de uma cena exterior foram desenvolvidas novas metodologias que filtram o movimento da água mantendo a detecção de objectos que não fazem parte da mesma. Foi ainda desenvolvido um esquema que permite melhorar a qualidade da segmentação, através de um segundo passo de segmentação na região ocupada pela *bounding box* expandida que envolve o objecto. No módulo de seguimento, que recebe o mapa binário resultante da segmentação de movimento é efectuado o seguimento de vários objectos em simultâneo com possibilidade de entradas e saídas bem como oclusões totais e parciais. Este processo utiliza um esquema baseado num algoritmo de correspondência não balanceado de similaridades entre objectos detectados no fotograma actual e previstos por um filtro de *kalman*. O seguimento de vários objectos é um passo fundamental para poder inferir o comportamento dos objectos, pois se não forem seguidos não existe um histórico credível das suas características. Por último foi desenvolvido um módulo de reconhecimento de objectos e análise do seu comportamento baseado nas características medidas a partir do seguimento dos mesmos. Este módulo assenta na classificação de padrões de características baseada na regra de *Bayes* a partir de classificadores concebidos por intermédio de dados de treino, ou seja, através de aprendizagem supervisionada. Sempre que padrões característicos de um afogamento são detectados em pessoas que se encontram dentro da piscina é accionado um alerta o qual corresponde ao objectivo do sistema.

6.2 Discussão

A entrada do sistema corresponde a sequências de imagens provenientes de uma câmara e a sua saída corresponde à identificação e reconhecimento de pessoas e objectos bem como à análise e inferência do seu comportamento no mundo exterior capturado por essa mesma câmara. Trata-se de um resumo da realidade através de uma sequência de 25 matrizes por segundo com milhares de pontos por matriz que podem variar entre 0 e $256^3 - 1$. Estas matrizes são uma projecção a duas dimensões do mundo exterior. A enormidade das conjugações faz com que seja intratável o varrimento de todos os casos, fazendo também com que todos os problemas de análise de imagem sejam hoje em dia um enorme desafio tecnológico. O sistema desenvolvido percorreu áreas de investigação de vanguarda no que respeita à análise de vídeo e processamento de imagem, seguimento de objectos em movimento, aprendizagem e reconhecimento de padrões. Deste modo, o facto de existirem ainda muitas lacunas nestes domínios desencadeia o aparecimento de novos caminhos e de novas soluções sendo áreas extremamente ricas no âmbito da investigação.

Como não podia deixar de acontecer o trabalho apresentado nesta tese deixa ainda algumas lacunas por solucionar, embora a comunidade científica já se tenha debruçado sobre algumas delas. O módulo de segmentação é o mais importante, pois todos dependem da saída que o mesmo gera, sendo notório que todos os erros nesta fase terão inevitavelmente repercussões drásticas nas saídas fornecidas pelos módulos subsequentes. Assim, neste trabalho ficou por tratar o problema das sombras, que apesar de não ser crítico, por estas serem consideradas *foreground* apenas no exterior da piscina, representa ainda assim alguns problemas para o módulo de seguimento. O volume de processamento do módulo de segmentação não é constante uma vez que o segundo passo de segmentação sobre os objectos detectados implica uma carga de processamento tanto maior quanto maior for o número de objectos detectados. No módulo de seguimento por vezes existem trocas de identidade dos objectos devido à longa duração das oclusões. Este módulo não é capaz de tratar por longos períodos de tempo as oclusões, pois existem muitas possibilidades no sentido da previsão da posição dos objectos escondidos. Para este problema a solução poderá passar pela colocação de mais que uma câmara capturando a mesma localização mas de um ângulo diferente, fazendo com a oclusão só apareça numa das imagens utilizando as outras para a resolver. No que respeita à remoção das sombras a solução de várias câmaras permite determinar se o objecto detectado tem volume ou não, detectando-se

assim as sombras, pois estas são projecções sem volume. Por último, o módulo de reconhecimento de comportamento poderia ser capaz de aprender automaticamente os padrões associados ao ambiente onde se encontra, *online learning*. Ao longo do tempo o sistema poderia recolher informações sobre o mundo exterior e criar padrões representativos do mesmo que lhe permitisse resumir a realidade à qual está exposto detectando comportamentos diferentes da normalidade. Esta última lacuna continua a ser a busca incessante que todos os investigadores de máquinas com inteligência natural, i.e., equiparada à do ser humano, procuram.

6.3 Trabalho Futuro

Tal como foi mencionado na secção anterior existe um conjunto de lacunas neste sistema que ao serem resolvidas aumentariam a sua fiabilidade e robustez. A resolução dessas lacunas é o trabalho futuro que mantém uma relação mais próxima com o trabalho desenvolvido nesta tese. Contudo, a transformação deste protótipo num produto lança novos desafios em outras áreas. Desta forma, o trabalho futuro passa inevitavelmente pela implementação do módulo de segmentação em hardware numa câmara IP inteligente. Esta câmara teria a possibilidade de fornecer uma saída directamente em HSV e o mapa binário resultante da segmentação já com as regiões identificadas. Através de uma rede destas câmaras e uma máquina central que efectuasse a fusão dos mapas binários seria possível executar os módulos subsequentes, nomeadamente o seguimento e a inferência de comportamento, com melhorias na eliminação de sombras e tratamento de oclusões. Também existe a possibilidade de colocar o sistema completo numa câmara IP, sendo esta a solução mais elegante no sentido que apenas seria necessária uma máquina para receber e gravar o vídeo analisado num formato comprimido, tal como o H264, podendo o acesso ser feito remotamente, inclusive para efectuar a parametrização do sistema. No fundo tratar-se-ia de uma câmara para detecção de afogamento, ou seja, um dispositivo fechado que pode ser utilizado em conjunto com um computador pessoal comum sem depender deste para efectuar a vigilância automática da piscina.

Bibliografia

A. Elgammal, R. D. (2002). Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. *Proc. IEEE*, *90.*

Adamson, P., Micklewright, J., & Wright, A. (2001). *A league table of child deaths by injury in rich nations.* Florence.

Adrienne, G., & John, D. A. (14 de Abril de 2010). *Concept Gallery | GEOG 486: Cartography and Visualization*. Obtido em 24 de Junho de 2010, de Lesson 2: Creating a Reference Map for Use in Emergency Management - Week 1: https://www.e-education.psu.edu/geog486/l2_p9.html

Avidan, S. (2001). Support vector tracking. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (pp. 184-191).

Beeck, E. v., Christine, B., Szpilman, D., Modell, J., & Bierens, J. (2006). Definition of Drowning. In C. Branche, C. Brewster, R. Brons, H. Daanen, D. Elliott, H. Gelissen, et al., *Handbook on Drowning - Prevention, Rescue and Treatment* (pp. 45-49). Joost J.L.M. Bierens.

Bertalmio, M., Sapiro, G., & Randall, G. (2000). Morphing active contours. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 733-737.

Beymer, D., McLauchlan, P., Coifman, B., & Malik, J. (1997). A real-time computer vision system for measuring traffic parameters. *Conf. on Computer Vision and Pattern Recognition, IEEE Computer Society.*

Bilmes, J. A. (1998). *A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models.* Berkeley.

Black, M., & Jepson, A. (1998). Eigentracking: Robust matching and tracking of articulated objects using a view-based representation. *International Journal of Computer Vision*, 63-84.

Boshuizen, H. C., Treurniet, F. H., & Marteloh, M. P. (1997). *Atlas of mortality in Europe.* Geneve: World Health Organization.

Bourgeois, F., & Lassalle, J.-C. (1971). An extension of the Munkres algorithm for the assignment problem to rectangular matrices. *Communications of the ACM*, (pp. 802-804).

Braun, C. L., & Smirnov, S. N. (1993). Why Is Water Blue? *Journal of Chemical Education*, 612-614.

Butler, D. E., Bove Jr., M. V., & Sridharan, S. (2005). Real-Time Adaptative Foreground/Background Segmentation. *EURASIP Journal on Applied Signal Processing*, 2292-2304.

C. Stauffer, W. E. (2000). Learning patterns of activity using real-time tracking. *IEEE Transactions Pattern Analysis Machine Intelligence.*

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 679-698.

Christensen, M., & Alblas, R. (2000). V2 - desig issues in distributed visual surveillance systems. Denmark.

Coblentz, A., Mollard, R., & Cabon, P. (2001). *Bibliographic Study on Lifeguard Vigilance.* Paris, France.

Collins, R. T., Lipton, A. J., Fujiyoshi, H., & Kanade, T. (2001). Algorithms for cooperative multisensor surveillance. *Proc. IEEE 1989.*

Collins, T. R., Lipton, J. A., Kanade, T., Fujiyoshi, H., Duggins, D., Tsin, Y., et al. (2000). *A System for Video Surveillance Monitoring.*

Comaniciu, D., & Meer, P. (2002). Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (pp. 603-619).

Comaniciu, D., Ramesh, V., & Meer, P. (2003). Kernel-based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 564-575.

Computer Aided Drowning Detection. (s.d.). Obtido em 26 de Junho de 2010, de Poseidon: http://www.poseidon-tech.com/us/system.html

Costa, L., & Cesar Jr., R. (2001). Simple Complexity Descriptors. In L. Costa, & R. Cesar Jr., *Shape Analysis and Classification - Theory an Practice.* CRC Press.

CPSC. (2001). *United States Consumer Product Safety Commission. How to plan for the unexpected: Preventing Child Drownings.* Washington.

Cucchiara, R., Grana, C., & Piccard, M. (2003). Detecting Moving Objects, Ghosts, and Shadows in Video Streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* Madison, USA.

Cucchiara, R., Grana, C., & Piccardi, M. (2001). Improving shadow suppression in moving object detection with HSV color information. *Proceedings of IEEE Intelligent Transportation Systems Conference.* Oakland, USA.

Dalal, N., & Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. *IEEE Computer Vision and Pattern Recognition.*

Dempster, A., Laird, N., & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, 39 (SeriesB)*, 1-38.

Detec. (2004). *DETEC - Detec AS*. Obtido em 22 de Dezembro de 2006, de http://www.detec.no

Duda, R. O., Hart, P. E., & Stork, D. G. (2000). Bayesian decision theory. In R. O. Duda, P. E. Hart, & D. G. Stork, *Pattern Classification* (pp. 3 - 65). John Wiley & Sons, Inc.

Elgammal, A., Hardwood, D., & Davis, L. (2000). Non-parametric model for background subtraction. *Proceedings ECCV*, (pp. 751-767). Dublin, Ireland.

Ellis, J., & White, J. (2000). *National Pool and Waterpark Lifeguard Training.* Jones & Bartlett Publishers.

Eng, H.-L., Toh, K.-A., Yau, W.-Y., & Wang, J. (2008). DEWS: A Live Visual Surveillance System for Early Drowning Detection at Pool. *IEEE Transactions on Circuits and Systems for Video Technology*, 196-210.

Eng, H.-L., Toh, K.-A., Yau, W.-Y., & Wang, J. (2008). DEWS: A Live Visual Surveillance System for Early Drowning Detection at Pool. *IEEE Transactions on Circuits and Systems for Video Technology*, 197-209.

Eng, H.-L., Wang, J., Kam, A. H., & Yau, W.-Y. (2004). Novel region based modeling for human detection within highly dynamic aquatic environment. *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2004).* Washington, DC.

Fei, L., Xueli, W., & Dongsheng, C. (2009). Drowning Detection Based on Background Subtraction. *International Conference on Embedded Software and Systems* (pp. 341-343). Las Vegas, USA: IEEE.

Fieguth, P., & Terzopoulos, D. (1997). Color-based tracking of heads and other mobile objects at video frame rates. *Computer Vision and Pattern Recognition (CVPR)*, (pp. 21-27).

Forney, G. D. (1973). The Viterbi algorithm. *Proceedings IEEE*, (pp. 263-278).

Ge, F., Wang, S., & Liu, T. (2007). New benchmark for image segmentation evaluation. *Journal of Electronic Imaging*, 033011-2.

GeoVision. (1996). *Digital Video Surveillance System - GeoVision Inc.* Obtido em 16 de Janeiro de 2007, de http://www.geovision.com.tw

Gevers, T., & Smeulders, A. W. (1997). Color Based Object Recognition. In *Image Analysis and Processing* (pp. 319-326). The Netherlands: Springer Verlin/Heidelberg.

Gonzalez, R. C., & Woods, R. E. (2001). *Digital Image Processing - Second Edition.* Prentice Hall.

Gotcha. (1996). *Video Surveillance Motion Detection Software for the personal computer*. Obtido em 6 de Janeiro de 2007, de http://www.gotchanow.com/

Hager, G., & Rasmussen, C. (2001). Probabilistic data association methods for tracking complex visual objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 560-576.

Heikkila, J., & Silven, O. (1999). A real-time system for monitoring of cyclists and pedestrians. *2nd IEEE Int. Workshop on Visual Surveillance.* Colorado.

Horn, B. (1997). Sequencial Labeling Algorithm. In B. Horn, *Robot Vision* (pp. 69 - 71). Cambridge, Massachusetts: McGraw-Hill Book Company.

Horn, B. K., & Schunk, B. G. (1980). Determining Optical Flow. *Artificial Intelligence*, 185-203.

Huttenlocher, D., Noh, J., & Rucklidge, W. (1993). Tracking nonrigid objects in complex scenes. *IEEE International Conference on Computer Vision (ICCV)*, (pp. 93-101).

Ivanov, Y., Stauffer, C., Bobick, A., & Grimson, W. (1999). Video surveillance of interactions. *2nd IEEE Int. Workshop onVisual Surveillance.* Colorado.

Jepson, A., Fleet, D., & Elmaraghi, T. (2003). Robust online appearence models for visual tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1296-1311.

Jian-Guang, L., Qi-Feing, L., Tie-Niu, T., & Wei-Ming, H. (2003). 3-D model based visual traffic surveillance. *Acta Automatica Sinica.*

Jilkov, V. P., Huimin, C., Li, R., & Nguyen, T. (2006). Feature Association for Object Tracking. *9th International Conference on Information Fusion*, (pp. 1-8). Florença.

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 35-45.

Kang, J., Cohen, I., & Medioni, G. (2004). Object reacquisition using geometric invariant appearence model. *International Conference on Pattern Recognition (ICPR)*, (pp. 759-762).

Koller, D., Weber, J., Huang, T., Malik, J., Ogasawara, G., Rao, B., et al. (1994). Towards Robust Automatic Traffic Scene Analysis in Real-time. *Proceedings ICPR 1994*, (pp. 126-131). Jerusalem, Israel.

Krug, E. (1999). *Injury: a leading cause of the global burden of disease.* Geneva: World Health Organization.

Krumm, J., Harris, S., Meyers, B., Brumit, B., Hale, M., & Shafer, S. (2000). Multi-camera multi-person tracking for easy living. *Third IEEE International Workshop on Visual Surveillance.* Ireland.

Kullback, S., & Leibler, R. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 79-86.

Lo, B. P., & Velastin, S. A. (2001). Automatic congestion detection system for underground platforms. *International Symposium on Intelligent Multimedia, Video & Speech Processing*, (pp. 158-161). Kowloon Shangri-La, Hong Kong.

Lowe, G. D. (2004). Distinctive Image Features from Scale-Invariant Key-points. *International Journal of Computer Vision*, 91-110.

Lu, W., & Tan, Y.-P. (2004). A Vision-Based Approach to Early Detction of Drowning Incidents in Swimming Pools. *IEEE Transactions on Circuits and Systems for Video Technology*, 159-178.

Luckav, R., & Plataniotis, K. N. (2007). Color Feature Detection. In T. Gevers, J. v. Weijer, & H. Stokman, *Color Image Processing - Methods and Applications* (pp. 206, 207). United Satates of America: Philip A. Laplante, Pennsylvania State University.

MacCormick, J., & Blake, A. (2000). Probabilistic exclusion and partitioned sampling for multiple object tracking. *International Journal of Computer Vision*, 57-71.

Makris, D., Ellis, T., & Black, J. (2004). Bridging the gaps between cameras. *International Conference Multimedia and Expo Taiwan.* Taiwan.

Marchesotti, L., Messina, A., Marcenaro, L., & Regazzoni, C. S. (2003). A cooperative multisensor system for face detection in video surveillance applications. *Acta Automatica Sinica.*

McIvor, A. (2000). Background subtraction techniques. *Proceedings of Image & Vision Computing New Zealand.* Auckland, New Zealand.

Micheloni, C., Foresti, G. L., & Snidaro, L. (2003). A co-operative multi-camera system for video-surveillance of parking lots. *Intelligent Distributed Surveillance Systems Symposium by IEE.* London.

Milestone. (2004). *Milestone Company*. Obtido em 23 de Fevereiro de 2006, de http://www.milestonesys.com

Mitchell, T. M. (1997). Bayesian Learning. In T. M. Mitchell, *Machine Learning* (pp. 154 - 200). McGraw-Hill.

Molesini, G., & Vannoni, M. (2008). Light reflection in a pool under falling rain droplets. *European Journal of Physics*, 403-411.

Mundhenk, T. N. (12 de Junho de 2008). *HSV And H2SV Color Space - ILabWiki*. Obtido em 27 de Setembro de 2008, de ILabWiki: http://ilab.usc.edu/wiki/index.php?title=HSV_And_H2SV_Color_Space&redirect=no

Naylor, M. (24 de Junho de 2003). *ADVISOR*. Obtido em 2007 de Maio de 7, de http://www-sop.inria.fr/orion/ADVISOR/

Nguyen, N. T., Venkatesh, S., West, G., & Bui, H. H. (2003). Multiple camera coordination in a surveillance system. *Acta Automatica Sinica.*

Nuuo. (2004). *NUUO Inc. - Global - The Intelligent Surveillance Solution*. Obtido em 4 de Fevereiro de 2007, de http://www.nuuo.com/

Nwagboso, C. (1998). User focused surveillance systems integration for intelligente transport systems. In C. S. Regazzoni, G. Fabri, & G. Vernazza, *Advanced Video-based Surveillance Systems.* Boston: Kluwer Academic Publishers.

Oliver, N. M., Rosario, B., & Pentland, A. P. (2000). A Bayesian computer vision system for modeling human interactions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 831-843.

Osinski, A. (2006). *Aquatic Consulting Services.* San Diego, California.

Paulidis, L., & Morellas, V. (2002). Two examples of indoor and outdoor surveillance systems. *Video-based Surveillance Systems.* Boston: Kluwer Academic Publishers.

Paulidis, L., & Morellas, V. (2002). Two examples of indoor and outdoor surveillance systems. In P. Remagnino, G. A. Jones, N. Paragios, & C. S. Regazzoni, *Video-based Surveillance Systems.* Boston: Kluwer Academic Publishers.

Peixoto, N. P., Cardoso, N. G., Cabral, J. M., Tavares, A. J., & Mendes, J. A. (2009). A Segmentation Approach for Object Detection on Highly Dynamic Aquatic Environments. *35th Annual Conference of the IEEE Industrial Electronics Society (IECON 2009). 24.* Porto, Portugal: IEEE.

Pelágio, L., Polacow, F., & Menezes, H. C. (2004). Acidentes por Submersão em Crianças Recortes de Imprensa - 2002 - 2003. *Campanha de Segurança na Água 2004.*

Pellegrini, M., & Tonami, P. (1998). Highway traffic monitoring. In C. S. Regazzoni, G. Fabri, & G. Vernazza, *Advanced Video-based Surveillance Systems.* Boston: Kluwer Academic Publishers.

Pia, F. (1974). Observations on the drowning of nonswimmers. *Journal of Physical Education*.

Piccardi, M. (2004). Background Subtraction Techniques: a review. *IEEE International Conference on Systems, Man and Cybernetics*, (pp. 3099-3104). The Hague, The Netherlands.

Ping, L., Lo, B., Sun, J., & Velastin, S. A. (2003). Fusing visual and audio information in a distributed intelligent surveillance system for public transport systems. *Acta Automatica Sinica.*

Power, W. P., & Schoonees, J. A. (2002). Understanding Background Mixture Models for Foreground Segmentation. *Proceedings Image and Vision Computing*.

Pozzobon, A., Sciutto, G., & Recagno, V. (1998). Security in ports: the user requirements for surveillance system. In C. S. Regazzoni, G. Fabri, & G. Vernazza, *Advanced Video-based Surveillance Systems.* Boston: Kluwer Academic Publishers.

Reid, D. B. (1979). An algorithm for tracking multiple targets. *IEEE Transactions on Automation and Control*, 560-576.

Remagnino, P., Baumberg, A., Grove, T., Hogg, D., Tan, T., Worral, A., et al. (1997). An integrated traffic and pedestrian model-based vision system. *BMVC97.* Israel.

Ronetti, N., & Dambra, C. (2000). Railway station surveillance: the Italian case. In G. L. Foresti, P. Mahonen, & C. S. Regazzoni, *Multimedia Video Based Surveillance Systems.* Boston: Kluwer Academic Publishers.

Rosales, R., & Sclaroff, S. (1999). 3D trajectory recovery for tracking multiple objects and trajectory guided recognition actions. *Conference on Computer Vision and Pattern Recognition (CVPR)*, (pp. 117-123).

Sato, K., & Aggarwal, J. (2004). Temporal spatio-velocity transform and its application to tracking and interaction. *Computer Vision and Image Understanding*, (pp. 100-128).

Seki, M., Wada, T., Fujiwara, H., & Sumi, K. (2003). Background subtraction based on cooccurrence of image variations. *Proceedings CVPR*, (pp. 65-72). Madison, USA.

Sethi, I., & Jain, R. (1987). Finding trajectories of feature points in a monocular image sequence. *IEEE Transactions on Pattern Analysis and Machine Intelligence 9*, 56-73.

Sethi, I., & Salari, V. (1990). Feature point correspondence in the presence of occlusion. *IEEE Transactions on on Pattern Analysis and Machine Intelligence*, 87-91.

Shafer, S. A. (1984). *Using Color to Separate Reflection Components.* New York.

Shafique, K., & Shah, M. (2003). A non-iterative greedy algorithm for multi-framepoint correspondence. *IEEE International Conference on Computer Vision (ICCV)*, (pp. 110-115).

Shi, J., & Tomasi, C. (1994). Good features to track. *Computer Vision and Pattern Recognition (CVPR)*, (pp. 593-600).

Soille, P., & Gratin, C. (2004). Fillhole. In P. Soille, *Morphological Image Analysis: Principles and Applications* (p. 208). Italy: Springer-Verlag.

Stauffer, C., & Grimson, W. (1999). Adaptive background mixture models for real-time tracking. *IEEE CVPR 1999*, (pp. 246-252). USA.

Streit, R. L., & Luginbuhl, T. E. (1994). Maximum likelihood method for probabilistic multi-hypotesis tracking. *Proceedings of the International Society for Optical Engineering (SPIE)*, (pp. 394-405).

Tan, Y.-P., & Lu, W. (2002). A Camera-Based Systemfor Early Detection of Drowning Incidents. *IEEE ICIP 2002*, (pp. 445-448). New York.

Tanizaki, H. (1987). Non-gaussian state-space modeling of nonstationary time series. *J. Amer Statistics Association 82*, 1032-1063.

Tao, H., Sawhney, H., & Kumar, R. (2002). Object tracking with bayesian estimation of dynamic layer representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 75-89.

Terzopoulos, D., & Szeliski, R. (2002). Tracking with kalman snakes. *Active Vision*, 75-89.

Toh, K.-A., Tran, Q.-L., & Srivivasan, D. (2004). Benchmarking a reduced multivariate polynomial pattern classifier. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 740-755.

Valera, M., & Velastin, S. A. (2005). Intelligent distributed surveillance systems: a review. *IEEE Proceedings Visual Image Signal Processing*, 192-204.

Veenman, C., Reinders, M., & Backer, E. (2001). Resolving motion correspondence for densely moving points. *IEEE Transactions on on Pattern Analysis and Machine Intelligence*, 54-72.

Velastin, S. A. (17 de Agosto de 2002). *CROMATICA - DIRC - Kingston*. Obtido em 22 de Fevereiro de 2006, de Digital Image Research Centre - CROMATICA: http://dilnxsrv.king.ac.uk/cromatica/

Wren, C., Azarbayejani, A., Darrell, T., & Pentland, A. P. (1997). Pfinder: real-time tracking of the human body. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 780-785.

Wu, F., Bell, J. W., & Schweitzer, H. (2002). Very fast template matching. *European Conference on Computer Vision (ECCV)*, (pp. 358-372).

Xu, M., Lowey, L., & Orwell, J. (2004). Architecture and algorithms for tracking football players with multiple cameras. *Proc. IEEE Workshop on Intelligent Distributed Surveillance Systems.* London.

Yilmaz, A., Javed, O., & Shah, M. (2006). Object Tracking: A Survey. *ACM Computer Surveys*.

Yilmaz, A., Li, X., & Shah, M. (2004). Contour based object tracking with occlusion handling in video acquired using mobile cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1531-1536.

Yuan, X., Sun, Z., Varol, Y., & Bebis, G. (2003). A distributed visual surveillance system. *IEEE Conference on Advanced Video and Signal based Surveillance.* Florida: 2003.

Zhang, H., Fritts, J. E., & Goldman, S. A. (2008). *Image Segmentation Evaluation: A Survey of Unsupervised Methods.* Elsevier.

Zhou, H., Yuan, Y., & Shi, C. (2008). Kernel-Based method for tracking objects with rotation and translation. *Journal of Computer Vision*.