

Faculdade de Engenharia da Universidade do Porto



FEUP

Detecção e seguimento de objectos em vídeo sob condições adversas

João Nuno Couto dos Santos

Relatório final da Preparação da dissertação realizada no âmbito do
Mestrado Integrado em Engenharia Electrotécnica e de Computadores
Major Telecomunicações

Orientador: Luís Corte-Real
Proponente: Pedro Carvalho

15 de Fevereiro de 2011

© João Nuno Couto dos Santos, 2011

Resumo

A Detecção e seguimento automático de objectos em vídeo é uma técnica utilizada nos sistemas de vigilância actuais. São técnicas que permitem que os sistemas funcionem de forma automática, minimizando assim o uso de recursos humanos.

Actualmente as imagens mais utilizadas são imagens naturais que têm algumas desvantagens em cenários com problemas de iluminação, não sendo assim apropriadas para casos onde é necessária a vigilância nocturna ou com condições ambientais adversas como chuva e nevoeiro.

Pretende-se com este trabalho desenvolver um algoritmo de detecção e seguimento de vídeo que obtenha bons resultados, sob condições adversas, nomeadamente condições de iluminação desapropriadas, incluindo luz brilhante e escuridão.

Serão abordados aspectos fundamentais da tecnologia, tais como técnicas de detecção e seguimento, imagens infravermelhas e técnicas de avaliação do algoritmo implementado.

Abstract

Automatic Detection and tracking of objects in video images is a technique used in surveillance systems today. These techniques allow systems to operate automatically, thus minimizing the need for human resources.

Currently, natural images are used more, and they have some disadvantages in scenes with poor lightning, so they aren't suitable for cases where night surveillance is needed or in adverse environmental conditions such as rain and fog.

The aim of this work is to develop a video detection and tracking algorithm that can have good results under adverse conditions, such as inappropriate lightning conditions, including bright light and darkness.

It will address fundamental aspects of the technology, such as techniques for detection and tracking, infrared images and evaluation techniques for the implemented algorithm.

Índice

Resumo	iii
Abstract.....	v
Índice.....	vii
Lista de figuras	viii
Abreviaturas e Símbolos	ix
Capítulo 1	1
1.1 - Introdução.....	1
1.2 - Objectivos.....	2
1.3 - Estrutura do documento	2
Capítulo 2	3
Estado da arte	3
2.1 - Trabalho relacionado.....	3
2.2 - Imagens infravermelhas	4
2.2.1 - Near-Infrared.....	5
2.2.2 - Far-Infrared	6
2.3 - Detecção e Seguimento.....	7
2.3.1 - Tracking usando Imagens Naturais	7
2.3.2 - Tracking usando Imagens Térmicas	10
2.3.3 - Tracking usando Imagens Naturais/Térmicas.....	15
2.4 - Avaliação de algoritmos	17
Capítulo 3	19
Trabalho Desenvolvido	19
Capítulo 4	21
Conclusões	21
Referências	22

Lista de figuras

Fig. 1 - Esquema da Radiação Solar	4
Fig. 2 - Espectro electromagnético.....	5
Fig. 3 - Exemplo de imagem infravermelha.....	5
Fig. 4 - Exemplo de imagens térmicas - Copos de plástico com líquidos com temperaturas diferentes (esquerda), imagem em escala de cinzento (meio), imagem com escala de cores (direita)	6
Fig. 5 - Arquitectura do sistema descrito em [7]	8
Fig. 6 - Processo de segmentação utilizado em [8].....	9
Fig. 7 - Fluxograma de segmentação utilizado em [13].....	11
Fig. 8 - (a) Região Candidata, (b) Histograma de intensidade, (c) Correção Vertical [14]	12
Fig. 9 - (a) Sub-região, (b) Sub-região binarizada, (c) Correção Horizontal [14]	13
Fig. 10 - Representação gráfica do algoritmo do método utilizado em [15]	14
Fig. 11 - Exemplo de imagens naturais e térmicas para ligeiro nublado (topo), vento forte e chuva (meio) e neve (baixo) [16]	15
Fig. 12 - Imagens utilizadas para teste [16]	16
Fig. 13 - Resultados: Ground Truth (1ª linha), imagens naturais (2ª linha), imagens térmicas (3ª linha) e combinação das duas (4ª linha) [16].....	17
Fig. 14 - Esquema da framework de avaliação	20

Abreviaturas e Símbolos

Lista de abreviaturas (ordenadas por ordem alfabética)

3D	Três Dimensões
BB	<i>Bounding Box</i>
CBIR	<i>Content-Based Image Retrieval</i>
FIR	<i>Far-Infrared</i>
FPS	Frames por segundo
GT	<i>Ground-Truth</i>
MIR	<i>Mid-wavelength Infrared</i>
MNRL	<i>Maximum Negative Run-Length</i>
NIR	<i>Near-Infrared</i>
NGT	<i>Nonground-Truth</i>
PD	<i>Partition Distance</i>
RS	<i>Reference Segmentation</i>

Lista de símbolos

μm	Micrómetros
---------------	-------------

Capítulo 1

1.1 - Introdução

As técnicas utilizadas para fazer a vigilância de sectores públicos ou privados da nossa sociedade, implicam que um funcionário ou grupo de funcionários, permaneçam durante longas horas em frente a ecrãs. Para as empresas, esta função implica um aumento de pessoal e consequentemente dos custos que não se revertem num aumento de qualidade do serviço. Isto porque um homem, ao contrário de uma máquina, é afectado por condições como o cansaço e distrações, que se propiciam em trabalhos monótonos e rotineiros, podendo comprometer a segurança do local. Outro factor a considerar é o tempo de resposta a uma ocorrência. Quanto maior o tempo decorrido entre a detecção da ocorrência e o alerta da mesma, mais comprometida pode ficar a segurança do local.

Para minimizar estes problemas, são utilizados métodos de visão computacional, que consistem em dotar um sistema automatizado com uma ferramenta de visão artificial para que possa tomar decisões consoante os acontecimentos que detectar. Isto só é possível graças ao constante desenvolvimento tecnológico que permite a utilização de técnicas cada vez mais avançadas.

Um exemplo deste tipo de visão são algoritmos que, de forma automática, fazem o seguimento de objectos em movimento. Actualmente grande parte destes métodos são eficientes em ambientes favoráveis, isto é, ambientes que têm um conjunto de características que facilitam a detecção, como por exemplo boas condições de iluminação, não havendo incidência de luz directa na câmara ou pouca iluminação. No entanto, se estas condições não estiverem presentes a eficácia do algoritmo torna-se reduzida e faz com que aumente a taxa de erros. Pretende-se então desenvolver soluções que consigam, através do uso de outras técnicas, uma boa eficácia de detecção quando as condições ambientais não são as mais indicadas, nomeadamente a presença de chuva, poeira, nevoeiro, e problemas de iluminação. Por

exemplo, um ambiente com luz muito brilhante e em excesso, ou por outro lado com pouca luz, não vai permitir obter uma imagem com qualidade suficiente para se obter bons resultados

1.2 - Objectivos

A eficiência da detecção e seguimento de objectos está actualmente muito dependente das condições de iluminação, atmosféricas ou de cenário. Torna-se assim importante investigar uma forma alternativa de melhorar o seguimento de objectos.

O objectivo desta Dissertação é desenvolver um algoritmo capaz de realizar a detecção e seguimento de objectos, quando as condições ambientais não se adequam ao bom funcionamento das técnicas de segmentação convencionais, nomeadamente com problemas de iluminação, fumos ou poeiras.

1.3 - Estrutura do documento

No capítulo seguinte serão abordadas com mais detalhe as tecnologias existentes sobre a captação de imagens infravermelhas, e o estado da arte sobre as técnicas de seguimento e detecção de pessoas. No capítulo 3 será descrito o trabalho desenvolvido até ao momento e finalmente serão apresentadas conclusões no capítulo 4.

Capítulo 2

Estado da arte

2.1 - Trabalho relacionado

Para a análise do estado da arte foram estudados artigos que permitiram adquirir conhecimentos sobre as tecnologias utilizadas para fazer a detecção e seguimento de objectos. Devido ao facto dos sistemas de detecção e seguimento serem muito sensíveis às constantes mudanças de iluminação, a que são expostos nos cenários exteriores, procurou-se uma alternativa às imagens naturais frequentemente utilizadas. Assim verificou-se que as imagens infravermelhas poderiam ser uma alternativa viável às imagens naturais, uma vez que são imunes às mudanças de iluminação.

Foi então definido que o tipo de algoritmo a usar na solução que se pretende estudar, deve ser capaz de seguir objectos em condições ambientais adversas. No algoritmo a implementar deverá ser tido em conta as capacidades actuais dos sistemas existentes e ultrapassar as dificuldades técnicas que alguns deles manifestam.

Estas tecnologias serão descritas mais detalhadamente nos subcapítulos seguintes.

2.2 - Imagens infravermelhas

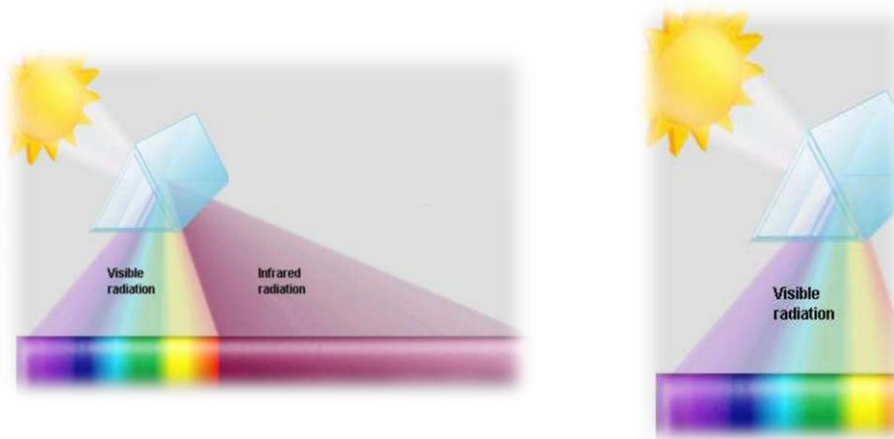


Fig. 1 - Esquema da Radiação Solar

O sol “bombardeia-nos” com uma extensa gama de energia, mas dessa gama apenas uma parte é detectada pelos nossos olhos.

Colocando um prisma num raio de luz solar é possível identificar as cores que fazem parte da radiação visível (Fig. 1). A radiação que não é detectada pelos nossos olhos denomina-se de radiação invisível, e é composta por diversos tipos de radiações, entre eles a radiação infravermelha.

O espectro electromagnético é o intervalo de todas as frequências da radiação electromagnética. O espectro electromagnético de um objecto é a distribuição característica da radiação electromagnética emitida ou absorvida por esse objecto em particular, estando dividido em vários campos (Fig. 2). A luz visível encontra-se após a radiação ultravioleta, entre os comprimentos de onda de $0,4\mu\text{m}$ e os $0,7\mu\text{m}$. A partir deste valor entra-se na gama dos raios infravermelhos, estando dividida em três principais zonas, a zona *Near-Infrared (NIR)* ($0,7$ a $2\mu\text{m}$), *Mid-wavelength Infrared (MIR)* (2 a $4\mu\text{m}$) e *Far-Infrared (FIR)* (4 a $1000\mu\text{m}$). Depois destes encontram-se as microondas. Existem dois grandes tipos de sistemas de visão infravermelha utilizados em sistemas de detecção e seguimento de vídeo que funcionam em gamas distintas do espectro, um funciona na gama NIR e outro na FIR. Cada um com as suas vantagens e desvantagens que serão descritas nos capítulos 2.2.1 e 2.2.2 respectivamente.

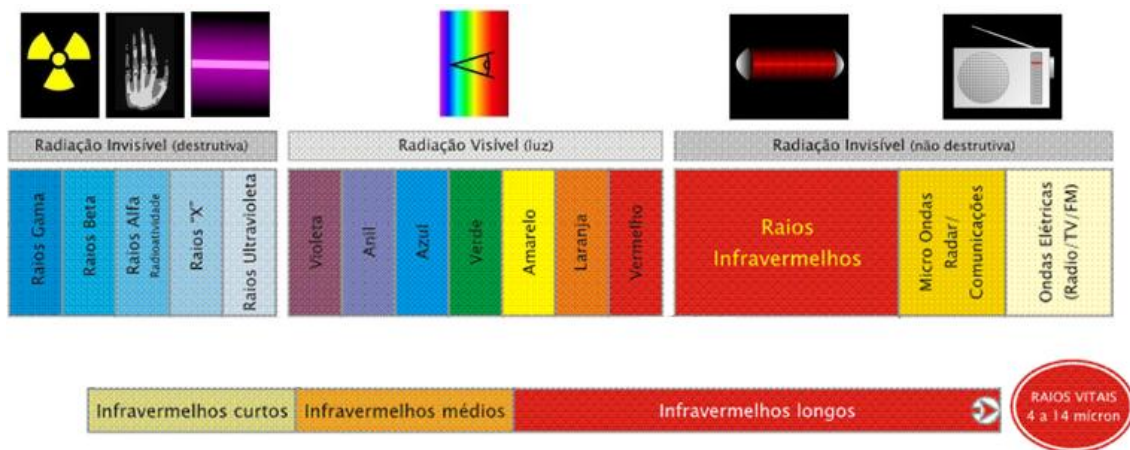


Fig. 2 - Espectro electromagnético

2.2.1 - Near-Infrared

Para o funcionamento das câmaras infravermelhas é necessária a existência de uma fonte de raios infravermelhos própria. Os raios emitidos pela fonte são depois reflectidos pelos objectos e é criada a imagem infravermelha. Este tipo de câmaras tem um consumo elevado de energia pois a fonte necessita de estar ligada para a aquisição das imagens, no entanto consegue-se facilmente distinguir o tipo de objectos presentes na imagem (Fig. 3).

As aplicações típicas destas câmaras passam pela vigilância em locais de acesso restrito, parques de estacionamento, e também pela monitorização de tráfego.



Fig. 3 - Exemplo de imagem infravermelha

2.2.2 - Far-Infrared

As imagens na gama FIR são também conhecidas como imagens térmicas. Estas câmaras absorvem as radiações emitidas pelos objectos e traduzem-nas numa imagem que pode ser a cores ou em escala de cinzento, onde cada intensidade corresponde a uma temperatura (Fig. 4). Segundo a lei de Stefan-Boltzmann, deduzida em 1879, a energia irradiada por um corpo negro é proporcional à sua temperatura absoluta e é dada por $E = \sigma T^4$, onde $\sigma = 5.67 \times 10^8 \text{ W}/(\text{m}^2 \cdot \text{K}^4)$ é a constante de Stefan-Boltzmann, sendo T a temperatura absoluta do corpo em Kelvin. Considerando que os objectos não são radiadores ideais, esta lei expressa-se na forma $E = \varepsilon \sigma T^4$, onde ε é a emissividade do objecto ($\varepsilon = 1$ para radiadores ideais) [1].

Pelo facto de estas câmaras captarem as radiações que são emitidas pelos objectos não necessitam de uma fonte própria de raios infravermelhos. Isto permite que o consumo de energia tenha valores reduzidos. As imagens resultantes têm pouca resolução e não se distingue facilmente todos os objectos que fazem parte da imagem, têm a vantagem de ser possível ver através de materiais como plástico, poeira, fumo, etc. No entanto, quanto maior for a distância ao objecto, menor é a qualidade da captação.

São tipicamente utilizadas como complemento a imagens infravermelhas e também em cenários de vigilância em locais de escuridão total ou com luzes brilhantes, uma vez que são imunes às condições de iluminação.

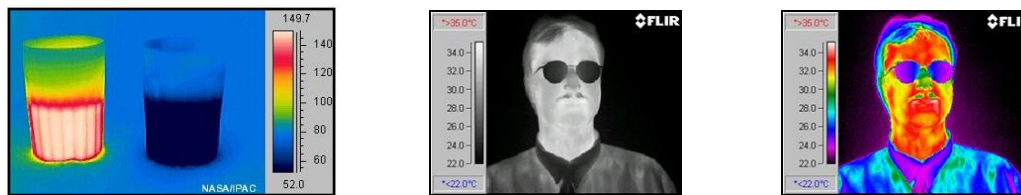


Fig. 4 - Exemplo de imagens térmicas - Copos de plástico com líquidos com temperaturas diferentes (esquerda), imagem em escala de cinzento (meio), imagem com escala de cores (direita)

Tendo em conta as características dos dois tipos de câmaras, quando o consumo de energia é um factor relevante de decisão, deve-se optar por usar as câmaras que funcionam na gama FIR.

2.3 - Detecção e Seguimento

A detecção e seguimento¹ de pessoas tem hoje em dia um papel importante em muitos campos de investigação, como a vídeo vigilância, interacção computador-pessoa [2], e sistemas de ajuda à condução [3]. Com a elevada investigação nesta área, desenvolveram-se vários métodos para realizar estas tarefas. Um destes métodos passa por usar imagens naturais para se fazer o tracking de pessoas. Este é o método mais utilizado uma vez que na maioria dos casos os cenários são espaços interiores, controlados e o uso deste tipo de imagens permite tirar o máximo partido do sistema [4]. Cada vez mais se recorre a sistemas que têm o objectivo de garantir a segurança de locais exteriores como parques de estacionamento, armazéns, etc. Como estes locais são expostos a vários tipos de condições (sol, chuva, nevoeiro, dia, noite), é necessário garantir que mesmo nestas situações as detecções são bem sucedidas, por isso em certos casos é vantajoso usar as imagens naturais em complemento com imagens térmicas [5]. No entanto, em certas situações, apenas o uso de imagens térmicas é suficiente para se obter os resultados desejados [6].

2.3.1 - Tracking usando Imagens Naturais

O tracking de pessoas usando imagens naturais já é utilizada há vários anos, seja para sistemas de vigilância, ou outros fins. Em [7], é apresentada uma técnica que consegue detectar e segmentar pessoas, seccionando toda a silhueta em vez de apenas detectar a posição da pessoa. Para tal, este método combina a análise da forma com o seguimento, localizando as pessoas e as suas partes (braços, pernas, cabeça e tronco) e criando modelos de aparência das mesmas, para que estas possam ser detectadas após ocultações em vez do normal uso de comparação de cores.

Num primeiro passo é utilizado um modelo estatístico de fundo para detectar os pixels do *foreground*, sendo depois agrupados em elementos e determinado um conjunto de características globais e locais de cada um. No segundo passo cada elemento é classificado numa de três classes: uma pessoa, pessoas num grupo ou outros objectos. No primeiro caso, sendo o elemento classificado como pessoa, aplica-se a este, uma análise de silhueta que vai permitir estimar a silhueta da referida pessoa, sendo possível também

¹ A detecção e seguimento será posteriormente referido como tracking

determinar se a pessoa está a segurar algum objecto. Caso tenha sido definido que é um grupo de pessoas então não é possível detectar partes individuais das pessoas, apenas é feita a segmentação do grupo em várias pessoas. Se finalmente for classificada como objecto então não se realiza análise de silhueta, apenas tenta seguir-se o objecto ao longo do vídeo.

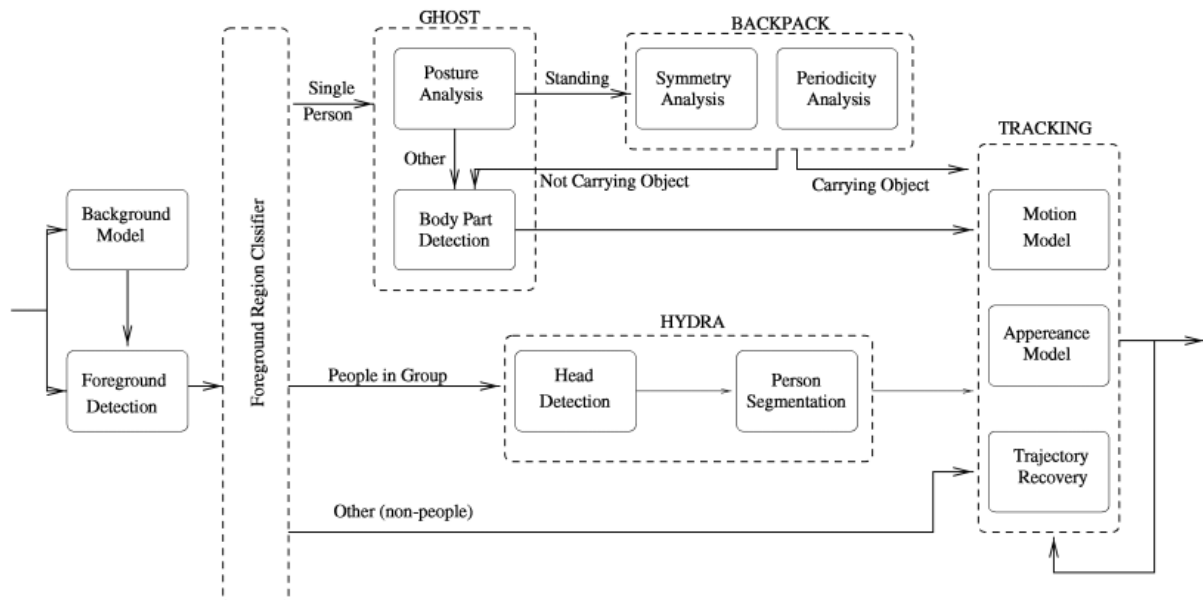


Fig. 5 - Arquitectura do sistema descrito em [7]

Este sistema tem como principais características a distinção de pessoas de outros objectos, detecção das várias partes de uma pessoa obtendo a sua silhueta e também o facto de detectar se uma pessoa está a transportar um objecto, segmentando-o para ser possível ser seguido mesmo quando trocado com outra pessoa.

A performance do sistema descrito acima depende da precisão do sistema de segmentação do fundo, e o sucesso do algoritmo é bastante afectado pela existência de sombras que dificultam a análise de silhueta. Este sistema consegue fazer a detecção num grande intervalo de escalas, no entanto os seus vários algoritmos de análise funcionam melhor quando as pessoas são maiores do que 75 x 50 pixels.

Uma base para o método de tracking de pessoas é descrito em [8]. Neste método utiliza-se uma câmara estática que está situada numa posição elevada, permitindo assim uma visão periférica do cenário, diminuindo o número de oclusões possíveis.

São utilizados vários modelos para segmentar correctamente as pessoas. O modelo de câmara permite calcular as posições físicas das pessoas na imagem, ajudando a calcular as posições 3D (altura e forma) das mesmas. O modelo de fundo ajuda a segmentar as pessoas e distingui-las do fundo. Com o modelo de aparência e de movimento humano é possível distinguir dentro do que foi segmentado, o que é pessoa e o que pode ser ruído, também o modo de locomoção (andar, correr, parado) e eliminar falsas hipóteses.

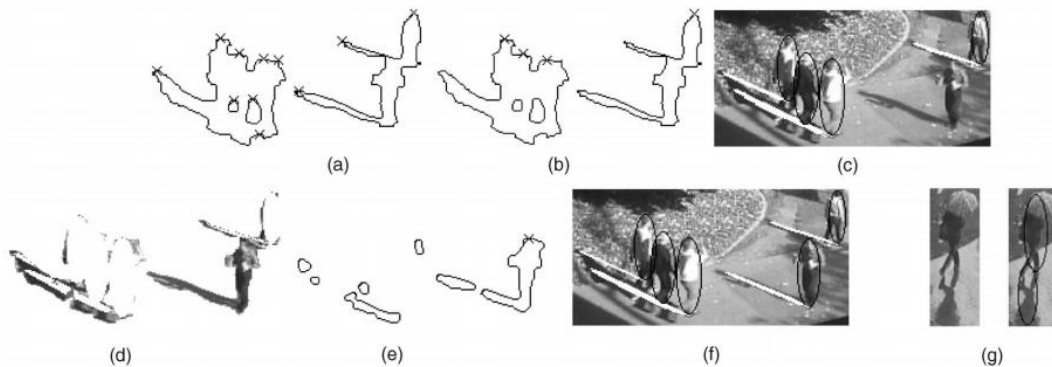


Fig. 6 - Processo de segmentação utilizado em [8]

Para se fazer a segmentação de várias pessoas primeiro procuram-se no *foreground* os pontos mais altos (na vertical) e considera-se candidato a cabeça com um espaçamento definido pelo tamanho médio da cabeça humana (Fig. 6a). Os topos que não têm tamanho suficiente, segundo o modelo de aparência, são descartados (Fig. 6b). As pessoas que são detectadas são retiradas do *foreground* (Fig. 6d) e o método é aplicado do início (Fig. 6e), até não ser encontrada nenhuma pessoa.

No seguimento das pessoas encontradas primeiro é utilizado um filtro de Kalman, que estima a posição onde vai estar a pessoa na frame seguinte, e procura-se a melhor correspondência numa área próxima do ponto estimado. Quando existem oclusões de objectos utiliza-se um mapa de ocupação que nos diz se o pixel que está a ser processado pertence a algum objecto que está a frente. Assim não se corre o risco de utilizar para a definição de um objecto, os pixels de outro que se encontra à sua frente. Finalmente para se definir se um objecto está oculto utiliza-se o parâmetro “ r ” definido por N_v/N_e onde N_v é o número de pixels do *foreground* visíveis e N_e é o número de pixels da máscara. Usam-se dois limiares T_{01} e T_{02} e se $T_{01} > r > T_{02}$ o objecto está parcialmente oculto, se $r < T_{02}$ o objecto está completamente oculto. Se

este se encontrar completamente oculto durante um certo número de frames, é descartado.

Este método produz resultados satisfatórios, mesmo com a presença de sombras, reflexões e ocultações, podendo no futuro ser melhorado para se atingir resultados ainda melhores.

2.3.2 - Tracking usando Imagens Térmicas

O uso de câmaras de infravermelho para detecção de pessoas já é usado em sistemas militares, mas os elevados custos não permitiam o seu uso comercial. Recentemente as câmaras de infravermelho ficaram mais acessíveis, permitindo que o tracking de pessoas usando tecnologia de infravermelhos ofereça uma abordagem alternativa ao desenvolvimento e implementação de sistemas de monitorização de pessoas.

A principal diferença na detecção de radiação térmica em oposição à radiação visível de uma pessoa, é que no caso da radiação térmica os objectos emitem a radiação, enquanto que na radiação visível os objectos reflectem essa radiação. Como a temperatura de uma pessoa é razoavelmente constante, isto faz com que a detecção de pessoas seja menos condicionada por mudanças de ambiente e assim mais fácil de implementar.

Existem, no entanto, algumas desvantagens das imagens térmicas em relação às visíveis. As imagens térmicas representam calor e não têm conteúdo espectral. O fundo térmico não é constante mas varia lentamente comparando com as mudanças de iluminação ambiente que podem ocorrer na detecção visível. O fundo térmico pode estar mais frio ou mais quente que as pessoas numa imagem térmica, este facto denomina-se por inversão de polaridade. Embora as pessoas mantenham uma temperatura constante, o vestuário afecta gravemente a representação e faz com que possa não haver ligação entre a cabeça e as mãos. As imagens térmicas são de menor resolução e contêm bastante ruído. Muitas câmaras térmicas reiniciam-se de alguns em alguns minutos, levando a descontinuidades no seguimento.

Existem várias formas de fazer o tracking de pessoas utilizando imagens térmicas [9-11], incluindo alguns métodos optimizados para a implementação destes sistemas em robots [12].

Em [13], para fazer a segmentação, é utilizado um método de subtracção pixel a pixel do fundo simultaneamente com a aquisição de dados Fig. 7.

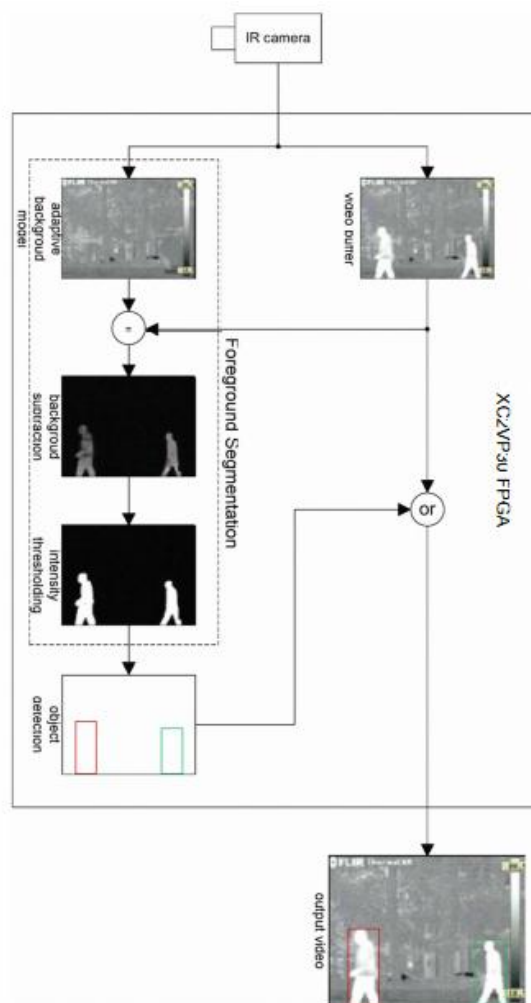


Fig. 7 - Fluxograma de segmentação utilizado em [13]

A característica principal deste sistema é que as pessoas podem ser detectadas à *frame rate* da câmara infravermelha (25 fps). A principal limitação deste sistema é a detecção de apenas um pedestre de cada vez. A localização de uma pessoa é mostrada com uma *Bounding Box* (BB) à volta da pessoa.

Em [14] é apresentada uma forma de se realizar a segmentação de pessoas em imagens térmicas. Primeiro é realizada uma mudança de escala para que todas as imagens sejam normalizadas e se possa trabalhar com valores similares. Depois são eliminados todos os pontos incandescentes que correspondem a lâmpadas, fontes de energia, aparelhos. Isto é feito através da aplicação de um *threshold*, uma vez que estes elementos se encontram mais brilhantes que as pessoas. É realizado posteriormente um novo *threshold*, desta vez com um valor que irá isolar os pontos candidatos a ser pessoas, e é realizada uma operação de abertura e outra de fecho que têm

como finalidade eliminar pontos isolados e unir áreas que tenham sido separadas.

Depois de encontradas as regiões candidatas a conter pessoas é necessário saber o número de pessoas que nelas estão inseridas. Para isso é feito uma correcção das regiões, tanto verticalmente como horizontalmente. É feita uma procura nas colunas pelo máximo valor de intensidade que corresponde à cabeça de uma pessoa, e o número de máximos locais obtidos corresponde ao número de pessoas inseridas nessa região (Fig. 8). Depois de dividida a região para conter apenas uma pessoa é feito a correcção horizontal, que consiste numa pesquisa nas linhas da sub-região e, uma vez mais, verificar quando que os valores de intensidade deixam de estar contidos no limite de intensidade do corpo, encontrando-se assim o comprimento da pessoa (Fig. 9). De notar que uma pessoa pode estar oculta, por exemplo, da cintura para baixo por se encontrar atrás de um objecto refinar.

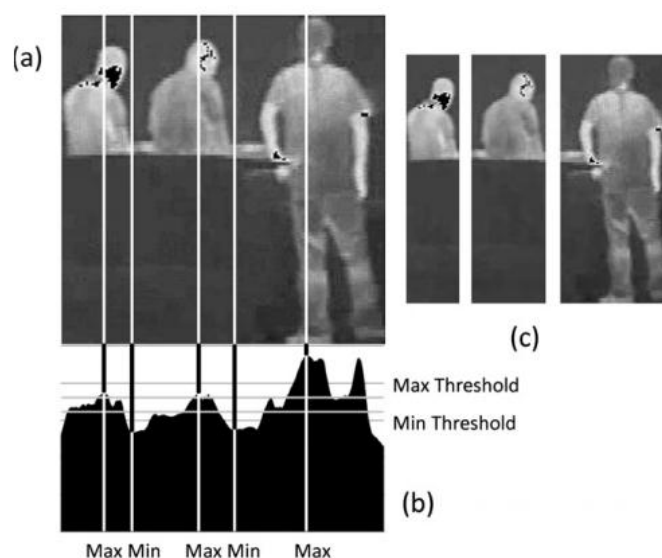


Fig. 8 - (a) Região Candidata, (b) Histograma de intensidade, (c) Correção Vertical [14]

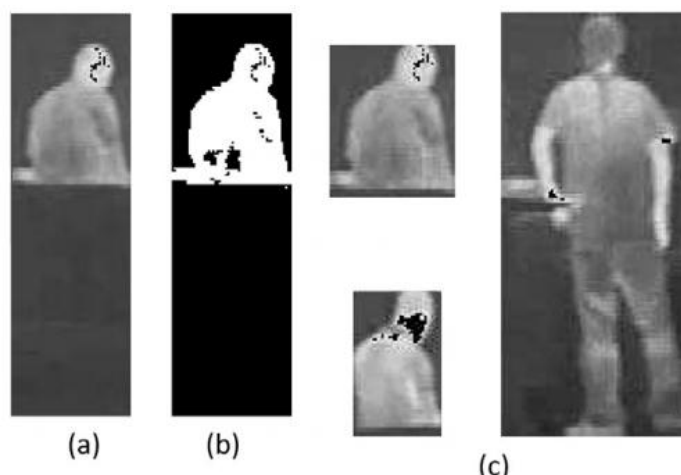


Fig. 9 - (a) Sub-região, (b) Sub-região binarizada, (c) Correção Horizontal [14]

Os resultados deste método de segmentação são positivos, conseguindo que na maioria dos casos os pontos de objectos quentes como lâmpadas, não interfiram na segmentação de uma pessoa, nem sejam interpretados como pessoas. Dependendo do tamanho das imagens térmicas o algoritmo consegue trabalhar a 6 fps para imagens de 768x576, a 13 fps para imagens de 537x403 e finalmente para as imagens de 384x288 o algoritmo trabalha com 26 fps o que significa que poderá estar inserido numa aplicação em tempo real.

É apresentado um método em [15] que tem como objectivo a identificação de uma pessoa e depois a recuperação da mesma em caso de perda devido a ruído, ocultação ou outro factor. A pessoa pode ser extraída da frame em análise usando um método de segmentação. Uma vez identificada a pessoa podem ser computados um conjunto de características, e estas características podem ser utilizadas para atribuir a pessoa a uma classe (de pé, abaixada, deitada) (Fig. 10). O resultado da pesquisa automática pode ser positivo, se o alvo for novamente encontrado, ou negativo, em caso do alvo ainda não ter sido encontrado ou identificado. No primeiro caso, o seguimento activo irá ser realizado novamente na próxima frame, enquanto que no segundo caso há a necessidade de continuar a procura na mesma frame. É definida uma condição de paragem para esta pesquisa após um número de frames. Quando esta condição é atingida com nenhum resultado positivo, a decisão final e o controlo do sistema é dado ao utilizador avisando da perda do alvo a seguir.

Para se identificar os alvos segmentados usam-se um conjunto de características, que estão agrupadas em quatro categorias: morfologia,

geometria, intensidade e semântica. Para realizar a pesquisa são utilizadas técnicas de *Content-Based Image Retrieval* (CBIR).

Este método pode funcionar em tempo real, uma vez que as operações mais demoradas são realizadas na etapa *off-line*, na qual é construída a base de dados para o CBIR. Entre as várias variáveis que fazem depender o número de operações realizadas, encontra-se a dimensão do alvo a ser seguido, isto é, objectos maiores requerem maior esforço para serem segmentados e caracterizados.

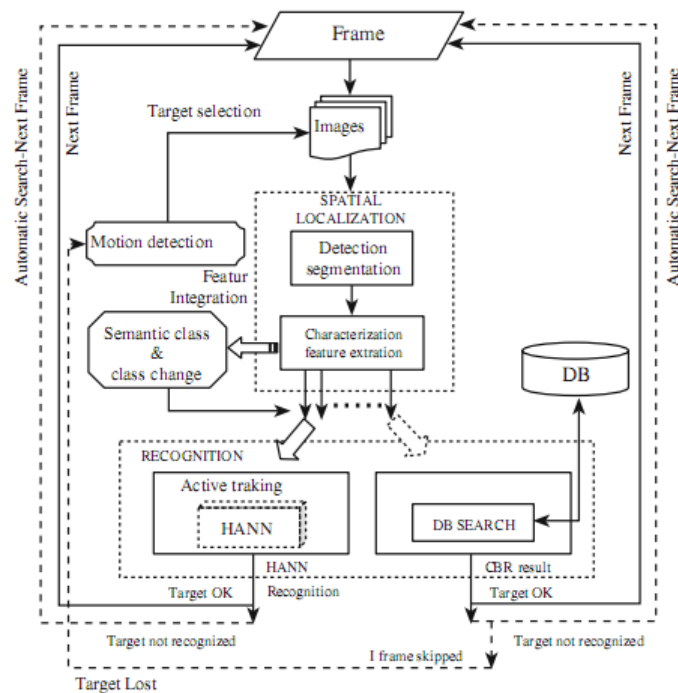


Fig. 10 - Representação gráfica do algoritmo do método utilizado em [15]

2.3.3 - Tracking usando Imagens Naturais/Térmicas



Fig. 11 - Exemplo de imagens naturais e térmicas para ligeiro nublado (topo), vento forte e chuva (meio) e neve (baixo) [16]

Como os algoritmos que utilizam imagens naturais têm uma grande sensibilidade às mudanças de iluminação e os que utilizam imagens térmicas muitas vezes têm uma resolução muito baixa, cada tipo de algoritmo é apropriado para determinados cenários e desapropriado para outros (Fig. 11). Para se tirar partido desta situação recorre-se à junção das imagens naturais com as imagens térmicas. Assim recorre-se às vantagens de cada uma em cada cenário e consegue-se obter um resultado melhor.

Um algoritmo deste tipo é apresentado em [17], que descreve um método de detecção *online* para duas *streams* sincronizadas, de vídeos de longo alcance, mas que estão sobre zooms diferentes.

Em [16] é combinada a informação das imagens naturais e térmicas ao nível do pixel, isto é, cada pixel é classificado como fundo ou *foreground* de acordo com a sua similaridade ao modelo de fundo natural-térmico. É aplicado um método de extracção de objectos de imagens naturais baseado em *codebook*. No *codebook* estão contidas *codewords* que contém um conjunto de informações sobre cada pixel numa determinada frame, o que permite ter um modelo dinâmico do fundo. Cada pixel é representado por uma *codeword* que contém os valores R, G, e B, a intensidade máxima e mínima, o número de correspondências, e o tempo da primeira e última

correspondência. Tem também um parâmetro *Maximum Negative Run-Length* (MNRL) que guarda o período de tempo em que a *codeword* não tem correspondência. Como os pixels do fundo são observados periodicamente, as *codewords* que correspondem a estes pixels tem um MNRL baixo enquanto que *codewords* geradas por uma presença temporária vão ter um MNRL alto. Pode-se por isso aplicar um *threshold* para se distinguir o fundo do *foreground*.

Para cada nova frame, cada pixel é associado à *codeword* mais similar baseado na sua cor e intensidade. Se não houver nenhuma *codeword* correspondente, então uma nova é criada e adicionada ao *codebook*. As *codewords* que não têm correspondência por um longo período de tempo são depois eliminadas.

Para integrar as imagens térmicas nesta solução é adicionado um parâmetro T às *codewords*, que corresponde à intensidade térmica de cada pixel.

Um pixel só é classificado como *foreground* se os seus valores de cor ou térmicos forem diferentes das *codewords* do fundo.

Utilizando um conjunto de imagens de teste (Fig. 12) para se avaliar os resultados deste método (Fig. 13), pode-se ver que as imagens naturais isoladamente apresentam melhores resultados nuns cenários e as imagens térmicas isoladas apresentam melhores resultados noutros cenários. Com a junção dos dois tipos é possível tirar partido das vantagens de cada tipo de imagem.



Fig. 12 - Imagens utilizadas para teste [16]

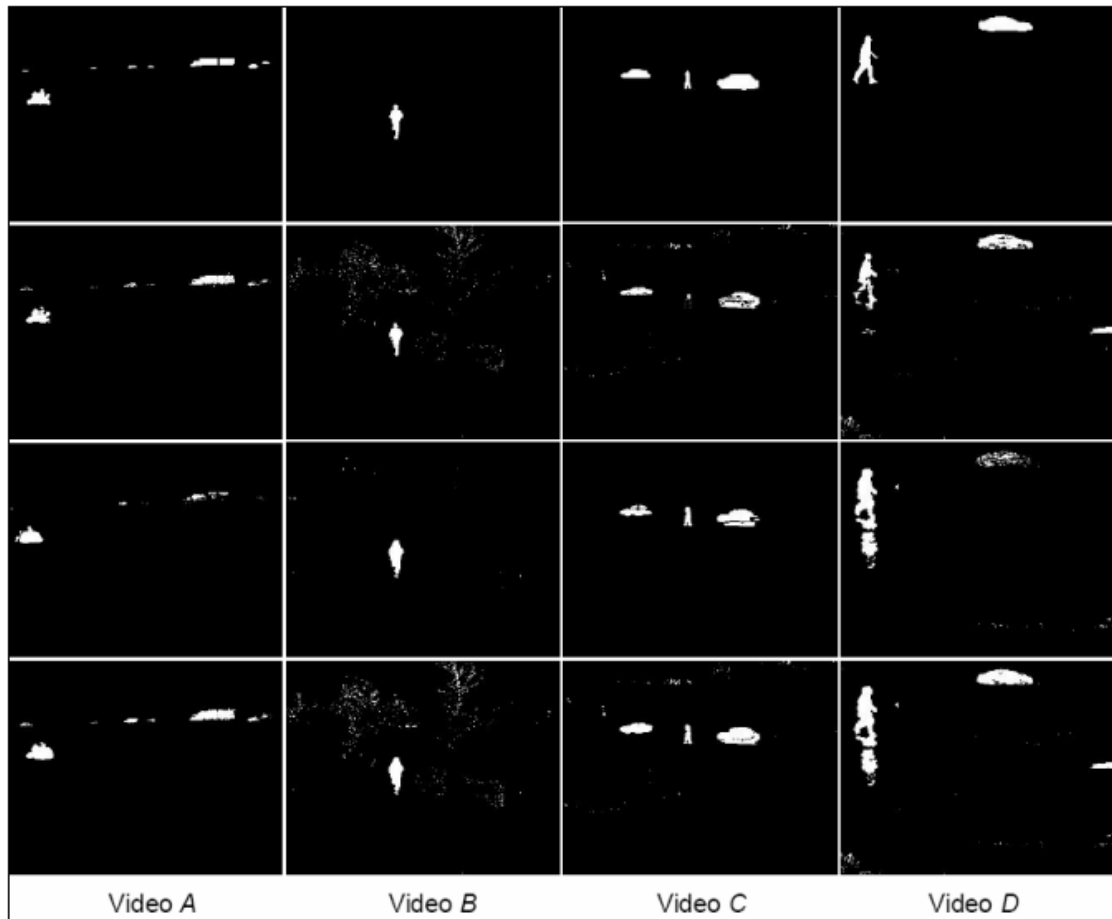


Fig. 13 - Resultados: Ground Truth (1ª linha), imagens naturais (2ª linha), imagens térmicas (3ª linha) e combinação das duas (4ª linha) [16]

Existem outros métodos baseados em *codewords* [18], utilizando no entanto formas diferentes de fazer o seguimento dos objectos detectados. É utilizado um modelo de Bayesian, que é uma das técnicas mais utilizadas para se fazer a estimação do movimento dos objectos [19]. De seguida é realizada a computação da probabilidade posterior que consiste na procura do estado candidato x' que, dado o último estado x conhecido, se ajusta melhor à observação corrente z e finalmente aplica um algoritmo chamado “*Jump-diffusion Dynamics*” que consiste no melhoramento da abordagem do filtro de partículas.

2.4 - Avaliação de algoritmos

A Framework de avaliação é uma ferramenta necessária, pois permite calcular de uma forma automatizada e objectiva a qualidade do algoritmo desenvolvido. Existem vários métodos para se calcular as métricas associadas

ao algoritmo. Em [20] o método utilizado envolve o uso de *Ground-Truth* (GT) para fazer a avaliação, propondo duas formas para se fazer a avaliação, uma baseada em frame e outra baseada no objecto. As métricas baseadas em frame são usadas para medir a performance do sistema em frames individuais do vídeo, onde cada frame é individualmente avaliado para ver se o número de objectos bem como o seu tamanho e localização correspondem à informação de GT desse frame particular.

Por outro lado as métricas baseadas no objecto têm em consideração toda a trajectória do objecto para testar a capacidade de seguimento do sistema, incluindo uma etiquetagem consistente do objecto. Aqui, o seguimento individual dos objectos, que são automaticamente detectados e depois seguidos durante o seu tempo de vida, são analisados como entidades diferentes.

É possível avaliar o algoritmo não tendo informação de GT, para isso utilizam-se as métricas *Nonground-Truth* (NGT). Em [21] estas métricas exploram características de cor e movimento na vizinhança do objecto segmentado. Uma característica é o contraste de cor espacial ao longo da fronteira de cada objecto. A segunda característica consiste nas diferenças do histograma de cor nos planos do objecto de vídeo. A terceira é baseada nas diferenças do vector de movimento ao longo da fronteira do objecto.

Uma alternativa baseada nas métricas *Partition Distance* (PD) é apresentada em [22]. Esta alternativa junta o uso de BB GT com *Reference Segmentation* (RS) minimizando a quantidade destas últimas, uma vez que são difíceis de obter. Designa-se assim por métrica híbrida.

Capítulo 3

Trabalho Desenvolvido

No decorrer do primeiro semestre foi iniciado algum trabalho no âmbito da dissertação a realizar no segundo semestre.

Foi identificado e desenvolvido o método que avalia a qualidade do algoritmo a desenvolver. Este método é muito importante uma vez que é através dele que se irá averiguar a qualidade da solução desenvolvida ao longo do semestre. Para fazer a integração das várias métricas foi desenvolvida uma framework (Fig. 14) que se irá encarregar de ler os ficheiros de configuração para as várias sequências que contém os parâmetros necessários para fazer a avaliação, e posteriormente irá aplicar as métricas que são possíveis, mediante os parâmetros de entrada fornecidos. Os módulos implementados foram o *launcher* que tem como funcionalidades, a leitura dos ficheiros de configuração das sequências que vão ser avaliadas, e a extracção dos parâmetros necessários para ser feita a avaliação e a parte do *Evaluation Manager* que faz a decisão de qual métrica executar, mediante os parâmetros de entrada que forem fornecidos. Tanto o launcher como o evaluation manager foram implementados em Python, no entanto umas métricas encontram-se em Python e outras em C++. Foi por isso necessário realizar uma ligação entre as duas linguagens para integrar todos os módulos da framework.

Para o acompanhamento da dissertação e disponibilização de informação sobre a mesma, encontra-se em desenvolvimento um *site* que irá ser actualizado ao longo do curso do projecto.

O desenvolvimento do projecto que terá por base a construção do algoritmo referido terá lugar no segundo semestre, e após a sua

implementação seguir-se-á a análise dos resultados e respectivas conclusões, bem como um olhar sobre perspectivas futuras.

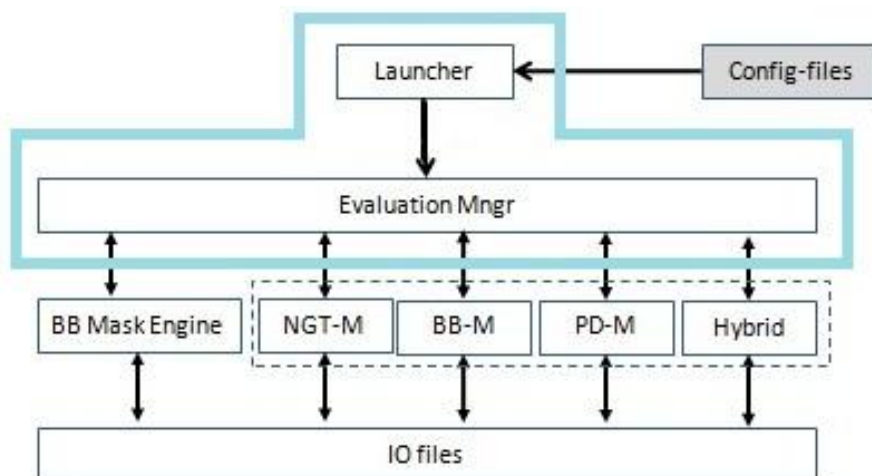


Fig. 14 - Esquema da framework de avaliação

Capítulo 4

Conclusões

Com este trabalho foi possível obter um conhecimento sobre as tecnologias existentes para usar em detecção e seguimento de imagens de vídeo.

Do que foi estudado, conclui-se que existem vários métodos que podem ser utilizados, mas que nenhum deles deve ser considerado como geral, já que cada um é aconselhado para determinadas condições de imagem ou cenários, e desaconselhado para outros.

Depois da análise aos vários métodos, e tendo em conta que o trabalho será desenvolvido em condições adversas, nomeadamente em condições de iluminação desfavoráveis, prevê-se que as imagens térmicas apresentem melhores resultados do que as imagens naturais, nomeadamente na detecção de pessoas quando estas condições estiverem presentes. No seguimento de pessoas com imagens térmicas torna-se difícil fazer a distinção de cada pessoa em certos cenários pois todas as pessoas têm uma descrição idêntica.

No decorrer do trabalho será também avaliado se a combinação de imagens térmicas com imagens naturais trará vantagens ao projecto.

Quanto à avaliação de algoritmos irão ser realizados testes com sequências de vídeo já existentes, para se aferir se a mesma é realizada correctamente e se os resultados estão de acordo com o esperado.

Referências

1. Zhihua, X., et al. *Infrared Face Recognition Based on Radiant Energy and Curvelet Transformation*. in *Information Assurance and Security, 2009. IAS '09. Fifth International Conference on*. 2009.
2. Wren, C.R., et al., *Pfinder: real-time tracking of the human body*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997. **19**(Compendex): p. 780-785.
3. Omar, M. and Z. Yi. *Pedestrian tracking routine for passive automotive night vision systems*. in *Sensor Review*. 2007. UK: Emerald.
4. Zhao, T., R. Nevatia, and B. Wu. *Segmentation and tracking of multiple humans in crowded environments*. in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2008. 445 Hoes Lane - P.O.Box 1331, Piscataway, NJ 08855-1331, United States: Inst. of Elec. and Elec. Eng. Computer Society.
5. Torresan, H., et al. *Advanced surveillance systems: Combining video and thermal imagery for pedestrian detection*. in *Thermosense XXVI, April 13, 2004 - April 15, 2004*. 2004. Orlando, FL, United states: SPIE.
6. Goubet, E., J. Katz, and F. Porikli. *Pedestrian tracking using thermal infrared imaging*. in *Infrared Technology and Applications XXXII, April 17, 2006 - April 21, 2006*. 2006. Kissimmee, FL, United states: SPIE.
7. Haritaoglu, I., D. Harwood, and L.S. Davis. *W4: Real-time surveillance of people and their activities*. in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2000. Los Alamitos, CA, United States: IEEE.
8. Zhao, T. and R. Nevatia. *Tracking multiple humans in complex situations*. in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2004: Institute of Electrical and Electronics Engineers Computer Society.
9. Wong, W.K., et al. *An effective surveillance system using thermal camera*. in *2009 International Conference on Signal Acquisition and Processing, ICSAP 2009, April 3, 2009 - April 5, 2009*. 2009. Kuala Lumpur, Malaysia: IEEE Computer Society.
10. Li, J. and W. Gong. *Real time pedestrian tracking using thermal infrared imagery*. in *Journal of Computers*. 2010. P.O.Box 40, FIN-90571, OULU, 90571, Finland: Academy Publisher.

11. Padole, C.N. and L.A. Alexandre. *Wigner distribution based motion tracking of human beings using thermal imaging*. in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, CVPRW 2010, June 13, 2010 - June 18, 2010*. 2010. San Francisco, CA, United states: IEEE Computer Society.
12. Treptow, A., G. Cielniak, and T. Duckett. *Real-time people tracking for mobile robots using thermal vision*. in *Robotics and Autonomous Systems*. 2006: Elsevier.
13. Walczyk, R., A. Armitage, and T.D. Binnie. *An embedded real-time pedestrian detection system using an infrared camera*. in *Signals and Systems Conference (ISSC 2009), IET Irish*. 2009.
14. Fernández-Caballero, A., et al. *Real-time human segmentation in infrared videos*. in *Expert Systems with Applications*. 2011.
15. Moroni, D. and G. Pieri. *Object tracking in video-surveillance*. in *Pattern Recognition and Image Analysis*. 2009: MAIK Nauka/Interperiodica distributed exclusively by Springer Science+Business Media LLC.
16. St-Laurent, L., X. Maldague, and D. Prevost. *Combination of colour and thermal sensors for enhanced object detection*. in *FUSION 2007 - 2007 10th International Conference on Information Fusion, July 9, 2007 - July 12, 2007*. 2007. Quebec, QC, Canada: Inst. of Elec. and Elec. Eng. Computer Society.
17. Torabi, A., G. Masse, and G.-A. Bilodeau. *Feedback scheme for thermal-visible video registration, sensor fusion, and people tracking*. in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, CVPRW 2010, June 13, 2010 - June 18, 2010*. 2010. San Francisco, CA, United states: IEEE Computer Society.
18. Leykin, A. and R. Hammoud. *Robust multi-pedestrian tracking in thermal-visible surveillance videos*. in *2006 Conference on Computer Vision and Pattern Recognition Workshops, June 17, 2006 - June 22, 2006*. 2006. New York, NY, United states: Institute of Electrical and Electronics Engineers Computer Society.
19. Dore, A., M. Soto, and C.S. Regazzoni, *Bayesian Tracking for Video Analytics*. *Signal Processing Magazine, IEEE*, 2010. 27(5): p. 46-55.
20. Bashir, F. and F. Porikli. *Performance Evaluation of Object Detection and Tracking Systems*. 2006.
21. Erdem, C.E., B. Sankur, and A.M. Tekalp. *Performance measures for video object segmentation and tracking*. in *Visual Communications and Image Processing 2003, July 8, 2003 - July 11, 2003*. 2003. Lugano, Switzerland: SPIE.
22. Carvalho, P., J.S. Cardoso, and L. Corte-Real. *Hybrid framework for evaluating video object tracking algorithms*. in *Electronics Letters*. 2010. Six Hills Way, Stevenage, SG1 2AY, United Kingdom: Institution of Engineering and Technology.