



INSTITUTO SUPERIOR DE ENGENHARIA DE LISBOA

Área Departamental de Engenharia de Eletrónica e Telecomunicações e de Computadores

Seguimento de Pessoas em Sistemas Multi-câmara

JOÃO ALBERTO MARTINS DIAS

(Bacharel)

Trabalho Final de Mestrado para obtenção do grau de Mestre em Engenharia de
Redes de Comunicação e Multimédia

Orientador:

Doutor Pedro Miguel Torres Mendes Jorge

Júri:

Presidente: Doutor Arnaldo Joaquim de Castro Abrantes

Vogal:

Doutor Jacinto Carlos Marques Peixoto do Nascimento

Dezembro de 2014



Resumo

Esta dissertação aborda o estudo, análise e implementação de sistemas de seguimento coerente de pessoas, em ambientes de múltiplas câmaras. Pretende-se que o sistema detete e siga pessoas nas imagens adquiridas por cada câmara de vigilância.

Posteriormente, estas trajetórias serão correlacionadas para permitir a determinação da trajetória total, que uma pessoa percorre enquanto permanecer na zona vigiada. De salientar o facto de que as áreas cobertas pelas várias câmaras, podem ser sobrepostas ou sem área comum.

O sistema relaciona os campos de visão das câmaras com o mapa da área monitorizada, para que as trajetórias sejam representadas num sistema de coordenadas comum a todas a câmaras (sistema de coordenadas do mundo).

As tarefas desenvolvidas nesta tese envolvem a calibração do sistema de múltiplas câmaras, a deteção e o seguimento de objetos, a correspondência das várias trajetórias locais detetadas em cada câmara e a representação destas trajetórias numa planta formada pelas várias vistas das câmaras.





Abstract

This thesis approaches the study, analysis and implementation of systems for tracking people in coherent systems of multiple cameras. It is intended that the system detect and follow people in the images acquired by each surveillance camera, these being fixed.

Later, these trajectories will be correlated to allow the determination of the total trajectory, a person travels while remaining in the surveillance zone. To emphasize the fact that, the areas covered by multiple cameras, can be no overlapping or common area.

The system relates the fields of view of the cameras with the map of the monitored area, so that these trajectories can be represented in a coordinate system common to all cameras (the world coordinate system).

The tasks developed in this thesis involve the calibration of multiple cameras system, the detection and tracking of objects and matching the various local trajectories detected in each camera and the representation of these trajectories in a plant formed by the various views of the cameras.





Agradecimentos

Em primeiro lugar gostaria de deixar um enorme agradecimento ao Doutor Pedro Miguel Torres Mendes Jorge, que me orientou neste desafio e que me incutiu perseverança e incentivo, as quais foram essenciais para a realização deste trabalho.

Um abraço de agradecimento aos meus colegas Carlos Nobre, Nuno Pinto e Sérgio Henriques, com quem partilhei horas de trabalho no mestrado.

Quero agradecer à Elsa e à Ana, toda a paciência que tiveram, pois eram as primeiras pessoas com quem eu desabafava.

Dedico este trabalho à Ana e à memória do meu Pai.





Índice

Lista de figuras.....	ix
Introdução.....	1
1.1 Motivação.....	1
1.2 Objetivos.....	1
1.3 Organização da tese.....	2
1.4 Contributos da tese.....	3
Estado da Arte.....	4
2.1 Introdução.....	4
2.2 Evolução dos sistemas.....	4
2.3 Topologias dos sistemas.....	5
2.4 Trabalho relacionado.....	5
2.4.1 Calibração de câmaras.....	6
2.4.2 Técnicas de deteção de movimento.....	7
2.4.3 Seguimento.....	8
2.4.4 Ligação de trajetórias.....	9
Considerações Teóricas.....	10
3.1 Calibração.....	10
3.1.1 Projeção central usando coordenada homogéneas.....	12
3.1.2 Rotação e translação da câmara.....	14
3.1.3 Determinação da matriz de projeção.....	15
3.1.4 Distorção das lentes.....	15
3.1.5 Homografias.....	16
3.2 Correspondência de blobs.....	17
3.2.1 Construção de trajetórias.....	20
Método Proposto.....	25
4.1 Introdução.....	25
4.2 Determinação da área a vigiar.....	26
4.3 Calibração de Câmaras.....	26



4.3.1	Calibração de câmara isolada	27
4.3.1.1	Homografia imagem mundo	30
4.3.1.2	Construção da imagem virtual de topo de uma câmara	30
4.3.2	Calibração de várias câmaras	32
4.3.2.1	Calibração com objeto calibrador comum	33
4.3.2.2	Calibração com objeto calibrador individuais	33
4.3.3	Construção da planta final	35
4.4	Deteção de trajetórias	38
4.4.1	Deteção de <i>blobs</i>	38
4.4.1.1	Segmentação da imagem	39
4.4.2	Correspondência de <i>blobs</i>	42
4.5	Relações entre trajetórias	44
4.5.1	Ligações de trajetórias numa câmara	46
4.5.2	Correspondências entre trajetórias entre câmaras	48
4.5.3	Ligação de trajetórias entre várias câmaras	52
4.6	Geração do grafo	53
4.6.1	Nós	53
4.6.2	Ligações	53
	Resultados	56
5.1	Tempos de execução	56
5.2	<i>Ground Truth</i>	58
5.3	Análise de resultados	59
	Conclusões e Trabalho Futuro	62
6.1	Conclusões	62
6.2	Trabalho Futuro	63
	Bibliografia	64
	Anexos	67
8.1	Grafos das trajetórias	67
8.2	Diagramas <i>UML</i>	78
8.2.1	Cálculo dos parâmetros intrínsecos	78
8.2.2	Cálculo da relação entre referenciais	79



Lista de figuras.

Figura 1: Geometria de uma câmara “ <i>pinhole</i> ” [28]	10
Figura 2: Projeção em X.....	11
Figura 3: Projeção em Y	11
Figura 4: Referencial da imagem.....	13
Figura 5: Transformação euclidiana entre coordenadas (figura retirada de [28]).	14
Figura 6: Homografia [28].....	16
Figura 7: Regiões de duas imagens consecutivas	17
Figura 8: <i>Blobs</i> de 2 imagens consecutivas	18
Figura 9: Trajetórias associadas	20
Figura 10: Nova trajetória.....	21
Figura 11: Troço de trajetória de dois <i>blobs</i> consecutivos	21
Figura 12: Criação de trajetória devido a <i>Merge</i>	21
Figura 13: Criação de trajetória devido a <i>Split</i>	22
Figura 14: Ocorrência de um <i>Leave</i>	22
Figura 15: Grafo que representa as trajetórias da Figura 7.....	23
Figura 16: Correspondência de <i>blobs</i> por intersecção de <i>bounding boxes</i>	24
Figura 17: Diagrama de blocos do método proposto.....	25
Figura 18: Áreas a vigiar	26
Figura 19: Objeto calibrador de precisão	27
Figura 20: Obtenções dos pontos dos cantos interiores do padrão.....	28
Figura 21: Imagem corrigida	28
Figura 22: Obtenção dos parâmetros extrínsecos	29
Figura 23: Criação de imagem topo virtual	31
Figura 24: Imagens das câmaras e respetivas imagens de vista virtual de topo.	32
Figura 25: Calibração com objeto calibrador comum	33
Figura 26: Calibração com objetos calibradores individuais.....	34
Figura 27: Relacionamento de referenciais	34
Figura 28: Relacionamento de referenciais	35
Figura 29: Obtenção da homografia $H_{\text{mundo} \rightarrow \text{topo}}$	36
Figura 30: Imagens das vistas de topo.....	37
Figura 31: Diagrama de blocos de deteção de <i>blobs</i>	38
Figura 32: Diagrama de blocos da segmentação	40
Figura 33: Imagens de fundo	41
Figura 34: imagens de uma sequência.....	41
Figura 35: Imagens segmentadas.....	41
Figura 36: Deteção de trajetórias.....	42
Figura 37: Sequência de imagens de <i>merges</i>	43



Figura 38: Sequência de imagens de <i>splits</i>	43
Figura 39: Representação de trajetórias e ligações merge/split.....	43
Figura 40: Relações entre trajetórias	44
Figura 41: Ligação de trajetórias numa câmara.....	46
Figura 42: Ligação em uma câmara	47
Figura 43: Representação gráfica de ligações de trajetórias numa câmara	47
Figura 44: Trajetórias nos instantes de tempo 5594	48
Figura 45: Correspondência de trajetórias	48
Figura 46: Correspondência de trajetórias (representação gráfica)	49
Figura 47: Correspondência de trajetórias	51
Figura 48: Representação gráfica de correspondência de trajetórias	51
Figura 49: Ligação em câmaras diferentes	52
Figura 50: Representação gráfica de ligações de trajetórias em câmaras diferentes.....	52
Figura 51: Trajetórias e relações	54
Figura 52: Trajetórias e relações com informação extra	55
Figura 53: Falha de detecção de <i>Split</i>	59
Figura 54: Falso <i>New</i>	60
Figura 55: Falha de detecção de relações entre câmaras.....	60
Figura 56: Falsa ligação numa câmara	61



Introdução

1.1 Motivação

Devido ao grande avanço tecnológico e ao contínuo decréscimo do preço dos equipamentos, é possível a instalação de sistemas de câmaras com alguma dimensão, sendo estes sistemas de vigilância cada vez mais utilizados para monitorização de zonas sensíveis, como por exemplo aeroportos, centros comerciais ou edifícios públicos. No entanto, devido ao grande número de câmaras colocadas, não é possível a sua constante monitorização por elementos da segurança.

Uma solução passa por dotar estes sistemas de algoritmos que realizam automaticamente determinadas tarefas, como por exemplo, deteção de objetos abandonados ou alerta de acesso a áreas interditas. Para a realização de grande parte destas tarefas é necessário o seguimento coerente de pessoas ou objetos ao longo de toda a zona vigiada. Este seguimento envolve a deteção e o reconhecimento de regiões ativas nos vários campos de visão das câmaras que formam o sistema de vigilância e posterior determinação da trajetória global.

Esta tarefa é complexa devido a um conjunto de problemas que dificultam a determinação das trajetórias globais, como por exemplo, devido a alterações da cena (luminosidade ou objetos com movimento esporádicos) ou alterações na aparência das pessoas provocadas por alteração de vista ou oclusões parciais.

1.2 Objetivos

Pretende-se nesta dissertação, desenvolver um sistema de videovigilância para o seguimento de pessoas numa determinada zona a vigiar, com recurso a várias câmaras fixas.



O objetivo principal será seguir as pessoas que evoluem no campo de visão das câmaras, operação que envolve a determinação da trajetória de pessoas isoladas e em grupo, e a extração de características visuais (cor, forma, textura, movimento, imagens de referência). Cada câmara tem a sua área de cobertura, podendo existir câmaras com áreas de cobertura sobrepostas e câmaras que não tenham sobreposição dos respetivos campos de visão.

No seguimento de pessoas, o maior desafio reside nas oclusões, podendo-se distinguir oclusões provocadas por elementos estáticos (pertencentes ao cenário), não sendo a pessoa detetada e oclusões devido à passagem da pessoa de um campo de visão de uma câmara para outra, não estando visível em nenhuma câmara.

A oclusão por interação de pessoas num grupo causa outro tipo de ambiguidades. É possível detetar o grupo mas torna-se difícil distinguir cada um dos seus elementos.

O seguimento é formulado com um processo de deteção e etiquetagem das regiões ativas em cada imagem da sequência de vídeo. Num primeiro passo são detetados segmentos de trajetória por associação das regiões ativas em imagem consecutivas. Este passo do algoritmo transforma as imagens da sequência de vídeo, num conjunto de segmentos de trajetória com as respetivas características extraídas do sinal (por exemplo, centro de massa, posição no mundo, cores dominantes ou velocidade médias).

Após a deteção de todos os segmentos de trajetória, o passo seguinte é verificar as relações existentes entre estes segmentos, permitindo a ligação de segmentos de trajetória que pertencem à mesma pessoa, detetadas na mesma câmara ou em diferentes câmaras.

1.3 Organização da tese

No capítulo 2 faz-se uma abordagem sobre os artigos científicos publicados e o que já foi desenvolvido sobre as várias fases deste trabalho, seguido de uma introdução teórica sobre cada tema, realizada no capítulo 3.

De seguida, no capítulo 4 apresenta-se o método desenvolvido.

Os resultados do sistema são apresentados no capítulo 5 e finalizando apresentam-se as conclusões no capítulo 6, onde também são dadas sugestões de trabalho futuro.



1.4 Contributos da tese

As contribuições presentes nesta tese são descritas no capítulo 4, sendo de destacar as seguintes:

- calibração de câmaras com auxílio de padrão calibrador;
- geração de vistas virtuais de topo de câmaras;
- criação de uma planta global das várias vistas virtuais de topo;
- relacionamento dos referenciais dos vários objetos calibradores através do uso de uma câmara auxiliar sendo esta retirada no final da calibração. Este contributo original permite, obter sistemas calibrados com câmaras sem vistas sobrepostas calibradas para o mesmo referencial;
- deteção de segmentos de trajetória através da análise de sequências de imagens;
- deteção de criação e separação de grupo;
- modelo para lidar com oclusões de pessoas numa câmara e entre câmaras,
- criação de um modelo para registo das relações de trajetórias;
- criação de grafos para visualização do relacionamento das trajetórias.



Estado da Arte

2.1 Introdução

Este capítulo apresenta um conjunto de trabalhos sobre a temática da videovigilância em vários domínios, abordando as técnicas desenvolvidas, tais como calibração de câmaras, segmentação ou seguimento de pessoas.

2.2 Evolução dos sistemas

Quando apareceram os primeiros sistemas de vigilância (anos 70), estes eram constituídos por câmaras analógicas colocadas em pontos fulcrais. Os sinais de vídeo eram transmitidos via rádio ou cabo até um a sala de controlo onde as imagens das câmaras podiam ser visualizadas, sendo igualmente armazenadas em vídeo gravadores analógicos. Para se poder ter resposta em tempo real de eventos estranhos, era necessário enormes recursos humanos. Nos casos de ocorrência de eventos, e se não houver informações sobre o instante de tempo onde este ocorreu, (por exemplo assalto noturno), obrigava a uma fastidiosa pesquisa nos registos magnéticos.

Nos finais da década de 80, os sistemas começaram a ser melhorados, com o aparecimento da compressão de vídeo, onde os sinais analógicos das câmaras eram processados localmente em placas de compressão de vídeo, sendo o sinal transmitido digitalmente. Esta digitalização permitiu o surgimento de processos rudimentares de deteção automática de eventos.

Com o desenvolvimento das redes de comunicação e o aparecimento de câmaras *IP*¹, é possível ter sistemas de vigilância com grande qualidade e em tempo real.

Atualmente existe um grande número de áreas de interesse em que os sistemas de videovigilância são aplicados. Existem trabalhos publicados em que propõem sistemas

¹ IP Internet Protocol



para monitorização de áreas florestais [1], controlo de aeroportos [2], controlo rodoviário [3] ou controlo das condições climáticas [4], entre outros.

2.3 Topologias dos sistemas

Os sistemas podem ser montados em diversos ambientes e com várias topologias.

Os sistemas de videovigilância podem ser utilizados para controlo de zonas interiores [5] [6] como por exemplo, centro comercial, galerias de arte, agências bancárias, ou para zonas exteriores [3] [4] como por exemplo, estádio, caminhos públicos, controlo de tráfego rodoviário.

Nos sistemas que monitorizam as zonas exteriores, existem um conjunto de variáveis não controláveis que obrigam a métodos mais exigentes na extração de informação. Um exemplo é a maior complexidade do sistema na obtenção de regiões ativas devido a mudanças de luminosidade, chuva intensa ou objetos pertencentes à imagem de fundo a moverem-se, como por exemplo os ramos das árvores ou sombras provocadas por exposição de fontes de luz.

Estes sistemas podem ser montados numa tipologia simples, estéreo ou múltiplas câmaras. Os sistemas simples são aqueles que usam somente uma câmara para vigiar uma determinada área. Estes sistemas são os mais simples de implementar e acarretam um menor peso computacional.

Os sistemas estéreo são compostos por duas ou mais câmaras a vigiar em simultâneo uma determinada área, colocadas em posições ligeiramente afastadas. Este sistema permite extrair informação tridimensional da área vigiada.

Nos sistemas de múltiplas câmaras temos várias câmaras a vigiar uma determinada área podendo os seus campos de visão estarem ou não sobrepostos.

2.4 Trabalho relacionado

A maioria dos trabalhos e artigos científicos não propõem métodos inovadores em todas as fases de desenvolvimento dum projeto de videovigilância, ficando estes focados somente numa determinada área.

Existem muitas áreas relacionadas com a videovigilância, como por exemplo a calibração, deteção de movimentos e regiões ativas, extração de características.



Com a obtenção das imagens das câmaras é necessário extrair informação, como por exemplo a detecção de regiões ativas, às quais são associadas um conjunto de características, como por exemplo, a sua posição, histogramas de cor ou velocidade.

Terá de existir um bloco que faça a correspondência entre essas regiões ativas detetadas em imagens consecutivas para obtenção dos segmentos de trajetória.

Através dos valores da calibração das câmaras, da informação extraída das regiões ativas, consegue-se efetuar a ligação dos segmentos de trajetórias que representam os mesmos objetos/pessoas.

O sistema deverá estar preparado para se conseguir extrair as trajetórias mais prováveis de cada pessoa.

De seguida, optou-se por realizar um estudo do trabalho relacionado em cada área descrita anteriormente, por forma a cobrir todas as temáticas de um sistema de videovigilância, como proposto nesta dissertação.

2.4.1 Calibração de câmaras

A calibração das câmaras é uma tarefa, que exige precisão na sua realização, para se poder obter resultados satisfatórios na obtenção das trajetórias das pessoas, em sistemas multi-câmara, podendo assim fazer-se o seguimento coerente em toda a área vigiada.

Hall [7], utiliza um conjunto de pontos não coplanares, baseando-se em projeção de perspetiva e no modelo ideal para a câmara (*pinhole*).

A mesma técnica baseada em projeção de perspetiva e pontos coplanares é utilizada por Tsai [8], mas agora tem em conta a distorção radial da lente.

Em [9] [10], são propostos algoritmos para calibração da câmara estéreo que se baseiam no método de Tsai [8] e Zhang [11], os quais usam um conjunto de pontos coplanares para obtenção das equações.

Existem outras abordagens em que a calibração é conseguida usando pontos de fuga [12] [13], em algumas das quais se tenta fazer a calibração de forma automática, sem nenhum elemento calibrador. Kusakunniran [13] apresentou um trabalho em que propunha a calibração através da observação do caminhar de pessoas, na qual se obtém a matriz principal através da obtenção do ponto de fuga vertical e da linha do horizonte.



Ruishuang [10], explora as funcionalidades de funções da biblioteca do *openCV*[14], mostrando que a partir de um objeto calibrador com determinadas características, se calibra uma câmara. São estimados os parâmetros intrínsecos e extrínsecos da câmara baseada no algoritmo de Zhang[11] e na ferramenta de calibração *BouguetMCT* [15]. Este processo necessita de um vetor de pontos 3D e as correspondentes projeções 2D. Consegue-se determinar estes pontos usando um objeto com uma geometria conhecida e facilmente detetável.

Behera e Kharade, [16] propuseram um sistema com multi-câmara com sobreposição dos campos de visão, relacionando as câmaras através de homografias, detetando para isso pontos referência partilhados por ambas as câmaras.

Takeshi[17] utilizou um robot móvel, com deteção de posicionamento no mundo, através de três pontos luminosos. O robot móvel desloca-se ao longo do campo de visão das várias câmaras, permitindo que o sistema faça a aquisição de imagens e relacione os pontos luminosos adquiridos nas imagens com as posições no mundo.

2.4.2 Técnicas de deteção de movimento

A deteção de movimento e a extração de zonas de interesse, também designadas por regiões ativas, é uma área que depende fortemente do meio onde está inserido. Assim, uma aplicação pode funcionar bem em determinados ambientes mas ser inapropriado para outros.

Um grupo de algoritmos baseia-se no método de gradiente temporal [18] [19]. Contudo estes métodos apresentam o problema de não detetar objetos temporariamente estacionários.

Outro grupo de algoritmos baseia-se na subtração de modelo de fundo em que a imagem corrente é subtraída a uma imagem que representa os elementos estáticos da cena (imagem de fundo) [16] .

Existem várias técnicas para atualização de fundo, essencialmente para compensar alterações de luminosidade no decorrer da cena.

Kharade [16] e Wang [20] utilizam a subtração de fundo com um modelo de mistura de k-gaussianas[21]. Cada um dos pixéis do modelo de fundo, é constituído por uma mistura de gaussianas que representam os valores de intensidade que mais ocorreram.



Um pixel é classificado como ativo, se o seu valor não for bem representado por este modelo.

Shengsheng [18] desenvolveu um método de detecção de movimento baseado na subtração dos contornos presentes na imagem corrente e na imagem de fundo.

2.4.3 Seguimento

Em cada imagem da sequência de vídeo, após a detecção de movimento e binarização da mesma, obtêm-se as regiões ativas.

O passo seguinte será a sua etiquetação, permitindo a extração de informação relevante para cada região, como por exemplo, área, centro de massa ou histograma de cor.

A região ativa conexa é usualmente designada por *blob*². A associação entre os *blobs* detetados em imagens consecutivas permite construir as trajetórias dos objetos e consequentemente fazer o seu seguimento.

A associação entre blobs envolve alguma incerteza, pois os objetos podem não manter uma forma rígida ao longo da sua trajetória, por exemplo, uma pessoa ao caminhar, devido ao movimento das pernas e dos braços, altera significativamente a forma da região ativa.

Existem métodos que predizem o estado de um objeto num instante t (por exemplo o centro de massa), tendo como base as observações até ao instante $t-1$. Os mais utilizados são o filtro de Kalman[22] e o filtro de partículas [23] [24].

No trabalho proposto por Shengsheng [18], após serem extraídas as regiões ativas, as pessoas são incorporados em *bounding box*³ sendo utilizando um método de correspondência das *bounding box* da imagem atual e da imagem anterior para extração de movimento.

Kumar [19], apresentou uma técnica que segue um alvo com uma câmara *PTZ*⁴, em que o alvo é representado por um histograma de cor, sendo associado igualmente o histograma da sua vizinhança pertencente à imagem de fundo.

² *BLOb* Binary Large Object

³ *Bounding box* Retângulo envolvente

⁴ *PTZ* Pan Tilt Zoom



Fuentes [25], apresenta um método simples que relaciona bidireccionalmente os *blobs* de duas imagens consecutivas através de matrizes de correspondência.

2.4.4 Ligação de trajetórias

No seguimento a ocorrência de oclusões faz com que as trajetórias sejam interrompidas, por isso as ligações de trajetórias é um bloco essencial para a construção completa da trajetória, pois permite ligar os segmentos de trajetória que lhe são correspondentes.

Walter[26], apresentou um método que permite fazer a correspondência de trajetórias de câmaras diferentes em que os vídeos não estão sincronizados. A reconstrução é obtida a partir de dados de movimento, através da comparação dos segmentos das trajetórias e estimação dos alinhamentos espaciais e temporais.

Pedro Jorge[27] propôs usar uma rede bayesiana para fazer o seguimento. As redes bayesianas permitem modelar a interação das trajetórias detetada usando heurísticas simples.



Considerações Teóricas

3.1 Calibração

Define-se como calibração de uma câmara o processo realizado para calcular o valor dos parâmetros intrínsecos, extrínsecos e os coeficientes de distorção da lente, que são usados na modelação de câmaras reais.

Considera-se calibração de múltiplas câmaras o processo que permite a calibração de várias câmaras e da determinação da relação entre estas.

A calibração de câmaras permite determinar os seus parâmetros intrínsecos e extrínsecos, representados através da matriz de projecção.

Através da calibração são determinados os parâmetros do modelo matemático que permite transformar pontos de coordenadas 3D do referencial da câmara em pontos de coordenadas 2D no plano da imagem.

Sistema de coordenadas no referencial da câmara

$$P \leftarrow (X^c, Y^c, Z^c)^T \qquad p \leftarrow (x^c, y^c)^T$$

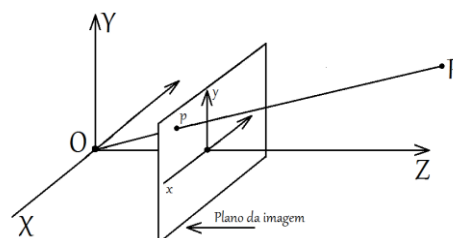


Figura 1: Geometria de uma câmara “pinhole” [28]



A Figura 1 mostra a geometria de uma câmara com foco finito (*pinhole*) em que um ponto P de coordenadas $(X^c, Y^c, Z^c)^T$, em relação ao referencial (X, Y, Z) centrado no centro óptico O , é projetado no ponto p de coordenadas $(x^c, y^c)^T$ em relação ao referencial (x, y) no plano da imagem.

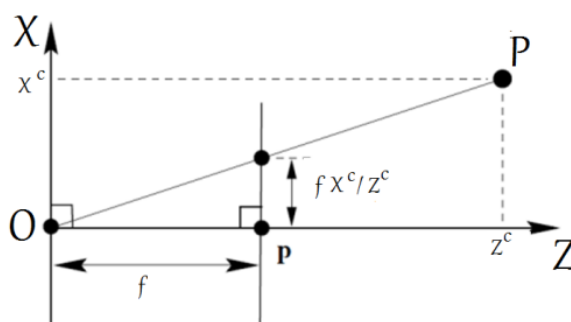


Figura 2: Projeção em X

$$x^c = f \frac{X^c}{Z^c}$$

Equação 1

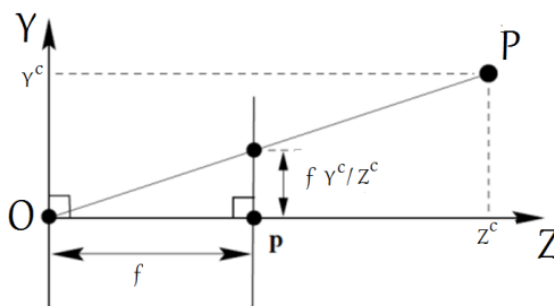


Figura 3: Projeção em Y

$$y^c = f \frac{Y^c}{Z^c}$$

Equação 2



As coordenadas dos pontos anteriores relacionam-se através das equações 1 e 2, e como mostram as figuras 2 e 3.

3.1.1 Projecção central usando coordenada homogéneas

As equações da projecção de perspectiva (equações 1 e 2) são não lineares quando expressas em coordenadas cartesianas, mas se representarmos as equações em coordenadas homogéneas, as equações passam a ser lineares.

Usando as propriedades das coordenadas homogéneas, dois vetores $(x_1, \dots, x_n, x_{n+1})^T$ e $(\lambda x_1, \dots, \lambda x_n, \lambda x_{n+1})^T$ representam o mesmo ponto para qualquer $\lambda \neq 0$, e um vetor com coordenadas homogéneas $(x_1, \dots, x_n, x_{n+1})^T$ representa o ponto com coordenadas cartesianas $(\frac{x_1}{x_{n+1}}, \dots, \frac{x_n}{x_{n+1}})^T$.

Se representarmos os pontos do mundo P e da imagem p através de vetores homogéneos, a projecção é expressa como um mapeamento linear entre as suas coordenadas homogéneas. Deste modo as equações 1 e 2 podem ser expressas do seguinte modo:

$$\begin{bmatrix} \lambda x^c \\ \lambda y^c \\ \lambda \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X^c \\ Y^c \\ Z^c \\ 1 \end{bmatrix} \quad \text{Equação 3}$$

No entanto, o ponto p no plano da imagem é normalmente expresso em píxeis sendo necessário a sua transformação para o referencial da imagem (u, v) , como ilustra a Figura 4.



Sistema de coordenadas no referencial da imagem

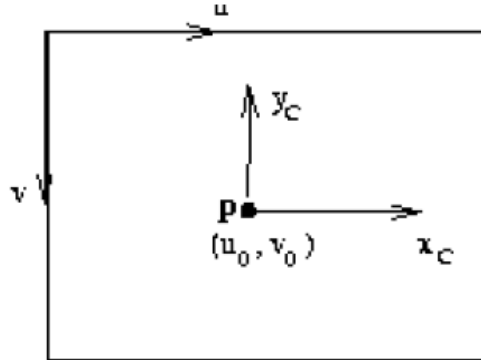


Figura 4: Referencial da imagem

$$\begin{aligned} k_u x^c &= u - u_0 \\ k_v y^c &= v_0 - v \end{aligned}$$

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \lambda u \\ \lambda v \\ \lambda \end{bmatrix} = \begin{bmatrix} k_u & 0 & u_0 \\ 0 & -k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \lambda x^c \\ \lambda y^c \\ \lambda \end{bmatrix} \quad \text{Equação 4}$$

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f k_u & 0 & u_0 & 0 \\ 0 & -f k_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X^c \\ Y^c \\ Z^c \\ 1 \end{bmatrix} \quad \text{Equação 5}$$

A matriz $\begin{bmatrix} f k_u & 0 & u_0 \\ 0 & -f k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}$ também designada por matriz **C**, representa os parâmetros intrínsecos e são quatro: $f k_u$ (também representado por α_u) e $-f k_v$ (α_v), que representam a relação entre coordenadas métricas e píxeis, e u_0 e v_0 , que representam as coordenadas do ponto principal no referencial da imagem (interceção entre o eixo ótico e o plano da imagem).



3.1.2 Rotação e translação da câmara

Geralmente os pontos do espaço 3D são expressos num referencial, designado por referencial do mundo. Os referenciais da câmara e do mundo estão relacionados por uma rotação e uma translação como mostra a Figura 5.

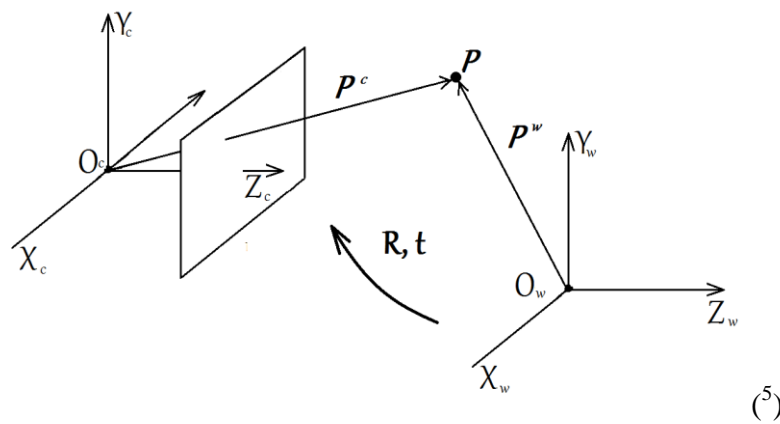


Figura 5: Transformação euclidiana entre coordenadas (figura retirada de [28]).

Seja P^w um vetor que representa um ponto P no referencial do mundo e P^c um vetor do mesmo ponto P no referencial da câmara, pode-se escrever:

$$P^c = RP^w + t \quad \text{Equação 6}$$

onde R representa a matriz de rotação de dimensão (3x3) e t o vetor de translação de dimensão 3.

Finalmente, relacionando as equações 5 e 6, obtém-se a seguinte expressão:

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = C(R|t) \begin{bmatrix} X^w \\ Y^w \\ Z^w \\ 1 \end{bmatrix} \quad \text{Equação 7}$$

⁵ Os eixos X, Y, Z da Figura 1 são nesta figura representados por (X_c, Y_c, Z_c)



O termo $C(\mathbf{R}/t)$ designa-se por matriz de projeção \mathbf{P} podendo a equação 7 ser expressa da seguinte forma:

$$\lambda \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \\ 1 \end{bmatrix} = \mathbf{P} \begin{bmatrix} \mathbf{X}^w \\ \mathbf{Y}^w \\ \mathbf{Z}^w \\ 1 \end{bmatrix} \quad \text{Equação 8}$$

Genericamente a matriz \mathbf{P} toma a forma da equação 9.

$$\mathbf{P} = \begin{bmatrix} \mathbf{p}_{11} & \mathbf{p}_{12} & \mathbf{p}_{13} & \mathbf{p}_{14} \\ \mathbf{p}_{21} & \mathbf{p}_{22} & \mathbf{p}_{23} & \mathbf{p}_{24} \\ \mathbf{p}_{31} & \mathbf{p}_{32} & \mathbf{p}_{33} & \mathbf{p}_{34} \end{bmatrix} \quad \text{Equação 9}$$

3.1.3 Determinação da matriz de projeção

Utilizando um conjunto de píxeis na imagem, cujo coordenadas 3D no referencial do mundo são conhecidas, é possível determinar a matriz de projeção.

Este cálculo envolve a determinação da solução de um sistema de equações que relaciona os referidos pontos. Existem várias técnicas para a resolução do sistema, como por exemplo, utilizando métodos de otimização não linear ou solução de mínimos quadrados.

A partir da matriz de projeção, através de técnicas de decomposição, obtêm-se a matriz dos parâmetros intrínsecos, de rotação e translação.

3.1.4 Distorção das lentes

O modelo da câmara para a geração de imagens utiliza, normalmente, o modelo de *pinhole* (buraco-da-agulha) que permite relacionar os pontos 3D do mundo e os píxeis da imagem.

Este modelo, quando conjugado com as coordenadas homogêneas, faz com que o mapeamento referido seja dado por uma transformação linear.

Todavia, se a aplicação em causa requerer elevados níveis de precisão, e as lentes da câmara utilizadas forem de baixo custo ou oferecerem um ângulo de visão amplo, o modelo referido falha, sendo a distorção radial das lentes a principal fonte de erros.



$$x^c_{corrected} = x^c(1 + k_1r^2 + k_2r^4 + k_3r^6) \quad \text{Equação 10}$$

$$y^c_{corrected} = y^c(1 + k_1r^2 + k_2r^4 + k_3r^6) \quad \text{Equação 11}$$

As equações 10 e 11 mostram as correções que se têm de efetuar nos píxeis no plano da imagem para que a imagem fique corrigida.

x^c e y^c são os píxeis (distorcidos) no plano da imagem, k_1 , k_2 e k_3 os coeficientes de distorção radial da lente e por fim $r = \sqrt{(x^c)^2 + (y^c)^2}$.

Nesta situação torna-se então necessário adicionar novos parâmetros intrínsecos às câmaras, de modo a considerar a distorção inserida pelas lentes.

3.1.5 Homografias

A estimação da matriz homográfica duma câmara, que se designa por matriz H , é um importante passo em muitas tarefas relacionadas com imagem, como por exemplo na reconstrução 3D, na geração de imagens panorâmicas, no próprio processo de calibração de câmaras ou para retificação métrica.

A matriz de homografia é uma simplificação da matriz de projeção pois considera que os pontos do mundo estão sobre um plano XY (ver figura 6), o que reduz o número de parâmetros a estimar em relação à matriz P .

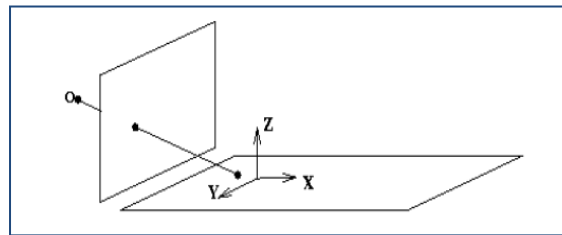


Figura 6: Homografia [28]

A equação 12 mostra o caso em que $Z^w=0$ (plano XY).

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} X^w \\ Y^w \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{14} \\ p_{21} & p_{22} & p_{24} \\ p_{31} & p_{32} & p_{34} \end{bmatrix} \begin{bmatrix} X^w \\ Y^w \\ 1 \end{bmatrix} \quad \text{Equação 12}$$



3.2 Correspondência de blobs

No trabalho proposto, a correspondência de *blobs* é realizada utilizando o método de *Fuentes* [25], que permite fazer correspondência entre *blobs* detetados em imagens consecutivas.

A matriz de correspondência tem dimensão $(n \times m)$, em que na correspondência de $(t-1)$ para (t) n é o número de *blobs* detetados na imagem $(t-1)$ e m o número de *blob* detetado na imagem (t) .

Existe um determinado número de características que podem ser usadas para fazer a correspondência. Poderá ser a distância dos centros de massa dos *blobs*, interseção dos retângulos envolventes (*bounding box*), o histograma de cor, ou a combinação de vários.

Se existir correspondência entre dois *blobs* em imagens consecutivas, a célula da matriz de correspondência associada a estes dois *blobs* é colocada ativa.

A Figura 7 mostra as regiões ativas detetadas em duas imagens consecutivas de uma sequência de vídeo, onde são representadas todas as possíveis situações consideradas no algoritmo.



Figura 7: Regiões de duas imagens consecutivas



Na Figura 8 as regiões ativas de cada imagem já se encontram numeradas e representadas pelas respectivas *bounding box*.

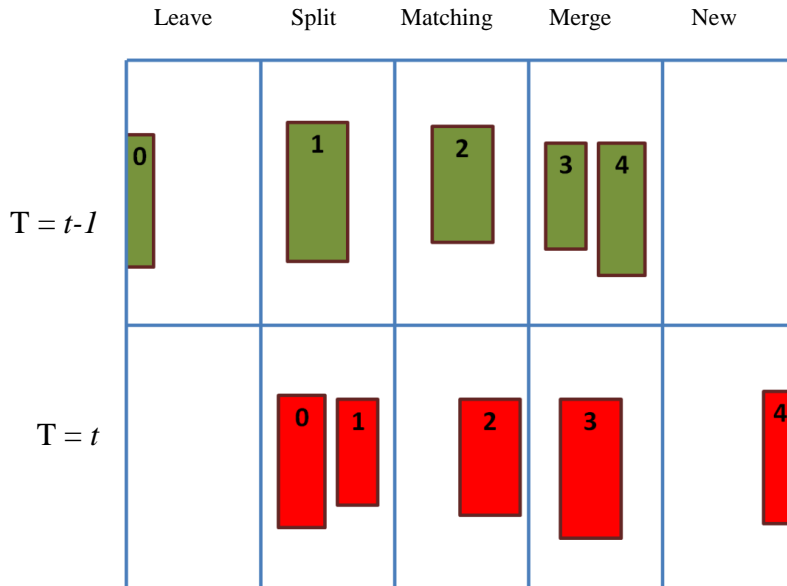


Figura 8: *Blobs* de 2 imagens consecutivas

Analisando a Figura 8, verifica-se que existem 5 situações possíveis de ocorrência:

Leave:

→ O *blob* 0 no instante ($t-1$) não tem nenhuma correspondência no instante (t).

New:

→ Não existe correspondência no instante ($t-1$) do *blob* 4 do instante (t).

Matching:

→ O *blob* 2 do instante ($t-1$) é correspondência do *blob* 2 do instante (t).

Merge:

→ Os *blobs* 3 e 4 do instante ($t-1$) têm correspondência com o *blob* 3 do instante (t).

Split:

→ O *blob* 1 do instante ($t-1$) têm correspondência com os *blobs* 0 e 1 do instante (t).



Cada duas imagens consecutivas originam duas matrizes de correspondência:

1. Matriz de correspondência da imagem no instante $(t-1)$ para a imagem no instante (t) a que corresponde a Tabela 1.
2. Matriz de correspondência da imagem no instante (t) para a imagem no instante $(t-1)$ a que corresponde a Tabela 2.

Tabela 1: (Anterior \rightarrow Corrente)

	0	1	2	3	4	
0	0	0	0	0	0	<i>leave</i>
1	1	1	0	0	0	<i>split</i>
2	0	0	1	0	0	<i>matching</i>
3	0	0	0	1	0	<i>merge</i>
4	0	0	0	1	0	

Tabela 2: (Corrente \rightarrow Anterior)

	0	1	2	3	4	
0	0	1	0	0	0	<i>split</i>
1	0	1	0	0	0	
2	0	0	1	0	0	<i>matching</i>
3	0	0	0	1	1	<i>merge</i>
4	0	0	0	0	0	<i>new</i>

A associação entre *blobs* envolve a análise das matrizes da Tabela 1: (Anterior \rightarrow Corrente) e da Tabela 2: (Corrente \rightarrow Anterior).

Pela análise das linhas da Tabela 1 temos:

Linha 0:

\rightarrow Não houve nenhuma correspondência, o que significa que o *blob* 0 do instante $(t-1)$ deixou de existir no instante (t) , a que corresponde a um *Leave*.

Linha 1:

\rightarrow Têm correspondência com a coluna 0 e 1 o que significa que houve *Split* do *blob* 1 do instante $(t-1)$ com o *blob* 0 e 1 do instante (t) .

Linha 2:

\rightarrow Tem correspondência unívoca com a coluna 2 o que significa que houve *Matching* do *blob* 2 do instante $(t-1)$ com o *blob* 2 do instante (t) .



Linhas 3 e 4:

→ Têm correspondência com a coluna 3, o que significa que houve *Merge* dos *blobs* 3 e 4 do instante ($t-1$) com o *blob* 3 do instante (t).

Analisando também a Tabela 2: (Corrente → Anterior), verifica-se que a linha 4 não tem nenhuma correspondência, o que quer dizer que o *blob* 4 na imagem corrente (instante (t)) é *New*.

3.2.1 Construção de trajetórias

As trajetórias são construídas através da análise das matrizes de correspondência ao longo da sequência de vídeo.

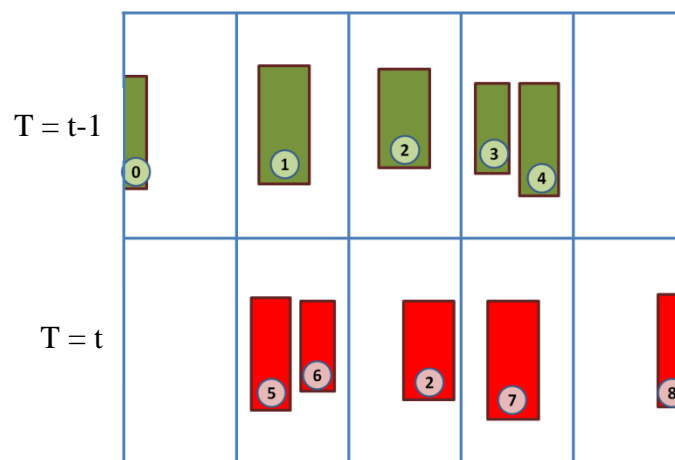


Figura 9: Trajetórias associadas

O número atribuído à trajetória vai sendo incrementado à medida que as trajetórias vão sendo formadas. A Figura 9 mostra a numeração das trajetórias em relação aos *blobs* representados na Figura 8.

Se na análise das matrizes de correspondência for detectado um *New* então uma nova trajetória é formada e a informação do *blob* correspondente (4) é guardada. A Figura 10 representa a formação de uma nova trajetória.

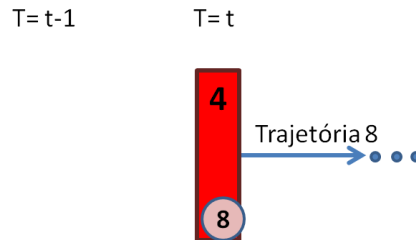


Figura 10: Nova trajetória

Uma trajetória vai sendo atualizada à medida que forem associados *blobs* em imagens consecutivas sendo a informação dos *blobs* também guardada. Isso acontece quando nas matrizes de correspondência ocorre uma situação de *Matching*. Na Figura 11 representa uma trajetória com dois *blobs*, no *blob 2* do instante $(t-1)$ é também armazenada a ligação para o *blob 2* do instante (t) .

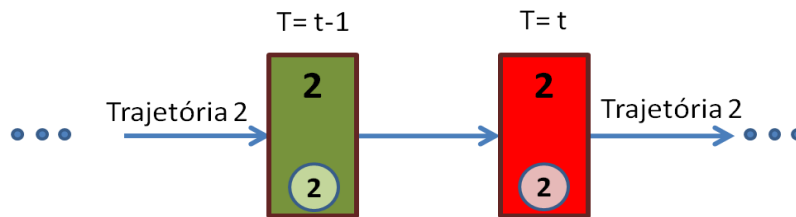


Figura 11: Troço de trajetória de dois *blobs* consecutivos

Quando se deteta um *Merge* as trajetórias respectivas terminam e é construída uma nova trajetória, como ilustra a Figura 12.

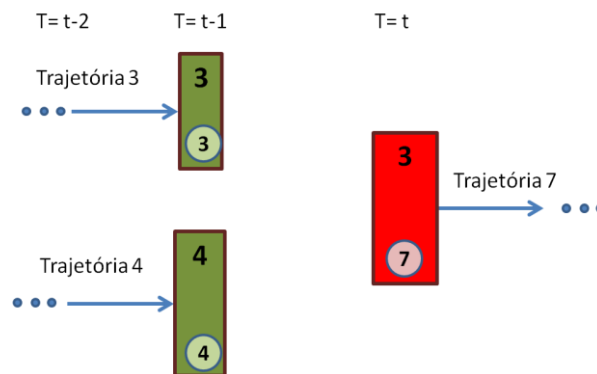


Figura 12: Criação de trajetória devido a *Merge*



Quando se deteta um *Split* a trajetória respectiva termina e são construídas novas trajetórias, como ilustra na Figura 13.

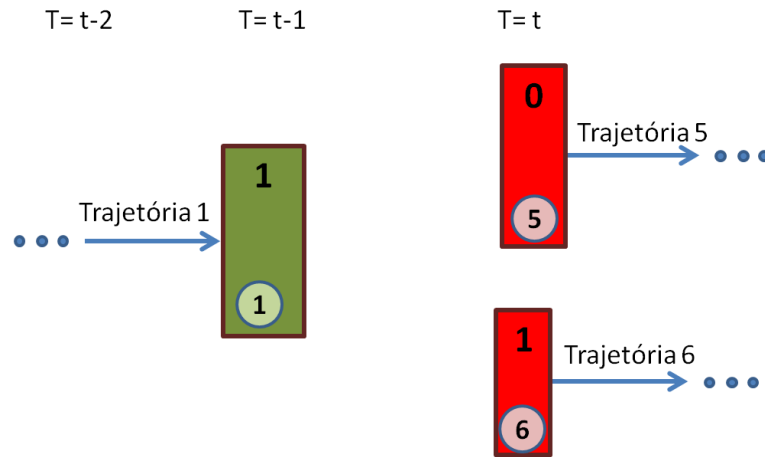


Figura 13: Criação de trajetória devido a *Split*

Quando se deteta uma situação de *Leave* o sistema não processa nenhuma informação uma vez que a trajetória já tinha sido atualizada no instante $(t-1)$ e não há nova informação associada, como ilustra a Figura 14.

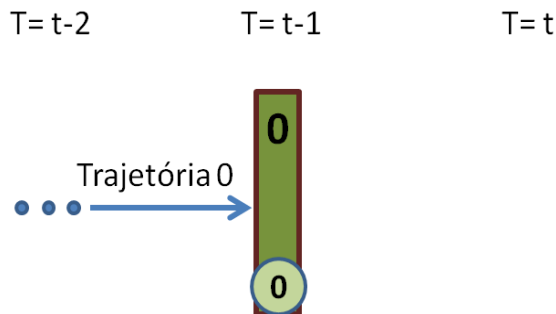


Figura 14: Ocorrência de um *Leave*



Após a construção das trajetórias, é necessário representar as relações que foram originadas por *Splits* ou *Merges*. Para armazenamento desta informação é criada uma tabela de ligações em que as linhas representam as trajetórias que terminam e as colunas as trajetórias que são geradas.

A Tabela 3 representa a ligação das trajetórias da Figura 7, onde se verificam as seguintes ocorrências

- *Split* da linha 1 (trajetória 1) com as colunas 5 e 6 (trajetórias 5 e 6, respetivamente);
- *Merge* da linha 3 e 4 (trajetórias 3 e 4, respetivamente) com a coluna 7 (trajetória 7).

Tabela 3: Tabela de ocorrências de Splits/Merges

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

Estes resultados podem ser representados graficamente através de grafos como se mostra na Figura 15, em que as trajetórias são representadas pelos nós e as ligações representam merge e *splits*.

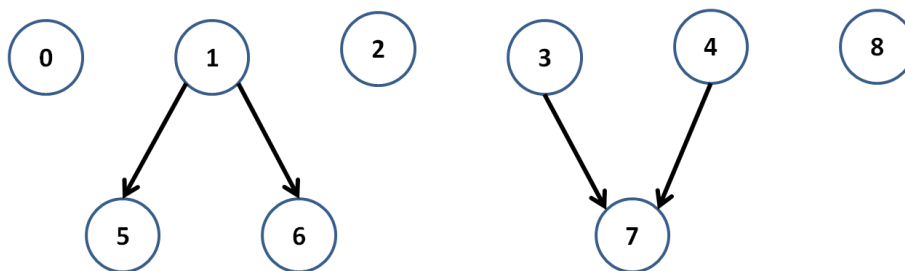


Figura 15: Grafo que representa as trajetórias da Figura 7



Para tornar mais robusto o processo de correspondência implementou-se um filtro de *Kalman* no movimento do centro de massas dos *blobs*, permitindo fazer a predição das *bounding boxes* dos *blobs* anteriores para a posição corrente. Esta técnica permite melhorar o algoritmo de correspondência de *blobs* uma vez que utiliza uma estimativa de velocidade aumentando a intersecção entre *bounding boxes*, como ilustrado na figura 16.

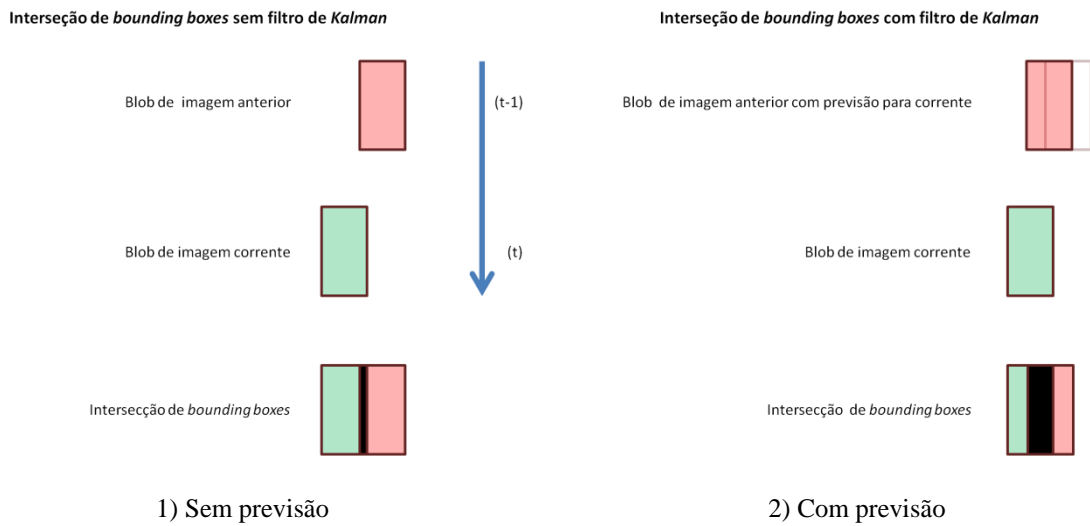


Figura 16: Correspondência de *blobs* por intersecção de *bounding boxes*



Método Proposto

4.1 Introdução

O método proposto permite através de várias câmaras vigiar uma determinada área de interesse, extraindo as trajetórias de pessoas que passam nessa área e relacionando essas trajetórias de modo a que se possa determinar as trajetórias totais.

Na Figura 17 apresenta o diagrama de blocos do trabalho proposto.

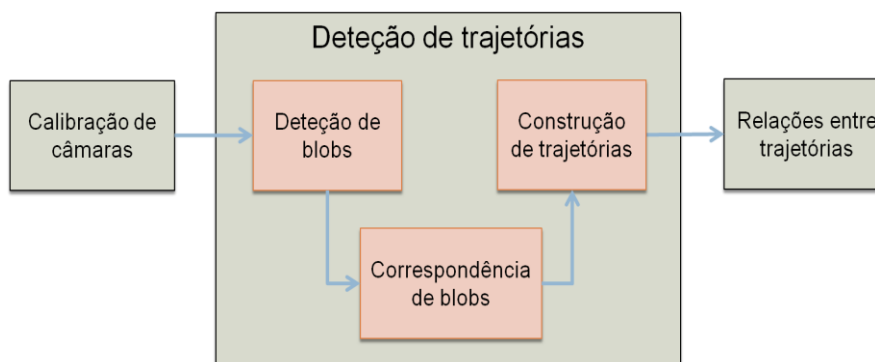


Figura 17: Diagrama de blocos do método proposto

No bloco *Calibração de câmaras* são obtidos os parâmetros intrínsecos, extrínsecos e coeficientes de distorção das lentes de cada câmara. São relacionados os referenciais associados a cada câmara, de modo a que se tenha um único referencial (referencial do mundo).

São extraídas as matrizes homográficas de modo a construir uma vista global virtual de topo que se designa por ‘planta’. Estas matrizes homográficas são também utilizadas para determinar as localizações das pessoas no referencial do mundo e na planta. Este



bloco fornece aos módulos seguintes as imagens corrigidas de cada câmara e as matrizes homográficas.

Após o módulo de *Calibração de câmaras* temos o bloco de *Deteção de trajetórias*, sendo este responsável pela segmentação, permitindo a deteção de zonas ativas (blobs) e a geração das trajetórias de cada câmara.

Por fim temos o bloco de *Relações entre trajetórias* onde são calculadas as relações entre trajetórias de modo a que seja possível fundir as trajetórias pertencentes à mesma pessoa ou grupo, que tenham sofrido oclusões.

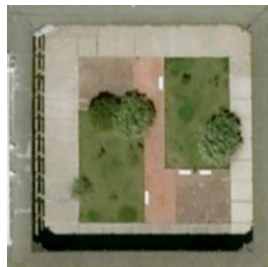
4.2 Determinação da área a vigiar

Num projeto de videovigilância em primeiro lugar é necessário definir qual a área que se pretende vigiar e dispor as câmaras de modo a que seja coberta o mais possível. Nas zonas de maior interesse é vantajoso a sobreposição dos campos de visão de câmaras para que se possam desambiguar algumas situações.

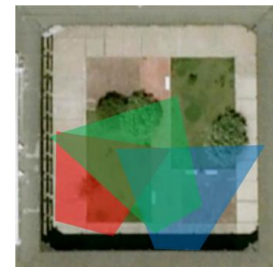
A Figura 18 mostra um exemplo de uma área onde se pretende fazer vigilância.



a) Imagem do local a vigiar



b) Imagem aérea, retirada de [29]



c) Exemplo de área vigiada com três câmaras

Figura 18: Áreas a vigiar

4.3 Calibração de Câmaras

Após o enquadramento das câmaras com a cena a vigiar, é necessário realizar a calibração do sistema. Poder-se-ia efetuar a obtenção dos parâmetros intrínsecos e extrínsecos simultaneamente, com as câmaras montadas no local definitivo, mas esse método iria provocar um erro bastante grande, pondo em causa a fiabilidade do sistema.



A extração dos parâmetros intrínsecos tem de ser efetuado com um objeto calibrador de precisão para que as características da lente, e os coeficientes de distorção sejam bem determinados. Assim, primeiramente as câmaras são calibradas de forma isolada e posteriormente é realizada a calibração conjunta do sistema.

4.3.1 Calibração de câmara isolada

Faz-se uma pré-instalação da câmara para definir a área de cobertura e ajustar a posição de *zoom* (distância focal) e focagem. As câmaras têm de ser montados com um sistema que permita removê-las e voltar a colocá-las sem alterar os parâmetros anteriores.

Feita a pré-instalação, retiram-se as câmaras sem alterar a sua posição de *zoom* e focagem, para obtenção dos parâmetros intrínsecos com um objeto calibrador de precisão, como mostra a Figura 19. Este procedimento tem de ser preciso, caso contrário a correção das imagens não se fará com qualidade, o que provocará uma grande distorção na imagem final e consequentes erros na obtenção das relações homográficas e posição da câmara.



Figura 19: Objeto calibrador de precisão

Este processo é realizado com base na biblioteca *openCV*, obtendo-se a matriz dos parâmetros intrínsecos, e os coeficientes de distorção. Esta função necessita de dois conjuntos de pontos relacionados entre si, os valores dos píxeis na imagem e os valores correspondente no referencial do objeto, sendo também necessário a dimensão da imagem em que se extraiu os respetivos pontos.

Estes conjuntos são calculados utilizando um algoritmo, onde são extraídos ordenadamente os valores dos píxeis, correspondentes aos cantos interiores do padrão apresentado na Figura 20, sendo posteriormente associados aos respetivos valores no referencial do objeto.

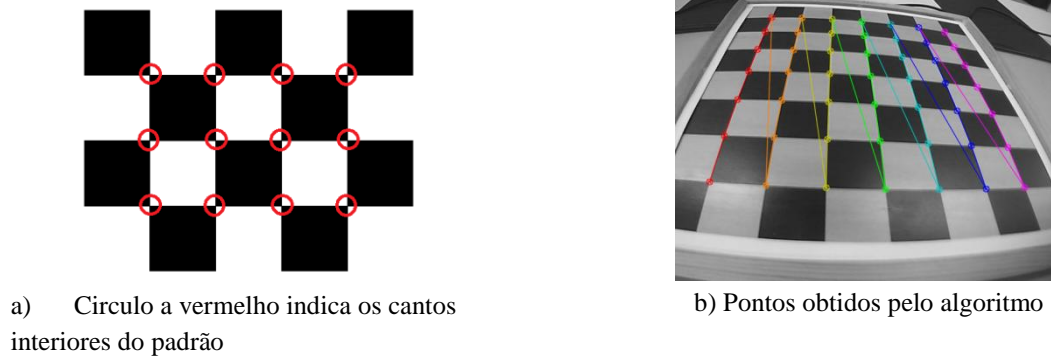


Figura 20: Obtenções dos pontos dos cantos interiores do padrão

Para uma maior precisão na calibração e dos coeficientes de distorção, o processo anterior é repetido com o objeto calibrador em várias posições, de modo a que se tenha um maior número de correspondências entre os pontos no objeto calibrador e os respetivos píxeis na imagem.

Aplicando os coeficientes de distorção obtidos no processo de calibração é possível corrigir as imagens das câmaras. A Figura 21 mostra a imagem corrigida da Figura 19.



Figura 21: Imagem corrigida

Após este passo de calibração, volta-se a montar a câmara na sua posição final, e através de um outro objeto calibrador, colocado na zona a vigiar extraem-se os parâmetros extrínsecos.

Através dos parâmetros extrínsecos é possível extrair a posição da câmara em relação ao referencial do objeto calibrador (matriz de rotação e vetor de translação), como ilustrado na Figura 22.d.

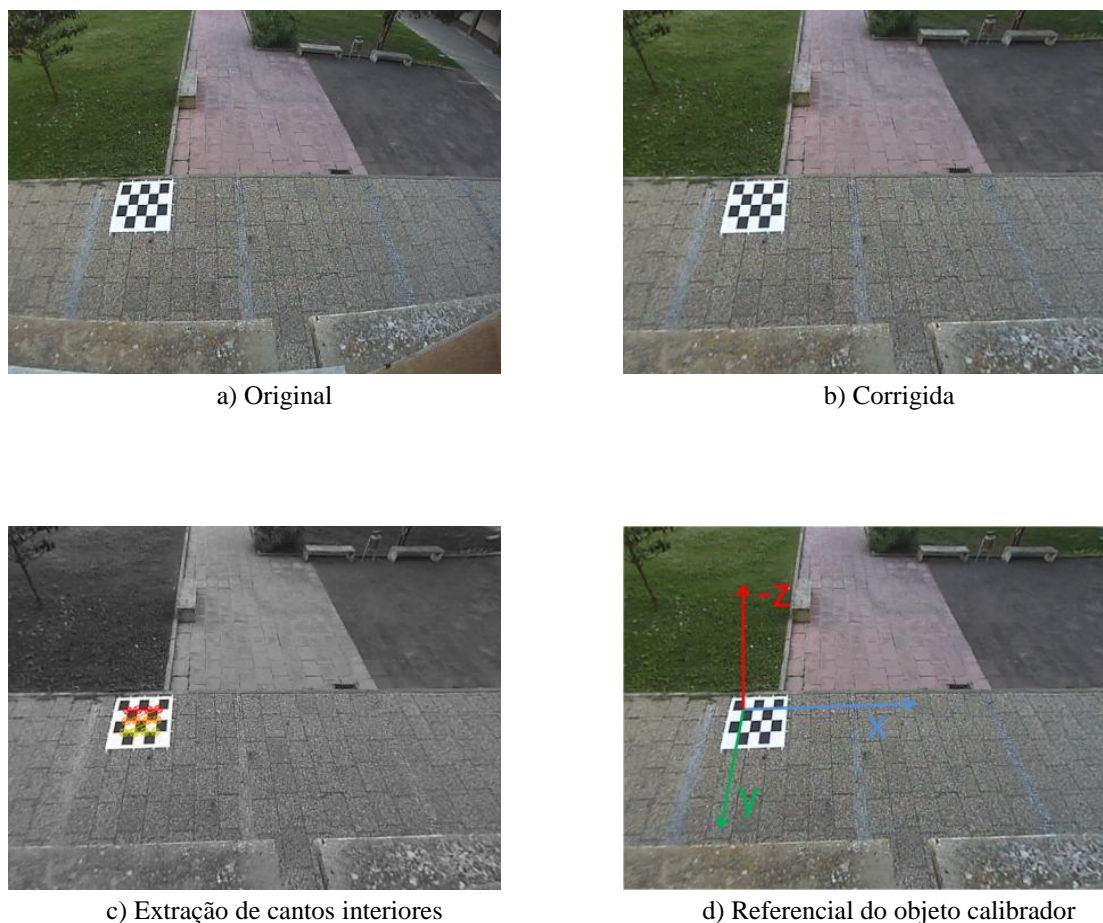


Figura 22: Obtenção dos parâmetros extrínsecos

Procede-se em primeiro lugar à correção da imagem da câmara, (Figura 22.b) e de seguida são extraídos os conjuntos de pontos (coordenadas na imagem e na referência do objeto calibrador) necessários para a estimação dos parâmetros extrínsecos (Figura 22.c).

Para obter os parâmetros extrínsecos utilizou-se um algoritmo do *openCV* que devolve o vetor de rotação e translação da câmara em relação à origem do referencial. Este algoritmo utiliza o método iterativo baseado na otimização de Levenberg-Marquardt. O objetivo deste método é encontrar a posição que minimiza o erro de retroprojeção, que é a soma dos quadrados das distâncias entre as projeções observadas.

Conhecendo os parâmetros extrínsecos calcula-se o espaço nulo da câmara (equação 13), que indica o centro ótico da câmara expresso no referencial do objeto calibrador.



$$\mathbf{O}_c = -\mathbf{R}^T \mathbf{t} \quad \text{Equação 13}$$

Considerando que a área a vigiar é plana e perpendicular ao eixo dos Z , (Figura 22.d), então a componente em Z do espaço nulo, devolve diretamente a altura da câmara e as componentes em X e Y a projeção da câmara no plano.

A matriz de R indica os valores de rotação que a câmara tem de efetuar em torno dos eixos do referencial do objeto calibrador, para que os referenciais do objeto calibrador e da câmara fiquem alinhados.

4.3.1.1 Homografia imagem mundo

Uma vez que se considera que a área a vigiar é plana, a relação entre as posições no mundo com as posições na imagem da câmara, é definida por uma homografia.

Esta relação homográfica vai ser obtida utilizando as correspondências entre os pontos do mundo e pontos na imagem (Figura 22.c).

Utilizando a matriz homográfica obtida, relaciona-se todos os píxeis na imagem com os valores correspondentes no mundo, mais concretamente, no plano XY .

4.3.1.2 Construção da imagem virtual de topo de uma câmara

Para criar uma imagem virtual com vista de topo, necessitamos de uma relação homográfica que relaciona os valores da imagem da câmara com a imagem de topo virtual.

Sabendo o valor da matriz $H_{\text{imagem} \rightarrow \text{mundo}}$, para se obter a matriz $H_{\text{imagem} \rightarrow \text{topo}}$, é necessário fazer o escalonamento das posições do mundo para as posições dos píxeis da imagem de vista virtual de topo cuja dimensão é pré-estabelecida, e que se optou por ser da mesma dimensão da imagem da câmara.

Foi implementada uma função que dada uma matriz homográfica $H_{\text{imagem_mundo}}$ e a dimensão da imagem, devolve a matriz homográfica $H_{\text{imagem_topo}}$ (Figura 23).

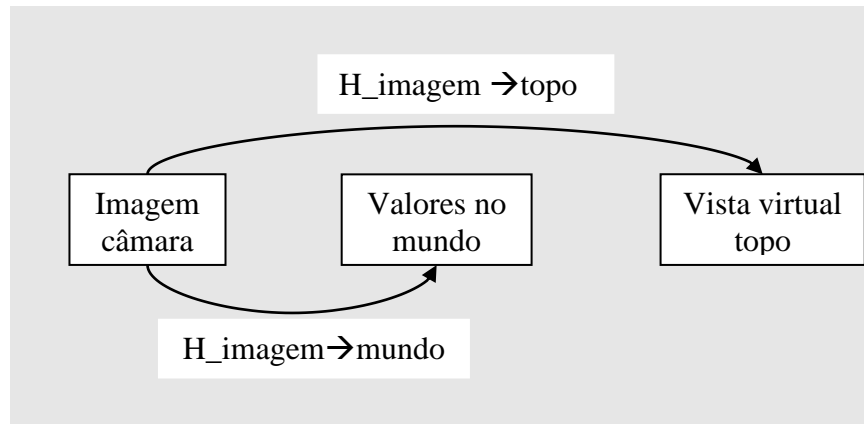
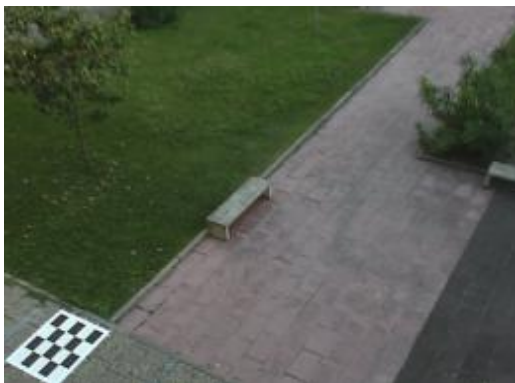


Figura 23: Criação de imagem topo virtual

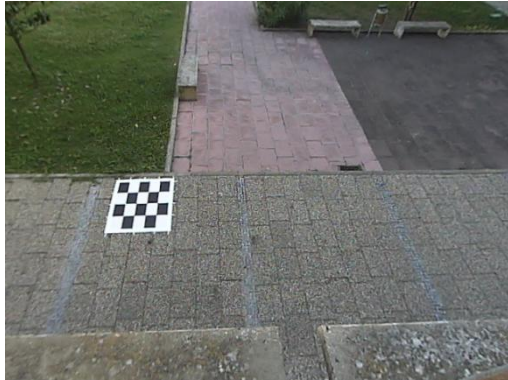
As imagens da Figura 24 mostram um exemplo do processo de calibração.



a) Câmera 1



b) Topo câmara 1



c) Câmera 2



d) Topo câmara 2



e) Câmera 3



f) Topo câmara 3

Figura 24: Imagens das câmaras e respetivas imagens de vista virtual de topo.

4.3.2 Calibração de várias câmaras

O objetivo da calibração de várias câmaras consiste na obtenção de um referencial comum a todas elas.

A calibração pode ser realizada por dois métodos distintos: O primeiro utiliza somente um objeto calibrador, e corresponde a situações em que tem vistas sobrepostas. O segundo é quando as vistas não são sobrepostas, sendo necessário utilizar mais do que um objeto calibrador.



4.3.2.1 Calibração com objeto calibrador comum

Quando a disposição das câmaras permite ter campos de visão com sobreposição, como mostra a Figura 25, é possível calibrar as câmaras com base no objeto calibrador colocado na mesma posição, ficando as câmaras com o mesmo referencial do mundo.

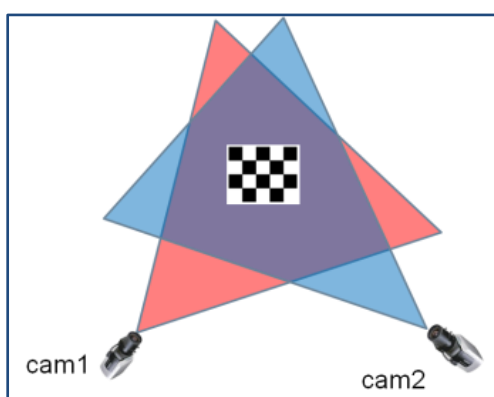


Figura 25: Calibração com objeto calibrador comum

4.3.2.2 Calibração com objeto calibrador individuais

Quando a disposição das câmaras não permite que tenham campos de visão sobrepostos, cada uma tem de ser calibrada com o objeto calibrador colocado em posições distintas (ver Figura 26). Deste modo cada câmara vai ficar referenciada para o seu objeto calibrador (referenciais do mundo diferentes) sendo necessário posteriormente relacionar os referenciais para corrigir as matrizes homográficas e posições das câmaras, de modos a terem o mesmo referencial.

Note-se que para se poder efetuar a calibração por este método, os objetos calibradores têm de ser iguais.

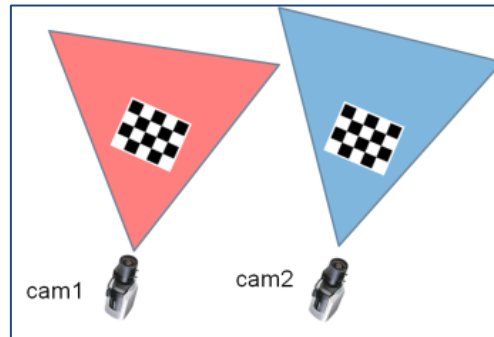


Figura 26: Calibração com objetos calibradores individuais

O processo que se adotou para relacionar os referenciais utiliza uma câmara auxiliar que permite a observação em simultâneo de ambos os objetos calibradores como mostra a Figura 27. Procede-se inicialmente à extração dos parâmetros intrínsecos e coeficientes de distorção desta câmara, como se procede à calibração de uma câmara isolada.

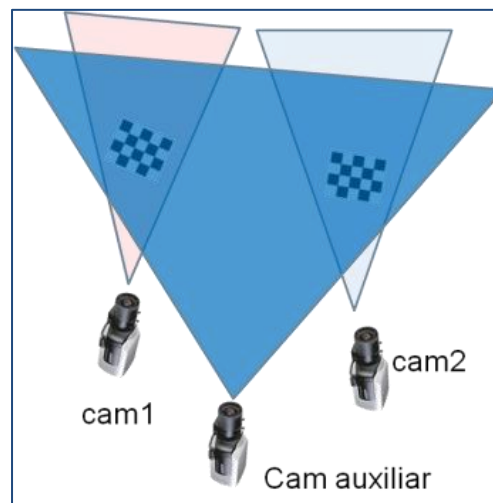


Figura 27: Relacionamento de referenciais

Após a captura da imagem contendo os dois objetos calibradores (Figura 28a), considera-se um dos objetos calibradores com o referencial do mundo global e calcula-se a matriz homográfica (imagem \rightarrow mundo) através do método explicado na seção (4.3.1.1).



Sabendo as posições na imagem dos cantos interiores do outro objeto calibrador, utiliza-se a mesma matriz homográfica (imagem \rightarrow mundo), para calcular todas as suas posições no mesmo referencial mundo global.

Relacionando agora as posições no referencial mundo global das posições dos cantos interiores, calculados anteriormente dos dois objetos calibradores (Figura 28b e Figura 28c), obtém-se a matriz homográfica que relaciona estes dois referenciais.



Figura 28: Relacionamento de referenciais

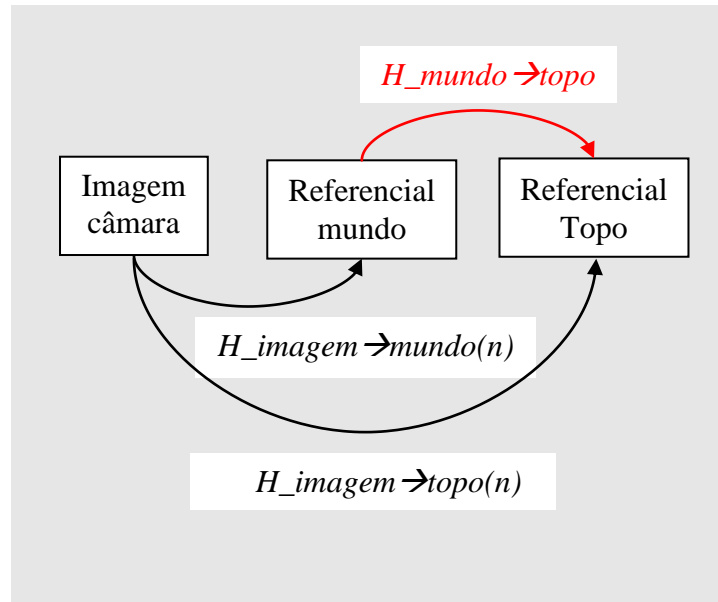
Para relacionar todas as câmaras para o referencial mundo escolhido como global, é necessário atualizar todos os valores das calibrações das câmaras que não estão referenciadas para o referencial mundo global.

Após a calibração completa do sistema os objetos calibradores e a câmara auxiliar são retirados.

4.3.3 Construção da planta final

Após todas as câmaras estarem calibradas para o mesmo referencial global procede-se à construção da imagem final virtual de topo e a homografia (mundo \rightarrow imagem).

Na seção (4.3.1.2) foram calculadas as homografias que permitem a construção das imagens virtuais de topo de cada câmara. A partir das homografias de qualquer câmara (Figura 29), calcula-se a homografia (mundo \rightarrow topo) (equação 14), de modo a incorporar toda a área vigiada e a posição das câmaras.

Figura 29: Obtenção da homografia $H_{\text{mundo} \rightarrow \text{topo}}$

$$H_{\text{mundo} \rightarrow \text{topo}} = H_{\text{imagem} \rightarrow \text{topo}(n)} * \text{inv}(H_{\text{imagem} \rightarrow \text{mundo}(n)})$$

Equação 14

Para construir a planta final calcula-se as homografias da imagem de cada câmara para a imagem final virtual de topo, obtendo-se as imagens parciais virtuais de topo, procedendo-se de seguida à sua colagem.

Na colagem das imagens testou-se dois métodos. O método 1 copia todos os píxeis diferentes de zero para a imagem final. A vantagem deste método é a rapidez, a desvantagem é que na construção as imagens ficam sobrepostas (Figura 30.a). O método 2 faz a média dos píxeis onde existe imagem. Este método exige mais processamento, no entanto a qualidade da imagem é superior (Figura 30.b).

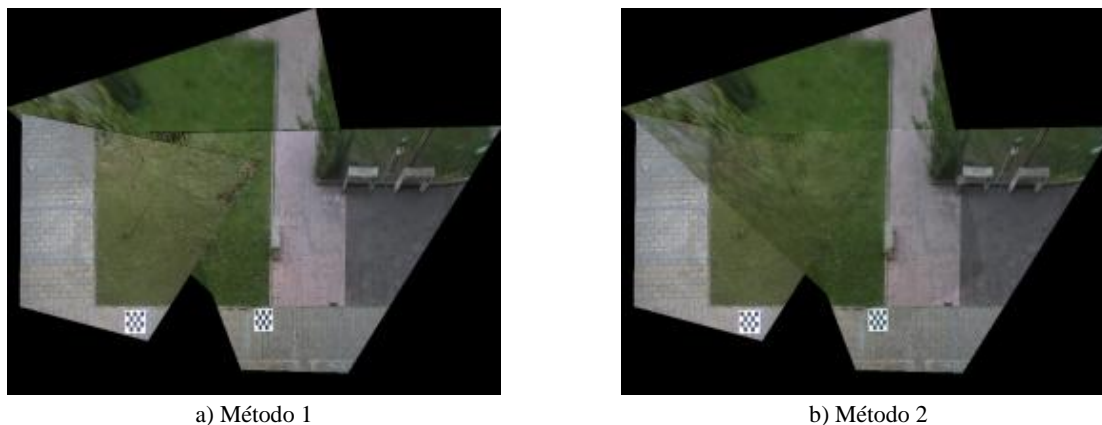


Figura 30: Imagens das vistas de topo

A Tabela 4 dá-nos a informação do tempo médio de execução da colagem das imagens parciais virtuais de topo.

Tabela 4: Tempos de execução da colagem de imagens

	Método 1	Método 2
Tempo de execução médio (ms)	49	100



4.4 Detecção de trajetórias

4.4.1 Detecção de *blobs*

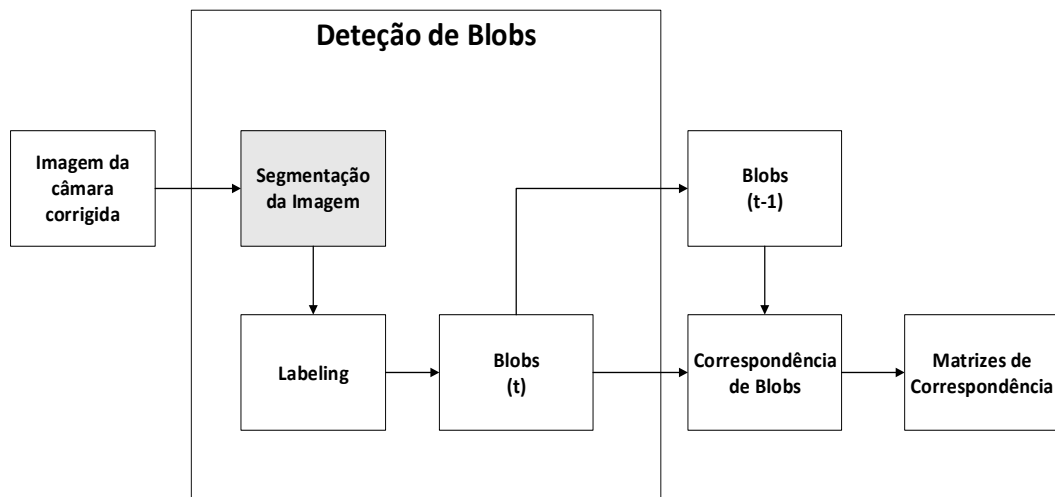


Figura 31: Diagrama de blocos de detecção de *blobs*

Para detecção de *blobs* é utilizado o conjunto de operações resumido no diagrama de blocos da Figura 31. No módulo de *Segmentação da Imagem* extrai-se as regiões ativas, no módulo de *Labeling* é aplicado um algoritmo de extração de componentes conexas, onde são etiquetadas as regiões ativas, e no módulo *Blobs(t)* são extraídos os *blobs* uteis e atualizada a informação referente a cada *blob*. Posteriormente esses *blobs* são memorizados em *Blobs (t-1)*, para que no bloco *Correspondência de Blobs* tenha-se acesso aos *blobs* de duas imagens consecutivas.



A Tabela 5 mostra algumas características relevantes, adquiridas na *Deteção de Blobs*, necessárias no processamento dos blocos seguintes.

Tabela 5: Variáveis atualizadas na deteção de *blobs*

Variável	Significado
pontosFronteira	Conjunto de pontos que limita a fronteira do <i>blob</i>
zonaInferior	Zona do <i>blob</i> situado na parte mais baixa da imagem
zonaSuperior	Zona do <i>blob</i> situado na parte superior da imagem
centroMassa	Zona do centro de massa do <i>blob</i>
num_frame	Número da imagem processada
num_blob	Número do <i>blob</i> atribuído no <i>labeling</i>
rectangle	Bounding box do <i>blob</i>
area	Área do <i>blob</i>
Hist (r,g,b)	Histograma de cor

4.4.1.1 Segmentação da imagem

A segmentação de imagem é realizada com base no conjunto de operações apresentadas no diagrama de blocos da Figura 32.

Considera-se que o módulo *Calibração de Câmara* (ver Figura 17) devolve as imagens já corrigidas.

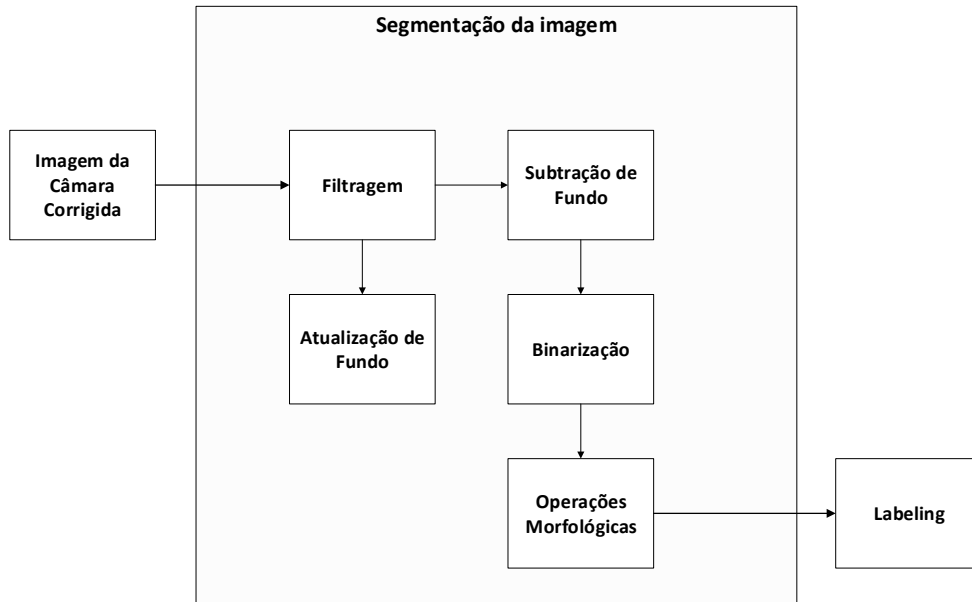


Figura 32: Diagrama de blocos da segmentação

A filtragem permite retirar algum ruído presente na imagem. Esta filtragem consiste na aplicação de um filtro passa-baixo com máscara gaussiana.

Para a detecção de movimento, é feita a subtração da imagem filtrada com a imagem de fundo, na qual obtemos a imagem diferença.

Posteriormente, à imagem diferença é aplicada uma binarização.

A imagem de fundo é atualizada para poder incorporar mudanças lentas de luminosidade e poder incorporar objetos que inicialmente não pertenciam à imagem de fundo.

Por fim, são aplicadas operações morfológicas que permitem suavizar a zona detetada e eliminar pequenas regiões ativas. Para este efeito são aplicadas as operações de abertura.

Para reduzir o problema da sobre-segmentação, desenvolveu-se um método usando a intersecção das *bounding boxes* da imagem do instante $(t-1)$ com as da imagem do instante (t) . Se houver mais do que uma *bounding box* do instante (t) que intersete outra *bounding box* do instante $(t-1)$ com valores de intersecção superiores a um limiar, então as do instante (t) é sobre-segmentação, sendo corrigida então as características inerentes num *blob* eliminando os restantes.



Posteriormente as imagens binárias são etiquetadas aplicado um algoritmo de extração de componentes conexas como já foi referido.

Para determinar o conjunto de blobs válidos, as respectivas áreas são testadas para eliminar as regiões ativas com um valor menor que um determinado limiar.

A Figura 35 mostra o resultado da segmentação para o conjunto de imagens da Figura 34.

Verifica-se na imagem da Figura 35, que algumas regiões ativas não foram consideradas em virtude da sua área ser pequena.



Figura 33: Imagens de fundo



Figura 34: imagens de uma sequência



Figura 35: Imagens segmentadas



4.4.2 Correspondência de *blobs*

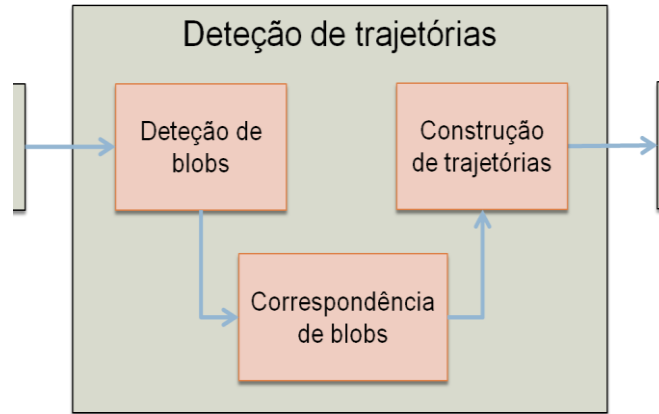


Figura 36: Detecção de trajetórias

O algoritmo adotado para fazer a correspondência de blobs baseou-se no método de Fuentes[25], em que o critério para correspondência foi a interseção de *bounding box*. Se o valor da interseção for superior a um determinado limiar, os *blobs* correspondem. Este método torna o processo independente da posição dos *blobs* na imagem.

Existem várias hipóteses de associação dos blobs, como foi descrito na seção 3.2 dando origem às matrizes de correspondência.

Estas matrizes de correspondência são em seguida analisadas no bloco Construção de trajetórias cujo funcionamento é descrito na seção 3.2.1.

Na Figura 37 mostra dois exemplos de ocorrência de *merges*; o primeiro *merge* corresponde aos *blobs* associados às trajetórias 8 e 9 da Figura 37.a, ao *blob* que irá ser associado à nova trajetória 11 da Figura 37.b. O segundo *merge* corresponde aos *blobs* associados às trajetórias 6 e 11 da Figura 37.c, ao *blob* que irá ser associado à nova trajetória 18 da Figura 37.d.

Na Figura 38 mostra um exemplo de ocorrência de dois *splits*, o primeiro *split* corresponde ao *blob* associado à trajetória 18 da Figura 38.a, aos *blobs* que irão ser associados às trajetórias 19 e 20 da Figura 38.a. O segundo *split* corresponde ao *blob* associado à trajetória 20 Figura 38.c, aos *blobs* que irão ser associados às trajetórias 22 e 23 da Figura 38.d.



Figura 37: Sequência de imagens de *merges*

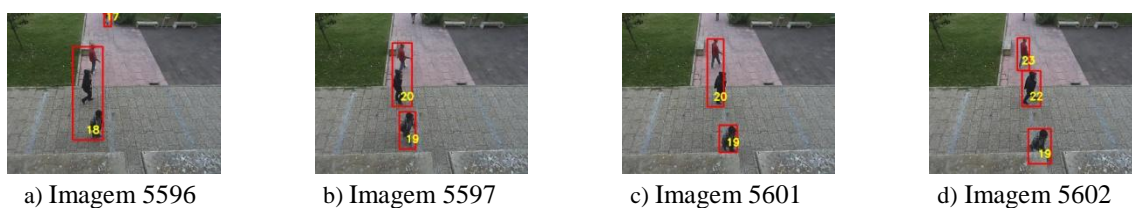


Figura 38: Sequência de imagens de *splits*

Adotou-se para representação gráfica das trajetórias grafos, em que no nó se insere um número, em que a parte inteira corresponde à identificação da câmara e a parte decimal corresponde à trajetória da respectiva câmara. Para melhor percepção visual, os nós, são preenchidos com uma cor distinta para cada câmara.

Na ligação entre os nós, a cor da seta é indicativa do tipo de ligação, que no caso de *splits* e *merges* convencionou-se de cor preta.

A Figura 39 ilustra graficamente a sequência das imagens da Figura 38.

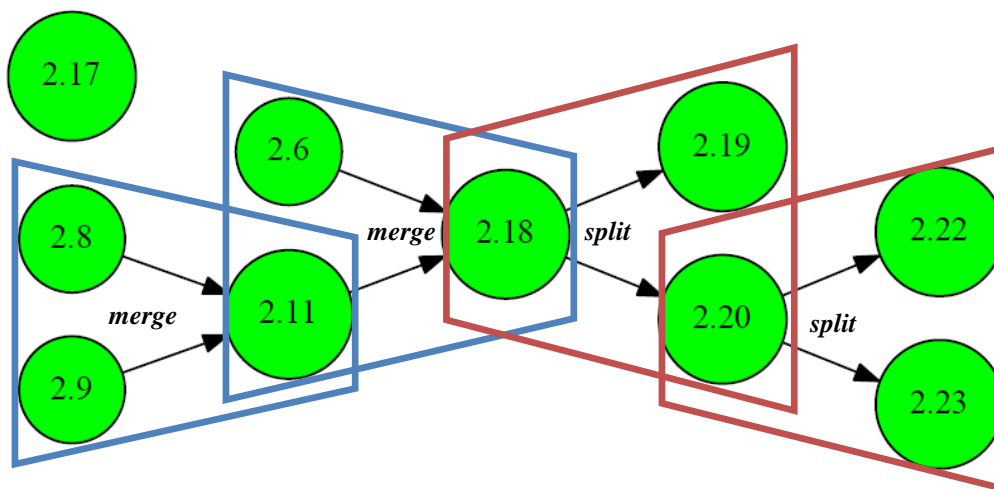


Figura 39: Representação de trajetórias e ligações merge/split



4.5 Relações entre trajetórias

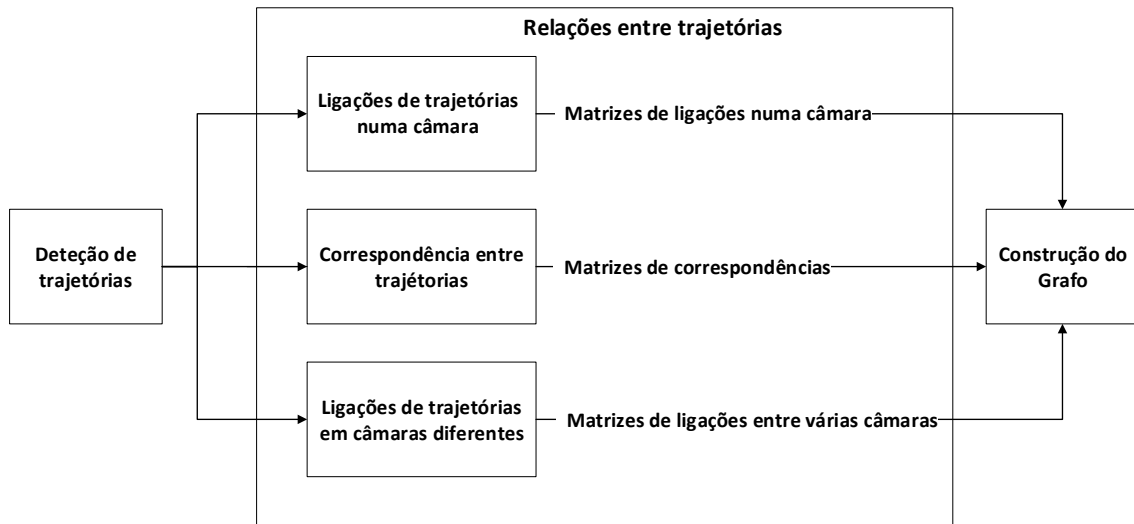


Figura 40: Relações entre trajetórias

Na detecção de trajetórias, o processamento é efetuado tomando como informação os píxeis da imagem da câmara. Nas relações de trajetórias, o processamento passa a ser efetuado com base na informação correspondente ao referencial do mundo.

Para simplificação dos algoritmos de implementação, orientaram-se as câmaras de modo a que as pessoas fiquem com a zona dos pés na parte inferior da imagem da câmara, ficando este ponto designado por ponto inferior. A zona da cabeça ficará então na parte superior da imagem das câmaras ficando este ponto designado por zona superior.

Podemos ter três tipos de relações conforme mostra na Figura 40.

As ligações das trajetórias numa câmara, que relacionam as trajetórias que ocorrem numa câmara e pertencem à mesma pessoa ou grupo. As correspondências entre trajetórias, que relaciona as trajetórias entre câmaras, pertencentes à mesma pessoa ou grupo e que ocorre no mesmo intervalo de tempo. Por fim temos as ligações das trajetórias entre câmaras que relaciona as trajetórias entre câmaras e que pertencem à mesma pessoa ou grupo.

Para construir estas relações são utilizadas matrizes de associação de trajetórias, em que os índices das linhas e colunas correspondem às trajetórias em estudo.



As matrizes de ligação das trajetórias da mesma câmara são quadradas e tem a dimensão do número de trajetórias de cada câmara. As linhas representam a zona da trajetória que termina e as colunas a zona das trajetórias que começam. Quando um elemento da matriz tem o valor unitário, significa que a trajetória com o número de linha poderá estar relacionada com a trajetória com o número de coluna.

As matrizes de correspondência entre trajetórias indicam aquelas que representam a mesma pessoa ou grupos, detetadas por mais de uma câmara. As linhas correspondem às trajetórias de uma câmara e as colunas às trajetórias de outra câmara. Quando um elemento da matriz tem o valor unitário significa que a trajetória com o número de linha da 1ª câmara corresponde à trajetória com o número de coluna da 2ª câmara.

Nas matrizes de ligação entre trajetórias de câmaras diferentes, as linhas correspondem às trajetórias que terminam numa câmara e as colunas às trajetórias que começam na outra câmara. Quando um elemento da matriz tem o valor unitário, significa que a trajetória com o número da linha da 1ª câmara poderá estar relacionada com a trajetória com o número da coluna da 2ª câmara.

A Tabela 6 indica o número de matrizes necessárias para a construção das relações. Em ligações numa câmara esse valor é sempre igual ao número de câmaras. Na Correspondência entre câmaras indica também o número máximo de matrizes, a que corresponde o caso de a área vigiada de todas as câmaras se sobreporem. Um outro caso particular será de não haver sobreposição da área vigiada em quaisquer câmaras, tornando o processamento e construção de matrizes relacionadas com a correspondência desnecessário. Por último em ligação de câmaras diferentes indica o número máximo de matrizes e está relacionado com a disposição das áreas vigiadas de cada câmara.

Tabela 6: Número de matrizes de relações para N câmaras

	Número de tabelas	2 Câmaras	3 Câmaras	10 Câmaras
Ligação numa câmara	N	2	3	10
Correspondência entre câmaras	$\sum_{n=1}^{N-1} n$	1	3	45
Ligação de câmaras diferentes	$2 * \sum_{n=1}^{N-1} n$	2	6	90



4.5.1 Ligações de trajetórias numa câmara

As ligações de trajetórias numa câmara permitem relacionar trajetórias que podem representar a mesma pessoa ou grupo que devido a oclusões ou falhas de segmentação, são interrompidas/partidas.

Para que haja ligação entre duas trajetórias numa mesma câmara, é necessário que obedeam a critérios de proximidade, i.e. as características relevantes de proximidade do último *blob* da trajetória que termina e do primeiro *blob* da trajetória que se inicia, têm de ser inferior a um limiar e ocorram dentro de uma janela temporal. É necessário também que a trajetória que termina, não tenha dado origem a *split/merge*, assim como a que se inicia não tenha sido originada por *split/merge*.

No exemplo da Figura 41 a trajetória 1 liga à trajetória 3 se Δx e Δt forem inferiores ao respetivos limiares.

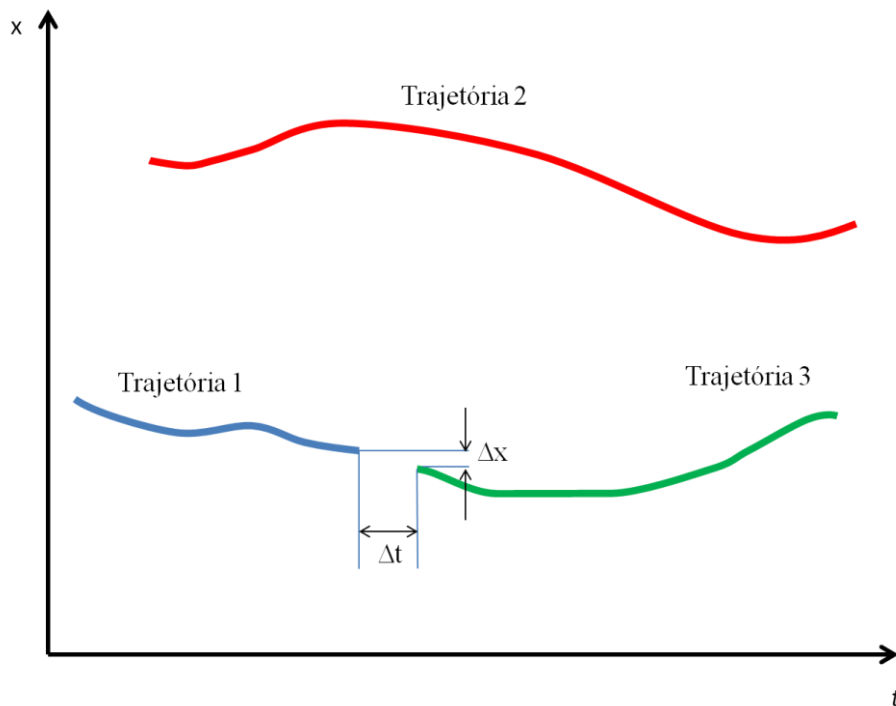


Figura 41: Ligação de trajetórias numa câmara



Na Figura 42 mostra um exemplo de ligações de trajetórias numa câmara.



Figura 42: Ligação em uma câmara

A Figura 43 ilustra graficamente a sequência das imagens da Figura 42. A ligação entre nós representa uma ligação de trajetórias numa câmara e convencionou-se de cor verde.

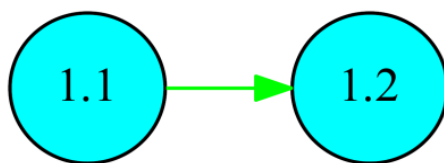
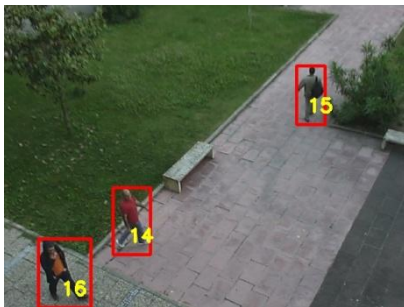


Figura 43: Representação gráfica de ligações de trajetórias numa câmara

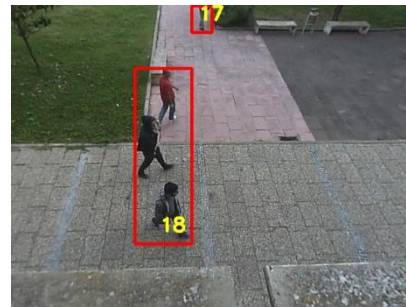


4.5.2 Correspondências entre trajetórias entre câmaras

Este tipo de ligação informa quais as trajetórias que acontecem em simultâneo no espaço temporal e espacial. Permite separar as pessoas de um grupo, existentes numa trajetória vista sob uma determinada câmara.



a) Imagem 5594 da câmara 1



b) Imagem 5594 da câmara 2

Figura 44: Trajetórias nos instantes de tempo 5594

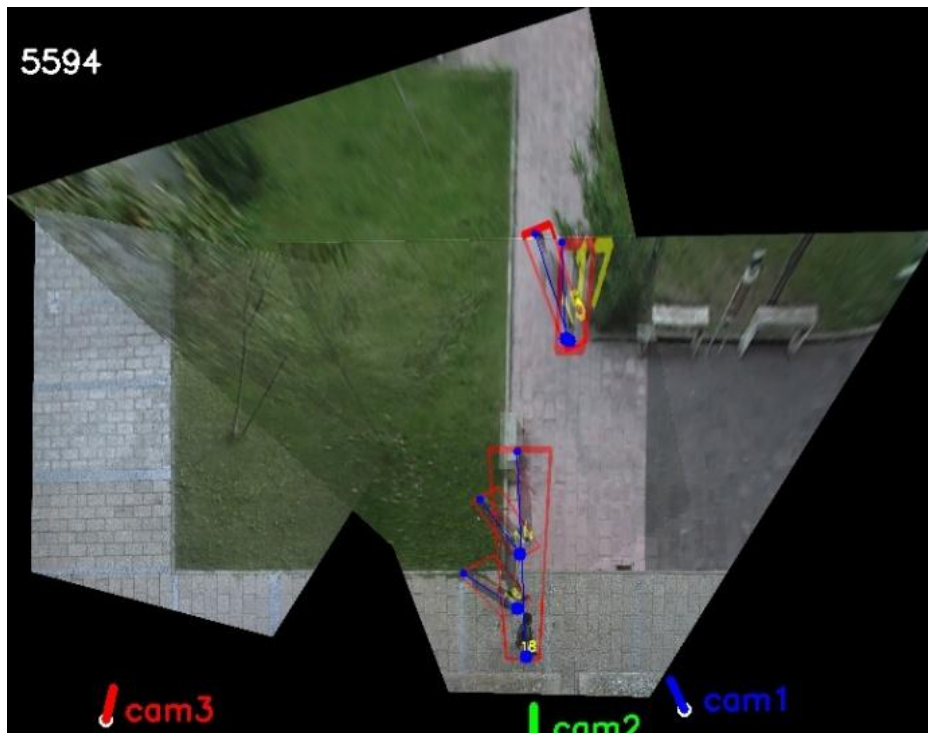


Figura 45: Correspondência de trajetórias



Na Figura 45 temos um exemplo de correspondência, esquematizada na Figura 46 em que na câmara 1 são visualizadas 3 pessoas a que vamos designar por a1,b1e d1 e na câmara 2 são visualizadas 4 pessoas a que vamos designar por a2,b2,c2, d2.

As trajetórias a que correspondem as pessoas a1,b1 e d1 nesse instante de tempo são respetivamente 1.14, 1.16 e 1.15.

Nesse mesmo instante de tempo as pessoas a2,b2 e c2 formam um grupo correspondente à trajetória 2.18 e a pessoa d2 corresponde à trajetória 2.17.

Para haver correspondência de trajetórias entre duas câmaras, além de terem de coexistir no mesmo instante tempo, as distâncias entre os pontos inferiores vistos na planta dos *blobs* representativos dessas trajetórias ou a distância entre o ponto inferior e o segmento de reta formado pelos pontos inferior e superior da outra também vistos igualmente na planta dos *blobs* representativos das trajetórias, têm de ser inferior a um determinado valor Δ .

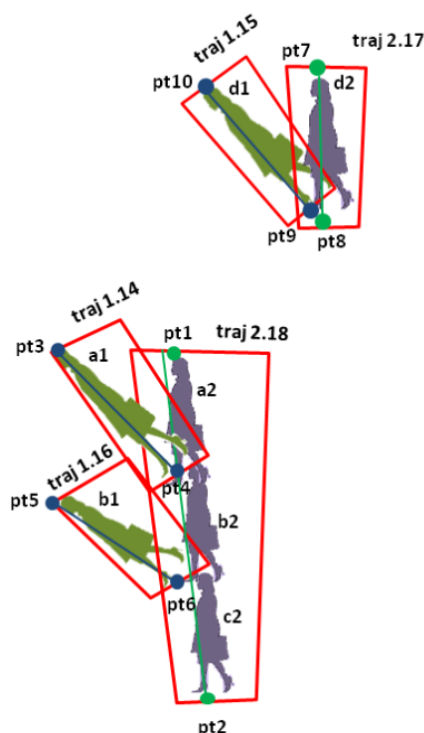
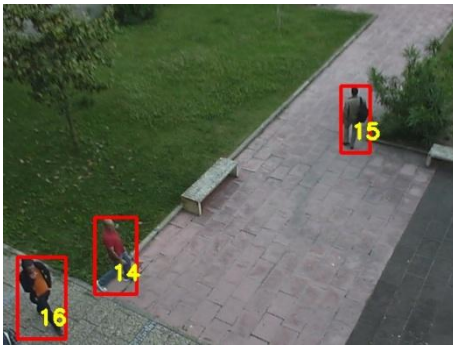


Figura 46: Correspondência de trajetórias (representação gráfica)

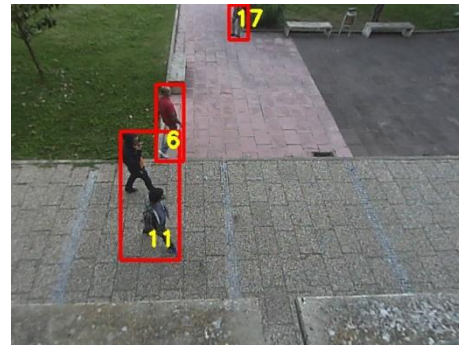


Analisando a Figura 46 verifica-se que a trajetória 1.14 tem correspondência com a trajetória 2.18, pois a distância do ponto pt4 ao segmento de reta formado pelos pontos pt1 e pt2 obedece às condições de proximidade. Também a trajetória 1.16 tem correspondência com a trajetória 2.18 pois neste caso é o ponto pt6 que se situa próximo do segmento de reta. Verifica-se outra correspondência entre as trajetórias 1.15 e 2.17 em que a condição de proximidade se verifica entre os pontos pt9 e pt8.

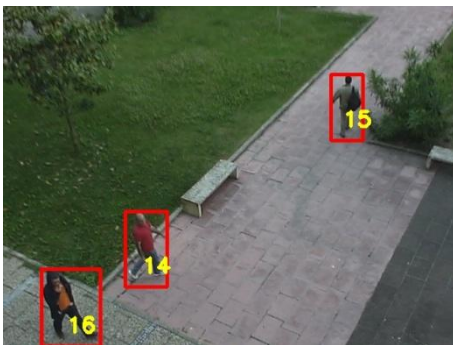
A Figura 47 ilustra um exemplo de uma sequência de imagens em que existem correspondência de trajetórias.



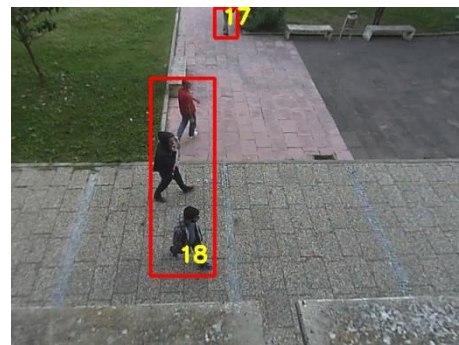
a) Imagem 5590 da câmara 1



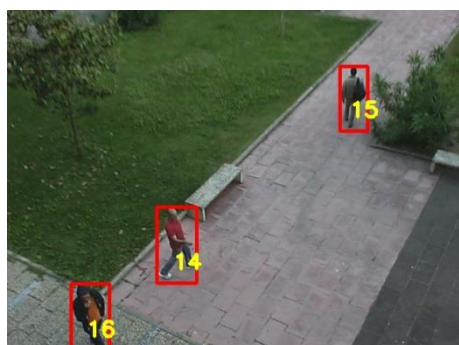
b) Imagem 5590 de câmara 2



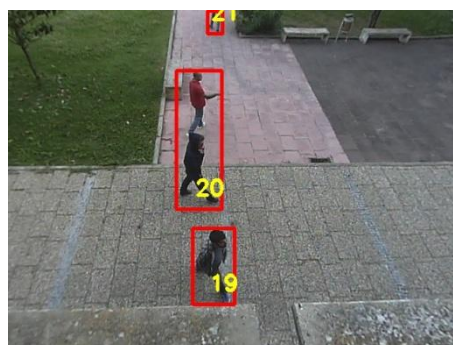
c) Imagem 5594 da câmara 1



d) Imagem 5594 da câmara2



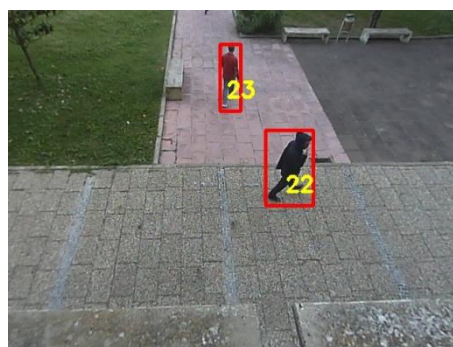
e) Imagem 5598 da câmara 1



f) Imagem 5598 da câmara 2



g) Imagem 5610 da câmara 1



h) Imagem 5610 da câmara 2

Figura 47: Correspondência de trajetórias

A Figura 48 representa graficamente as correspondências das trajetórias do exemplo da Figura 47.

A ligação entre nós representa uma correspondência de trajetórias entre câmaras e convencionou-se de cor vermelha sendo a sua representação é bilateral.

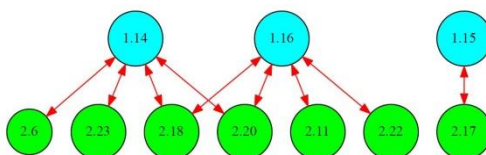


Figura 48: Representação gráfica de correspondência de trajetórias



4.5.3 Ligação de trajetórias entre várias câmaras

Este tipo de ligação tem o mesmo princípio de funcionamento ao da ligação de trajetória uma câmara, mas agora as ligações são entre câmaras, ou seja permitem relacionar trajetórias que podem representar a mesma pessoa ou grupo em que devido à passagem por zonas não cobertas, essas trajetórias são partidas. O critério de proximidade tem uma maior variância implicando o aumento de incerteza da ligação.

A Figura 50 representa graficamente as correspondências das trajetórias do exemplo Figura 49.

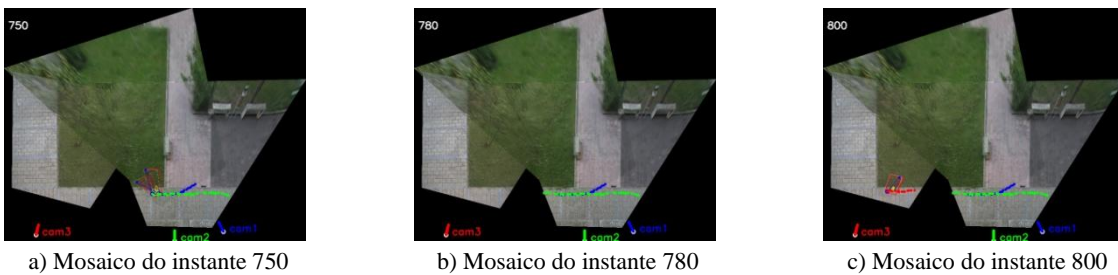


Figura 49: Ligação em câmaras diferentes

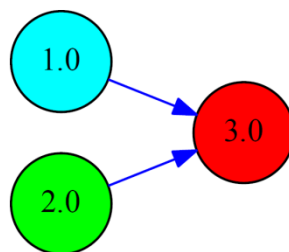


Figura 50: Representação gráfica de ligações de trajetórias em câmaras diferentes



4.6 Geração do grafo

A estrutura do grafo para posteriormente ser visualizada é construída recorrendo às matrizes de correspondência e às matrizes de relações de trajetórias (ligações e correspondência). A obtenção do grafo é obtida através de um visualizador[30].

Os elementos constituintes do grafo são os nós e as ligações.

4.6.1 Nós

Os nós representam as trajetórias, têm a identificação da câmara que representa a trajetória de cada nó e o número de trajetória dentro dessa câmara. Para melhor visualização representa-se as trajetórias de cada câmara com uma cor diferente. Pode-se incluir outros elementos identificativos dentro do nó, como por exemplo, início e fim da ocorrência da trajetória, velocidade média e altura das regiões ativas.

4.6.2 Ligações

As ligações representam o tipo de relação entre as trajetórias sendo diferenciadas através da sua cor.

- a) Ligações a preto representam os *splits* e *merges* entre as trajetórias. A informação para estas ligações é obtida em tempo real.
- b) As ligações a verde representam as ligações entre trajetórias numa mesma câmara. Quando um nó A está ligado a um nó B através de um seta a verde no sentido de A para B, significa que a trajetória B poderá ser a continuação da trajetória A em que A e B pertencem à mesma câmara.
- c) As ligações a vermelho representam as correspondências entre trajetórias. Quando um nó E está ligado a um nó F significa que as trajetórias E e F representam a mesma pessoa ou grupo e coexistem no referencial espacial e temporal. A seta é representada em duplo sentido porque se E corresponde F implicitamente F também corresponde E.
- d) As ligações a azul representam as ligações entre trajetórias mas em câmaras diferentes. Quando um nó C está ligado a um nó D através de uma seta a azul signifi-



ca que a trajetória D poderá ser a continuação da trajetória C em que C e D pertencem a câmaras diferentes.

A Figura 51 mostra um exemplo de um grafo de maior complexidade e a Figura 52 mostra as mesmas trajetórias com a maior quantidade de informação sobre a trajetórias.

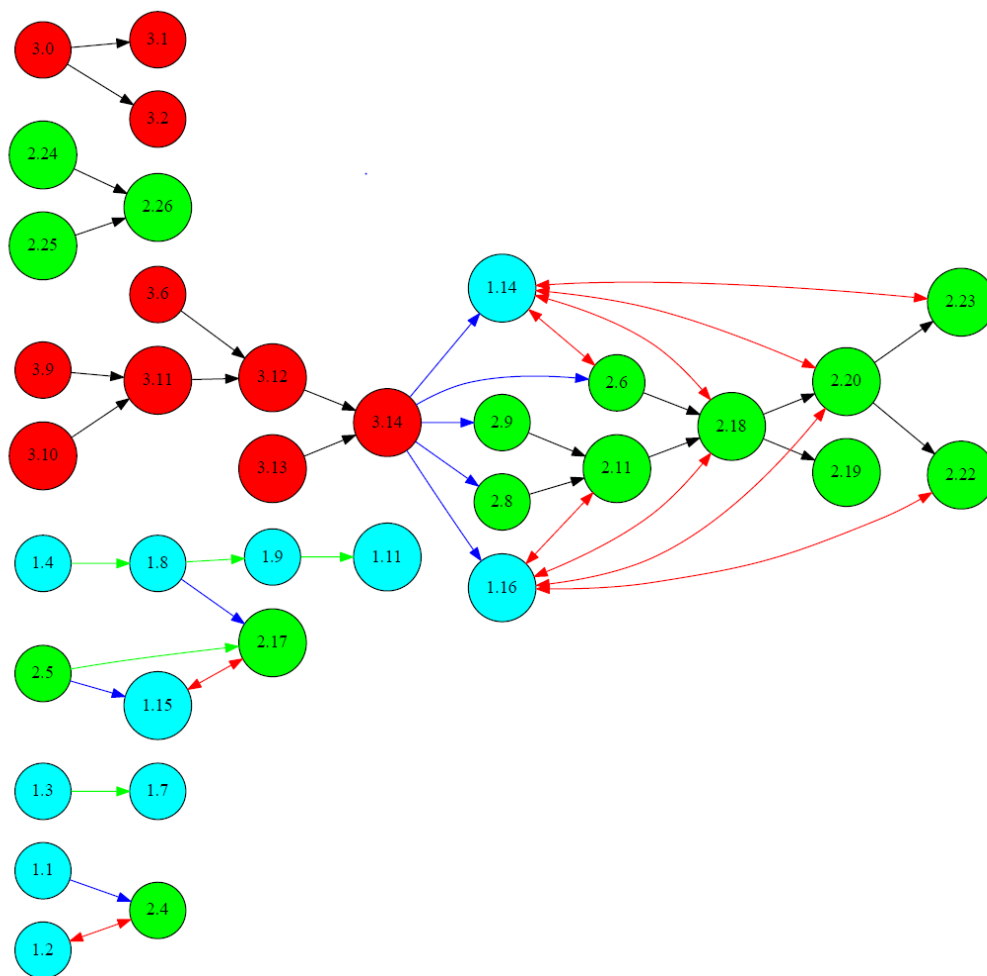


Figura 51: Trajetórias e relações

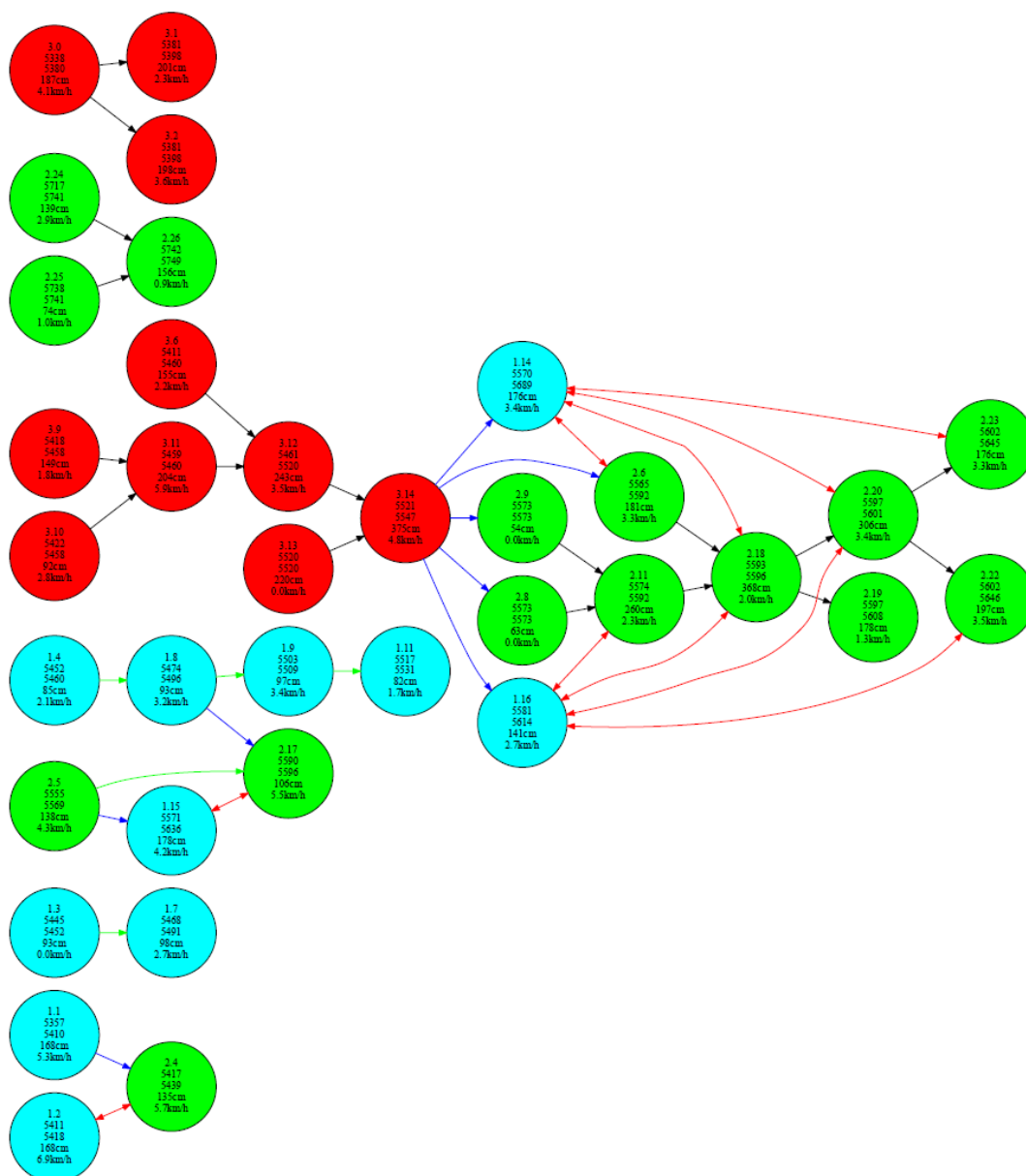


Figura 52: Trajetórias e relações com informação extra



Resultados

5.1 Tempos de execução

Um dos principais objetivos deste sistema é ser capaz trabalhar em tempo real que está relacionado diretamente com o bloco de Detecção de trajetórias.

As sequências de imagens utilizadas no presente trabalho foram adquiridas com uma velocidade de sete imagens por segundo, utilizando câmaras IP sincronizadas por *software*.

Coloca-se então a questão: Considerando sete imagens por segundo a velocidade de aquisição, consegue o nosso sistema processar essa quantidade de informação em tempo real?

Na Tabela 7 encontram-se alguns tempos de execução com um equipamento (*i5@2.4GHz*). O teste foi efetuado com as sequências de imagem já gravadas em disco e foi calculada o tempo médio de execução de alguns componentes do sistema.

Tabela 7: Tempos médios de execução

		<i>Milissegundos</i>	
Tempo execução total	Mínimo	192	
	Média	243	
	Máximo	329	
Captura sequência de imagens	Mínimo	74	
	Média	77	
	Máximo	121	
Detecção de trajetórias	Detecção de <i>Blobs</i>	Mínimo	102
		Média	114
		Máximo	160
	Correspondência de <i>Blobs</i>	Mínimo	8
		Média	52
		Máximo	128



Dos dados obtidos conclui-se que a aquisição de imagens e a Detecção de *Blobs* é relativamente constante ao longo do processo.

No bloco de Correspondência de *Blobs*, verifica-se que o tempo gasto no processamento das sequências de imagens aumenta com o número de *blobs* detetados. O valor mínimo é bastante baixo devido a inexistência de *blobs* nessas sequências.

Para a sequência de imagens analisada obteve-se uma média de processamento de 243_{ms}, concluindo-se que não é possível com o sistema proposto e hardware disponível a construção em tempo real. Com os dados obtidos o máximo, para que o sistema funcione em tempo real o número máximo de imagens de captura de cada câmara é de 4.

Do bloco de Detecção de trajetórias a componente de Detecção de *Blobs* é a que tem um maior peso computacional.

Para se conseguir melhorar a velocidade de processamento, mantendo o mesmo tipo de hardware existem várias soluções.

Baixar a resolução de câmaras, o que permite uma maior rapidez na obtenção das sequências das imagens e segmentação. Esta solução não melhora o módulo da Correspondência de *blobs* e Construção de trajetórias, pois não dependem da resolução da imagem.

Efetuar processamento distribuído. Esta solução seria a mais viável, pois o módulo de Detecção de trajetórias é independente para cada câmara, porque as trajetórias são construídas só com a informação da imagem. Esta opção permite igualmente não degradar o sistema, com o aumento do número de câmaras.

Na Tabela 8: *Ground truth* apresentam-se os resultados finais de Detecção e Relação de trajetórias devolvidas pelo sistema, comparadas pela análise manual.

Para facilidade de análise e rapidez do processo, dividiu-se a sequência total de imagens em 25 subsequências independentes.



5.2 Ground Truth

Tabela 8: Ground truth

Sequences	Merges				Splits				Connections in one chamber				Matching between cameras				Connections between cameras			
	GT	V	F ₊	F ₋	GT	V	F ₊	F ₋	GT	V	F ₊	F ₋	GT	V	F ₊	F ₋	GT	V	F ₊	F ₋
1 (700- 900)	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	2	2	0	0
2 (990-1153)	2	2	0	0	2	1	0	1	0	0	0	0	9	8	0	1	0	0	0	0
3 (1272-1457)	2	2	0	0	1	0	0	1	0	0	0	0	1	1	0	0	6	6	0	0
4 (1569-1716)	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
5 (1838-2083)	2	2	0	0	2	2	0	0	2	2	0	0	10	10	0	0	7	7	0	0
6 (2222-2383)	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
7 (2511-2688)	0	0	0	0	0	0	0	0	3	3	0	0	2	2	0	0	2	2	0	0
8 (2882-3175)	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	2	2	0	0
9 (3227-4057)	1	1	0	0	1	1	0	0	1	1	0	0	3	3	0	0	9	8	0	1
10 (4205-4481)	1	1	0	0	0	0	0	0	3	3	0	0	0	0	0	0	1	1	0	0
11 (4545-4623)	3	3	0	0	0	0	0	0	2	2	0	0	5	5	0	0	6	6	0	0
12 (4646-4766)	0	0	0	0	0	0	0	0	4	4	0	0	1	1	0	0	2	2	0	0
13 (4820-5164)	1	1	0	0	0	0	0	0	4	4	0	0	0	0	0	0	0	0	0	0
14 (5172-5312)	4	4	0	0	3	2	0	1	0	0	0	0	11	10	0	1	5	4	0	1
15 (5335-5753)	6	6	0	0	4	3	0	1	7	7	0	0	8	8	0	0	8	8	0	0
16 (5777-6052)	1	1	0	0	1	1	0	0	1	1	0	0	5	5	0	0	1	1	0	0
17 (6064-7900)	12	13	1	0	9	9	2	2	35	35	0	0	23	23	0	0	24	24	0	0
18 (7901-8500)	5	5	0	0	4	3	0	1	16	17	1	0	9	9	0	0	14	18	4	0
19 (8590-9927)	0	0	0	0	0	0	0	0	2	2	0	0	9	9	0	0	6	6	0	0
20 (10095-10244)	3	5	2	0	4	5	2	1	1	1	0	0	0	0	0	0	8	9	1	0
21 (10249-11050)	3	0	0	3	3	3	0	0	3	4	1	0	12	11	0	1	7	7	0	0
22 (11164-11322)	5	5	0	0	2	2	0	0	2	2	0	0	7	6	0	1	3	3	0	0
23 (11326-11502)	3	4	1	0	3	3	1	1	0	0	0	0	8	8	0	0	0	0	0	0
24 (11691-12210)	0	0	0	0	0	0	0	0	0	0	0	0	2	2	0	0	2	2	0	0
25 (12310-12690)	4	4	0	0	3	2	0	1	6	6	0	0	12	12	0	0	15	14	0	1
	58	59	4	3	42	37	5	10	92	94	2	0	141	137	0	4	130	132	5	3



A Tabela 8 dá informações sobre a fiabilidade do nosso sistema, em que GT representa as relações existentes no mundo, V as ligações que o sistema detetou e são verdadeiras, F_- representa as ligações existentes no mundo mas que o sistema não detetou e por fim, F_+ representa as relações que o sistema detetou mas que não existem no mundo. A equação 15 mostra a relação entre o *ground truth* com as deteções e os falsos positivos e negativos.

$$GT = V + (F_-) - (F_+) \quad \text{Equação 15}$$

Tabela 9: Total de deteções

<i>Ground Truth</i>	463	
Detetadas	459	
Detetadas falsas	16	36
Falhas deteção	20	

A Tabela 9 permite calcular as taxas de erro e a fiabilidade do sistema

Taxa de erro de falsas deteções (F_+) $36/463 \cong 8\%$
Taxa de erro de relação não detetadas (F_-) $1/463 \cong 8\%$

5.3 Análise de resultados

São mostrados de seguida alguns exemplos de falhas no sistema.

➔ Sequência 2 (990-1153)



a) Câmara 2 T=1079



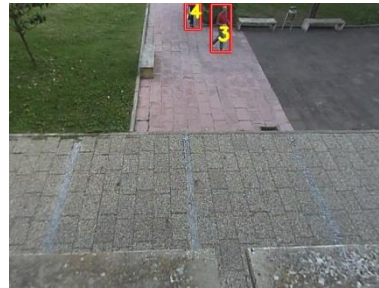
b) Câmara 2 T=1080

Figura 53: Falha de deteção de *Split*

Na Figura 53, existe uma falha de deteção de *Split* da trajetória 2, devido a não ter havido correspondência entre o *blob* associado à trajetória 2 da Figura 53.a e os *blobs*



associados às trajetória 2 e 3 da Figura 53.b. Esta situação provoca um acréscimo de velocidade do *blob* associado à trajetória 2 da Figura 53.b devido à variação brusca do centro de massa, fazendo com que na imagem seguinte o seu valor predito seja incorreto. Como na matriz de correspondência são usados os valores preditos das *bounding boxes*, em vez de ocorrer um *Matching* ocorre um *Leave* e *New* (Figura 54) provocando o fim da trajetória 2 e o começo da trajetória 4.



(Câmara 2 T=1081)

Figura 54: Falso *New*

→ Sequência 9 (3306-4057)



a) Câmara 1 T=3941



b) Câmara 3 T=4037

Figura 55: Falha de detecção de relações entre câmaras

Na Figura 55, existe uma falha de detecção de relação de trajetórias entre câmaras.

A pessoa com a trajetória 6 na câmara 1 (Figura 55.a) é a mesma da trajetória 7 da câmara 3 (Figura 55.b) O motivo desta falha é a mudança de direção da pessoa entre o percurso da câmara 1 e o da câmara 3, falhando assim a previsão.



→ Sequência 25 (12310-12690)



a) Câmara 2 T=12358



b) Câmara 2 T=12381

Figura 56: Falsa ligação numa câmara

Na Figura 56 verifica-se uma falsa ligação entre a trajetória 3 e a trajetória 9, isto é devido ao algoritmo prever a posição do *blob* correspondente ao fim trajetória 3 nessa região. Para contornar esta situação além da posição prevista torna-se necessário acrescentar mais características na correspondência.



Conclusões e Trabalho Futuro

6.1 Conclusões

Obtiveram-se resultados satisfatórios na deteção e relação de trajetórias.

Verificaram-se algumas falhas na deteção de *Merges* e *Split*, que se deveram essencialmente ao facto de haver falhas na captura das imagens, provocando “saltos”.

Na relação de trajetórias, as ligações de trajetória entre câmaras, devido a se considerar somente o critério do deslocamento não permite a mudança de direção e velocidade quando uma pessoa passa por uma zona não coberta pelos campos de visão das câmaras.

Dos resultados obtidos conclui-se que com o equipamento e métodos utilizados o máximo de captura de imagens para se poder fazer o seguimento em tempo real é de 4 imagens por segundo.



6.2 Trabalho Futuro

Neste trabalho abordou-se o seguimento de pessoas até à geração e relação de trajetórias.

Possibilidade de geração de grafos intermédios com o sistema em funcionamento.

Em trabalhos futuros existe o desafio de a partir dessas trajetórias e das relações entre as trajetórias conseguir inferir, qual a probabilidade de uma pessoa passar em cada trajetória. Sabendo a probabilidade em cada trajetória infere-se de seguida qual a trajetória mais provável que determinada pessoa efetuou.

Desenvolvimento de sistemas distribuídos em videovigilância para permitir um processamento mais rápido, em que cada câmara teria uma máquina dedicada.

Desenvolvimento de sistemas de câmaras *3D*.



Bibliografia

- [1] X.-Y. Xiong and B. Wang, "Study on forest fires recognition and moving target tracking in video surveillance system," in *2011 International Conference on Electronic & Mechanical Engineering and Information Technology*, Harbin, Heilongjiang, China, Aug. 2011 .
- [2] D. Bloisi, L. Iocchi, D. Nardi, M. Fiorini and G. Graziano, "Ground traffic surveillance system for Air Traffic control," in *ITS Telecommunications (ITST), 2012 12th International Conference on*, Taipei, Taiwan, Nov. 2012.
- [3] Z. Jinglei and L. Zhengguang, "A Vision-Based Road Surveillance System Using Improved Background Subtraction and Region Growing Approach," in *Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2007. SNPDP 2007. Eighth ACIS International Conference on*, Qingdao, China, Aug. 2007.
- [4] Z. Jia, H. Wang and R. Cabalero, "A two-step approach to see-through bad weather for surveillance video quality enhancement," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, Shanghai, China, May 2011.
- [5] C. Jaynes, A. Kale, ., N. Sanders and E. Grossmann, "The Terrascope dataset: scripted multi-camera indoor video surveillance with ground-truth," in *Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005. 2nd Joint IEEE International Workshop on*, Beijing, 2005.
- [6] R. Cucchiara, "Computer vision system for in-house video," in *Vision, Image and Signal Processing, IEE Proceedings -*, april 2005.
- [7] E. L. Hall, "Measuring Curved Surfaces for Robot Vision," in *Computer*, University of Tennessee, 1982.
- [8] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," in *IEEE JOURNAL OF ROBOTICS AND AUTOMATION*, 1987.
- [9] C. Shanshan, Z. Wuheng and Z. Lijun, "Camera Calibration via Stereo Vision Using Tsai's Method," in *Education Technology and Computer Science, 2009. ETCS '09. First International Workshop on*, Wuhan, Hubei, 2009.
- [] Y. Xin, Xin, Z. Ruishuang and L. Su, "A Calibration Method Based on



- 10] OpenCV," in *Intelligent Systems and Applications (ISA), 2011 3rd International Workshop on*, May 2011.
- [Z. Zhang, "A Flexible New Technique for Camera Calibration," in *Pattern*
- 11] *Analysis and Machine Intelligence, IEEE Transactions on*, 2000.
- ["Camera calibration using two or three vanishing points," in *Computer Science*
- 12] *and Information Systems (FedCSIS), 2012 Federated Conference on*, Wroclaw, Sept. 2012.
- [W. Kusakunniran, H. Li and J. Zhang, "A direct method to self-calibrate a
- 13] surveillance camera by observing a walking pedestrian," in *Digital Image Computing: Techniques and Applications, 2009. DICTA '09.*, Melbourne, VIC, 2009.
- [[Online]. Available: <http://opencv.org/>. [Accessed 2014].
- 14]
- [J.-Y. Bouguet, "Camera Calibration Toolbox for Matlab," 2 Dezembro 2013.
- 15] [Online]. Available: http://www.vision.caltech.edu/bouguetj/calib_doc/. [Accessed 21 Julho 2014].
- [R. K. Behera, P. Kharade, S. Yerva, P. Dhane, A. Jain and K. Kutty, "Multi-
- 16] Camera Based Surveillance System," in *Information and Communication Technologies (WICT), 2012 World Congress on* , Trivandrum , Nov 2012 .
- [T. Sasaki and H. Hashimoto, "Camera Calibration using Mobile Robot in
- 17] Intelligent Space," in *SICE-ICASE International Joint Conference 2006*, Korea, 2006.
- [S. Yu, X. Chen, W. Sun and D. Xie, "A robust method for detecting and
- 18] counting people," in *Audio, Language and Image Processing, 2008. ICALIP 2008. International Conference on*, July 2008.
- [P. Kumar, A. Dick and T. S. Sheng, "Real Time Target Tracking with Pan Tilt
- 19] Zoom Camera," in *Digital Image Computing: Techniques and Applications, 2009. DICTA '09.*, Melbourne, VIC, 2009.
- [F. Wang and S. Dai, "Adaptive background update based on mixture models of
- 20] Gaussian," in *Information and Automation, 2009. ICIA '09. International Conference on*, China, June 2009.
- [C. Stauffer and W. Grimson, "Adaptive background mixture models for real-
- 21] time tracking," in *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on*, Fort Collins, CO, June 1999.
- [S. Mittal, T. Prasad, S. Saurabh, X. Fan and H. Shin, "Pedestrian Detection and
- 22] Tracking Using Deformable Part Models and Kalman Filtering," in *SoC Design Conference (ISOCC), 2012 International*, Jeju Island, Nov. 2012.
- [Z. Khan, T. Balch and F. Dellaert, "An MCMC-Based Particle Filter for
- 23] Tracking Multiple Interacting Targets," in *8th European Conference on Computer Vision*, Prague, Czech Republic, May 2004.

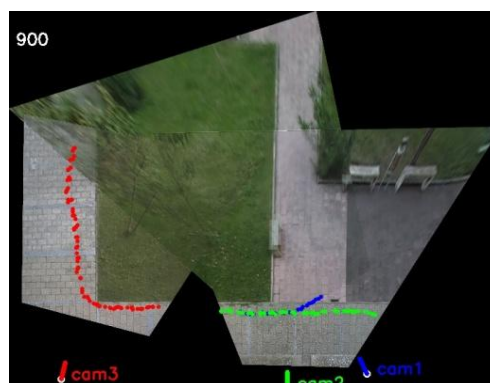
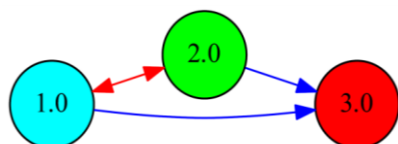


- [K. Nummiaro, E. Koller-Meier and L. V. Gool, "An adaptive color-based
24] particle filter," in *Image and Vision Computing 21 (2003)*, January 2003.
- [L. M. Fuentes and S. A. velastin, "People tracking in surveillance applications,"
25] in *Proceedings 2nd IEEE Int. Workshop on PETS, Kauai, Hawaii, USA,*, Hawaii,
USA, December 2001.
- [W. Nunziati, S. Sclarof and A. D. Bimbo, "Matching Trajectories between Video
26] Sequences by Exploiting a Sparse Projective Invariant Representation," in *IEEE
TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*,,
MARCH 2010.
- [P. M. Jorge, J. M. S. and A. J. Abrantes, "Estimation of the Bayesian network
27] architecture for object tracking in video sequences," in *Pattern Recognition, 2004.
ICPR 2004. Proceedings of the 17th International Conference on* , Aug. 2004.
- [R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, UK:
28] Cambridge University Press, 2004, p. 670.
- [google, "www.google.pt," [Online]. Available:
29] <https://www.google.pt/maps/@38.7557153,-9.1164587,75m/data=!3m1!1e3>.
[Accessed 11 Julho 2014].
- ["Graphviz - Graph Visualization Software," [Online]. Available:
30] <http://www.graphviz.org/>. [Accessed 09 07 2014].

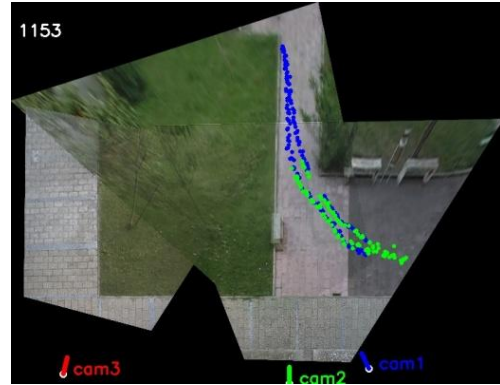
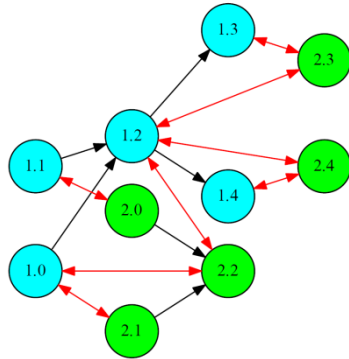


Anexos

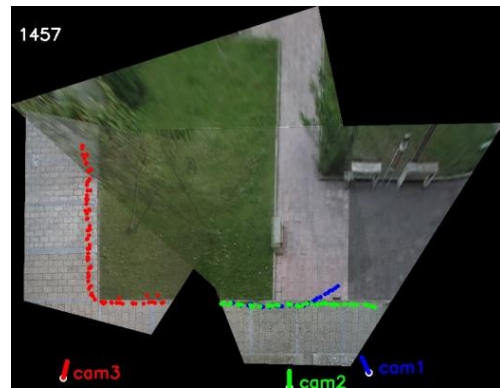
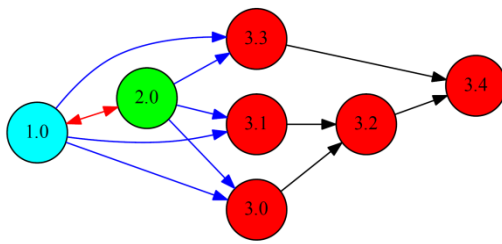
8.1 Grafos das trajetórias



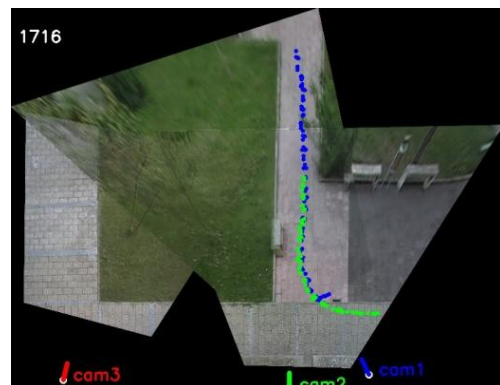
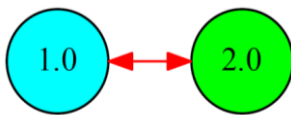
Grafo 1: Trajetórias entre a imagem 700 e a imagem 900



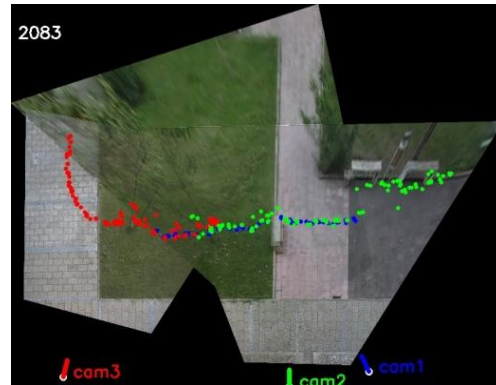
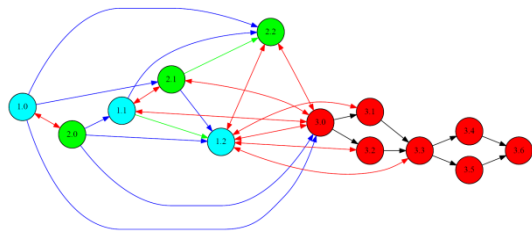
Grafo 2: Trajetórias entre a imagem 990 e a imagem 1153



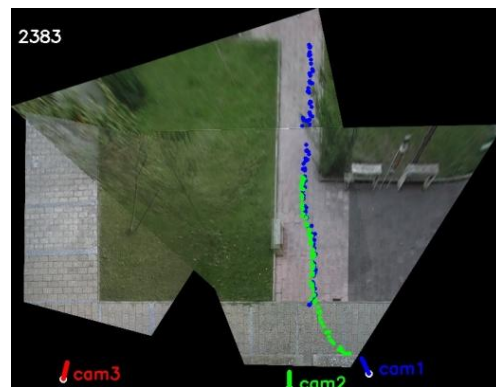
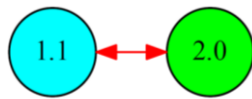
Grafo 3: Trajetórias entre a imagem 1272 e a imagem 1457



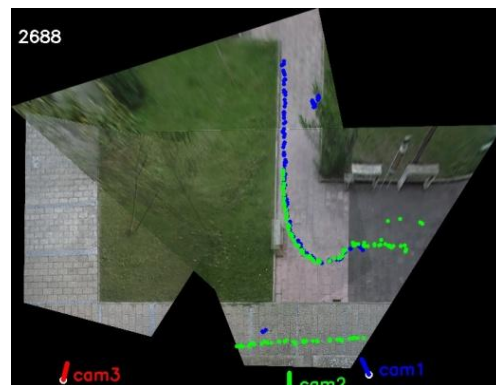
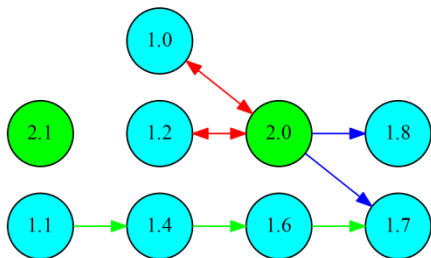
Grafo 4: Trajetórias entre a imagem 1569 e a imagem 1716



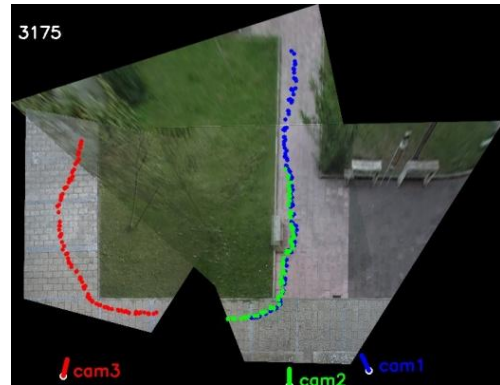
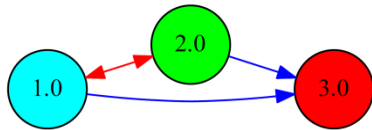
Grafo 5: Trajetórias entre a imagem 1838 e a imagem 2083



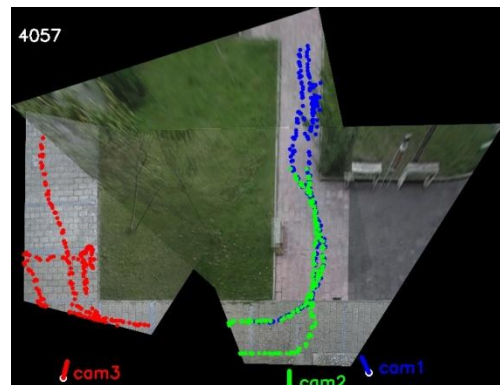
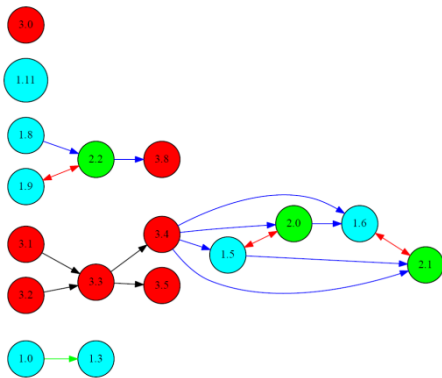
Grafo 6: Trajetórias entre a imagem 2222 e a imagem 2383



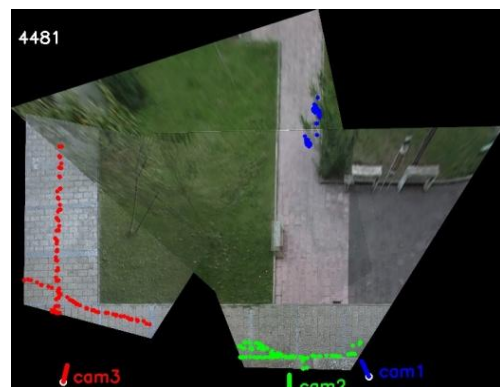
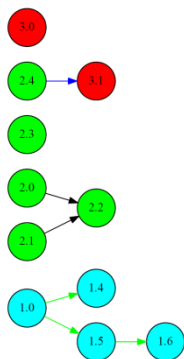
Grafo 7: Trajetórias entre a imagem 2511 e a imagem 2688



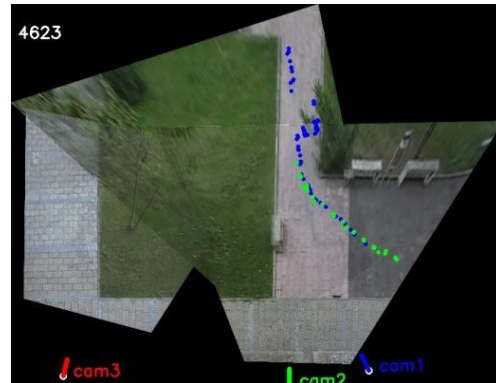
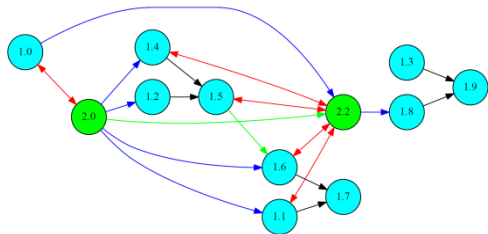
Grafo 8: Trajetórias entre a imagem 2882 e a imagem 3175



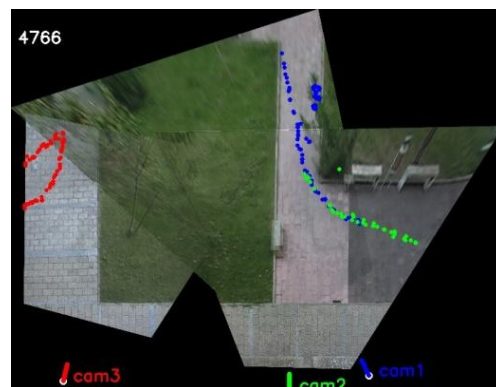
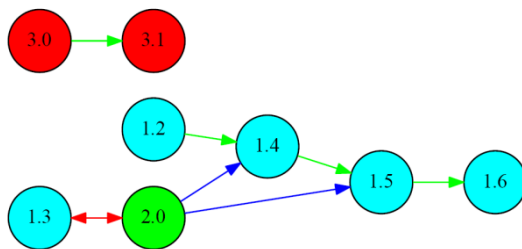
Grafo 9: Trajetórias entre a imagem 3227 e a imagem 4057



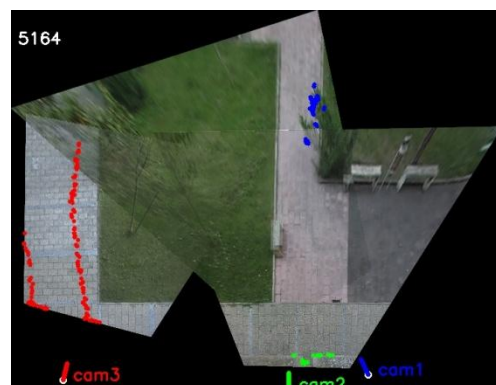
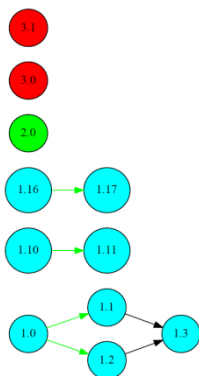
Grafo 10: Trajetórias entre a imagem 4205 e a imagem 4481



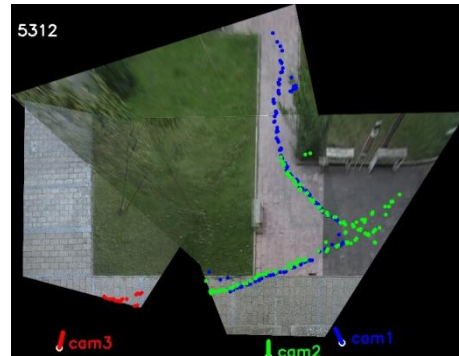
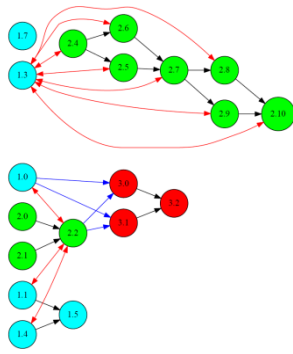
Grafo 11: Trajetórias entre a imagem 4545 e a imagem 4623



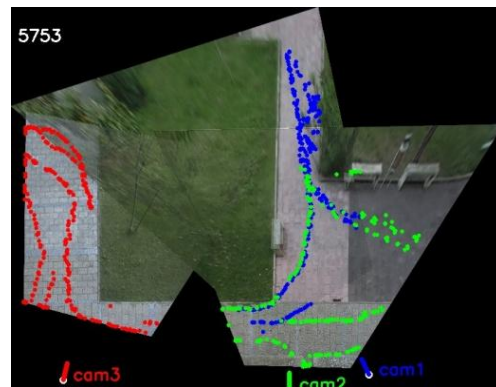
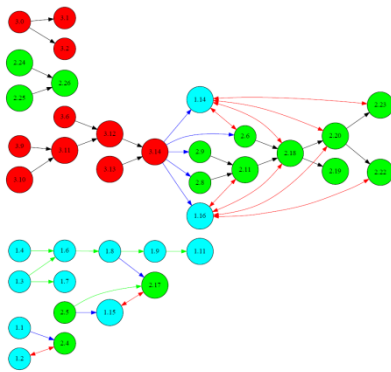
Grafo 12: Trajetórias entre a imagem 4646 e a imagem 4766



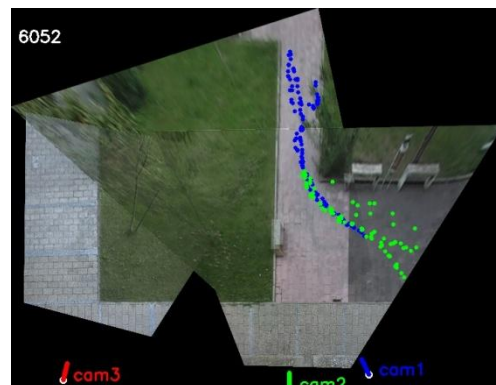
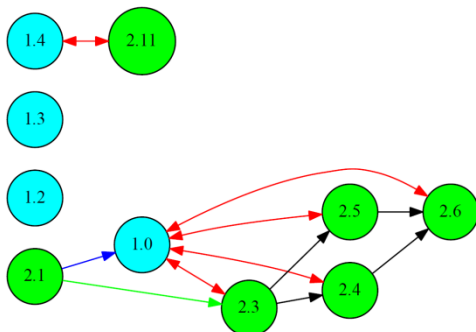
Grafo 13: Trajetórias entre a imagem 4820 e a imagem 5164



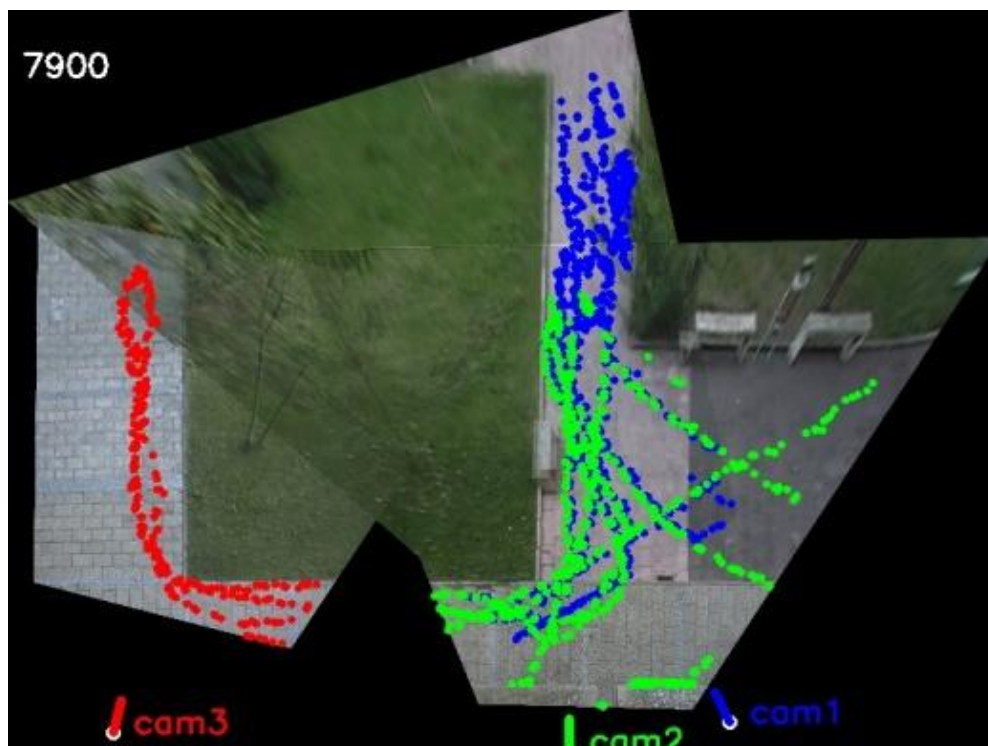
Grafo 14: Trajetórias entre a imagem 5172 e a imagem 5312



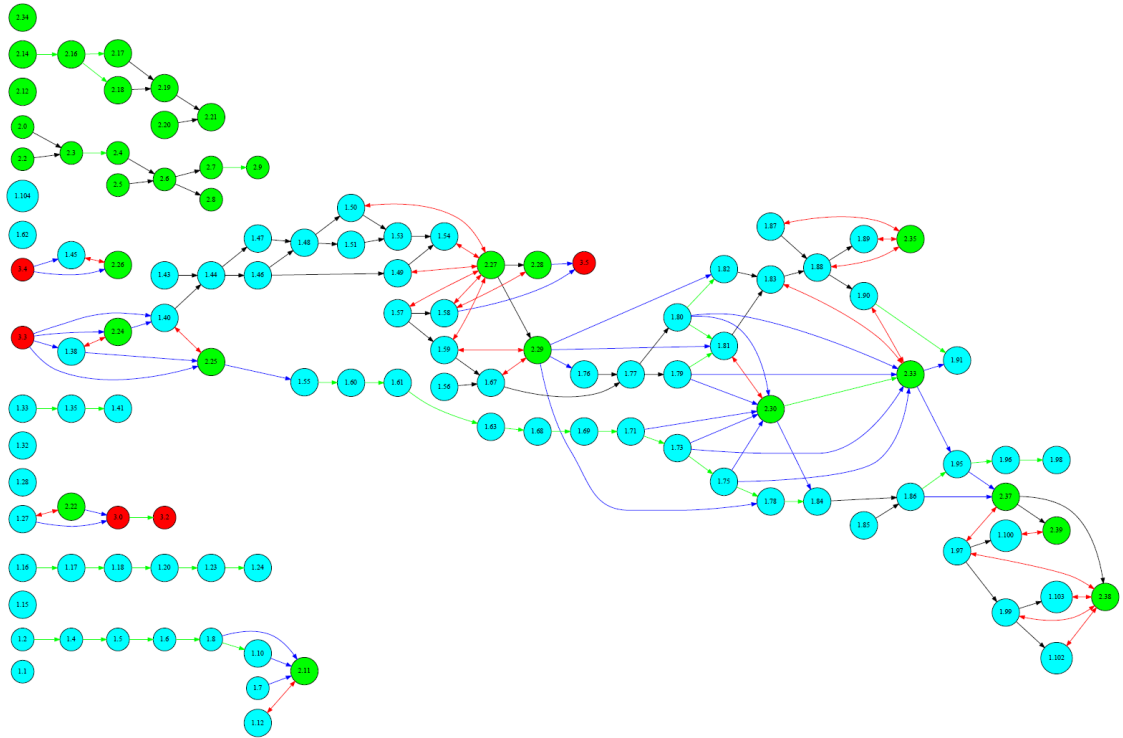
Grafo 15: Trajetórias entre a imagem 5335 e a imagem 5753



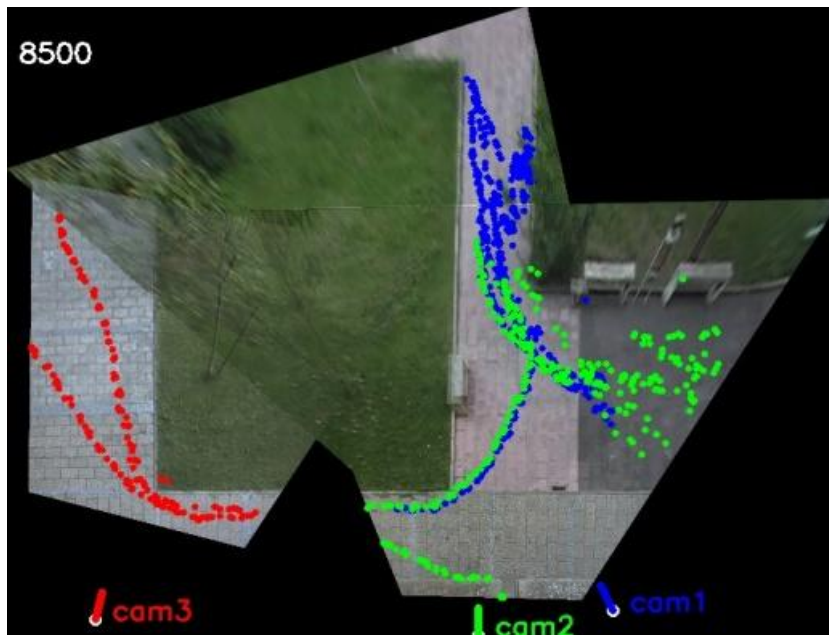
Grafo 16: Trajetórias entre a imagem 5777 e a imagem 6052



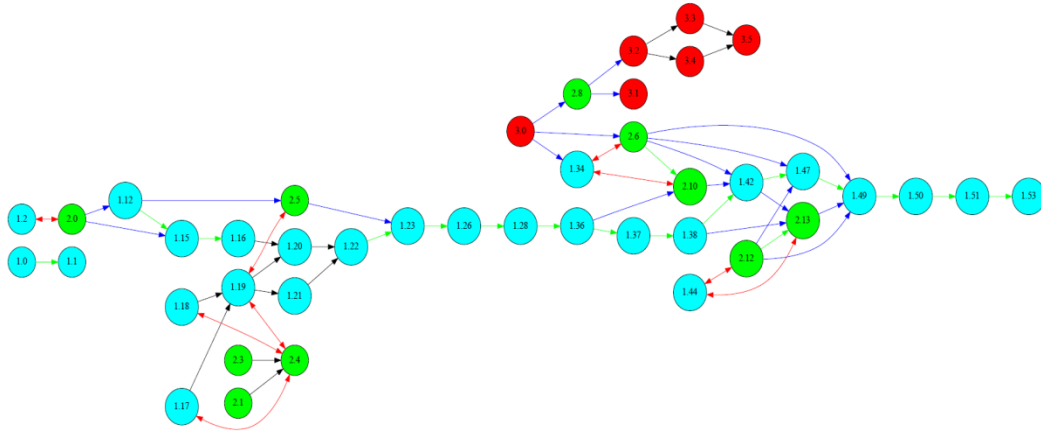
Trajatórias entre a imagem 6064 e a imagem 7900



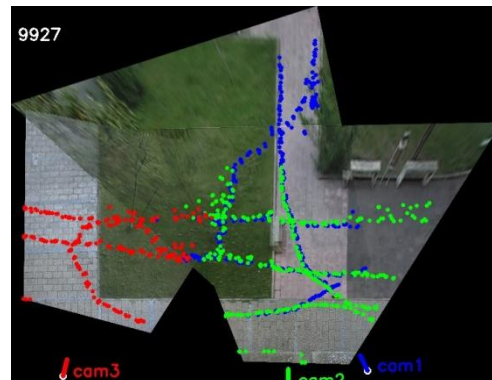
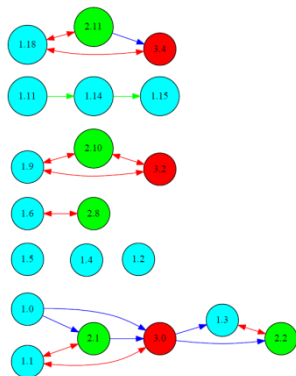
Grafo 17: Trajetórias entre a imagem 6064 e a imagem 7900



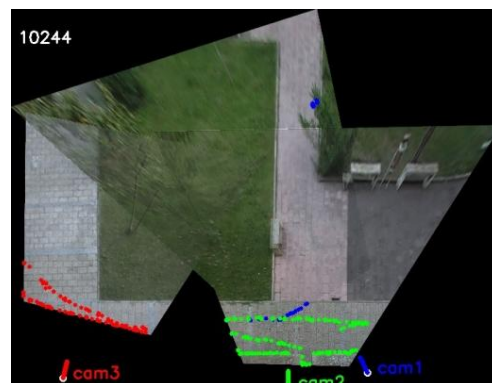
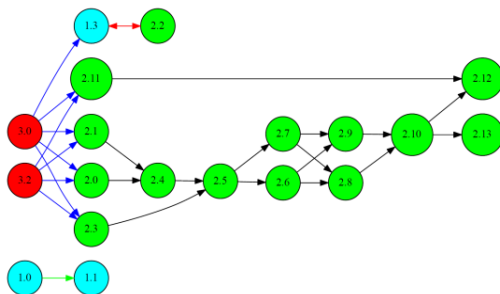
Trajetórias entre a imagem 7901 e a imagem 8500



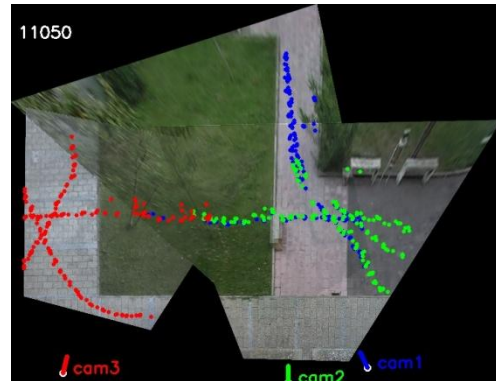
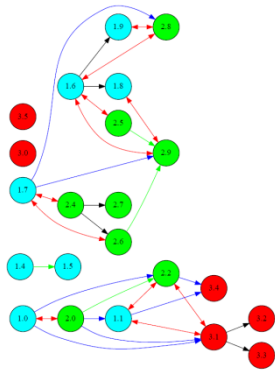
Grafo 18: Trajetórias entre a imagem 7901 e a imagem 8500



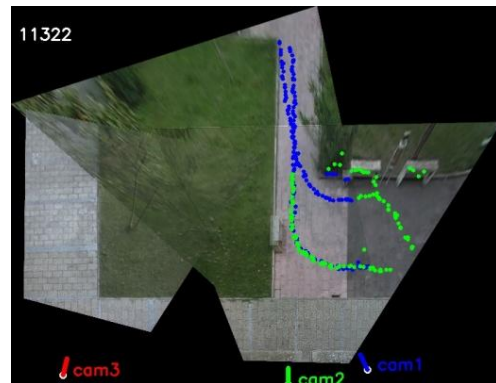
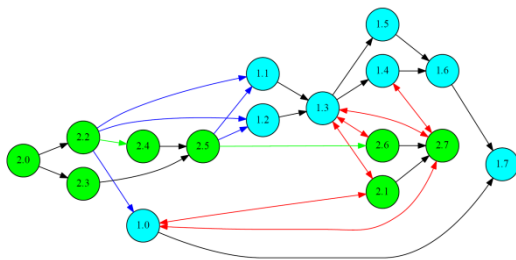
Grafo 19: Trajetórias entre a imagem 8590 e a imagem 9927



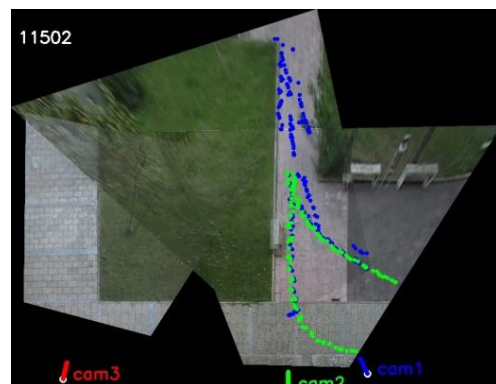
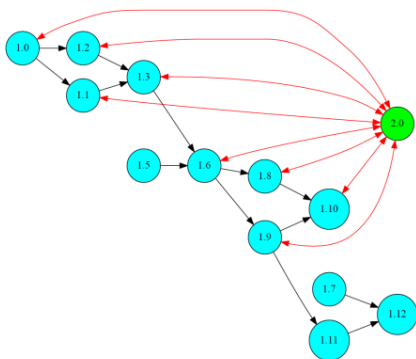
Grafo 20: Trajetórias entre a imagem 10095 e a imagem 10244



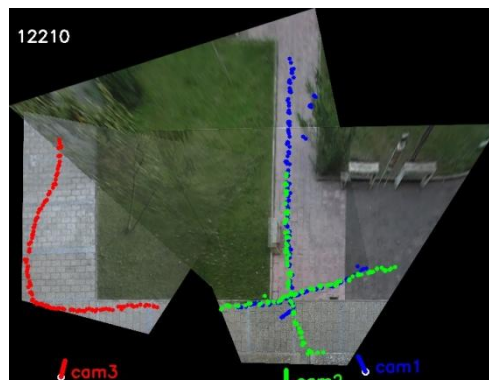
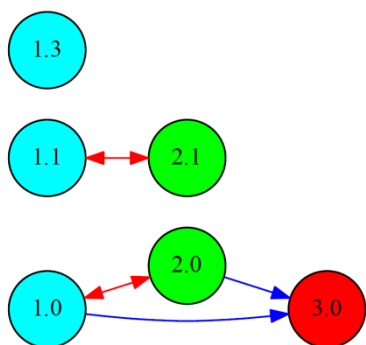
Grafo 21: Trajetórias entre a imagem 10249 e a imagem 11050



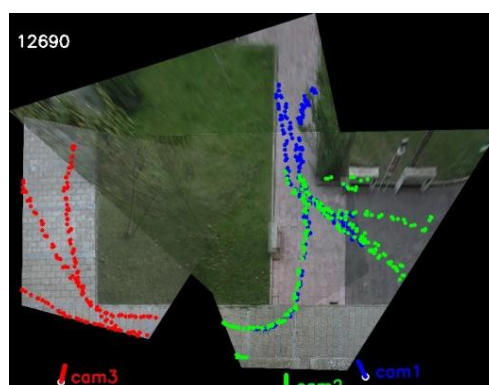
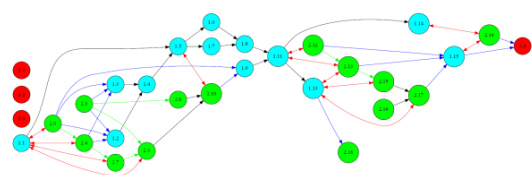
Grafo 22: Trajetórias entre a imagem 11164 e a imagem 11322



Grafo 23: Trajetórias entre a imagem 11326 e a imagem 11502



Grafo 24: Trajetórias entre a imagem 11691 e a imagem 12210

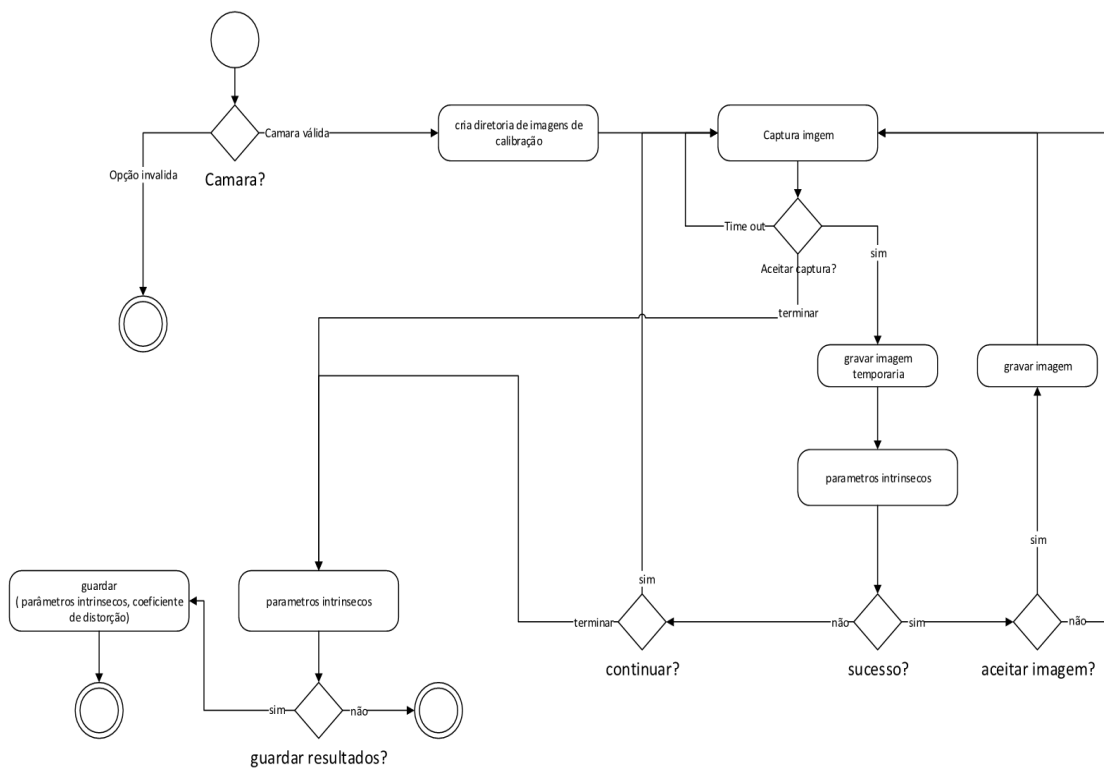


Grafo 25: Trajetórias entre a imagem 12310 e a imagem 12690



8.2 Diagramas UML

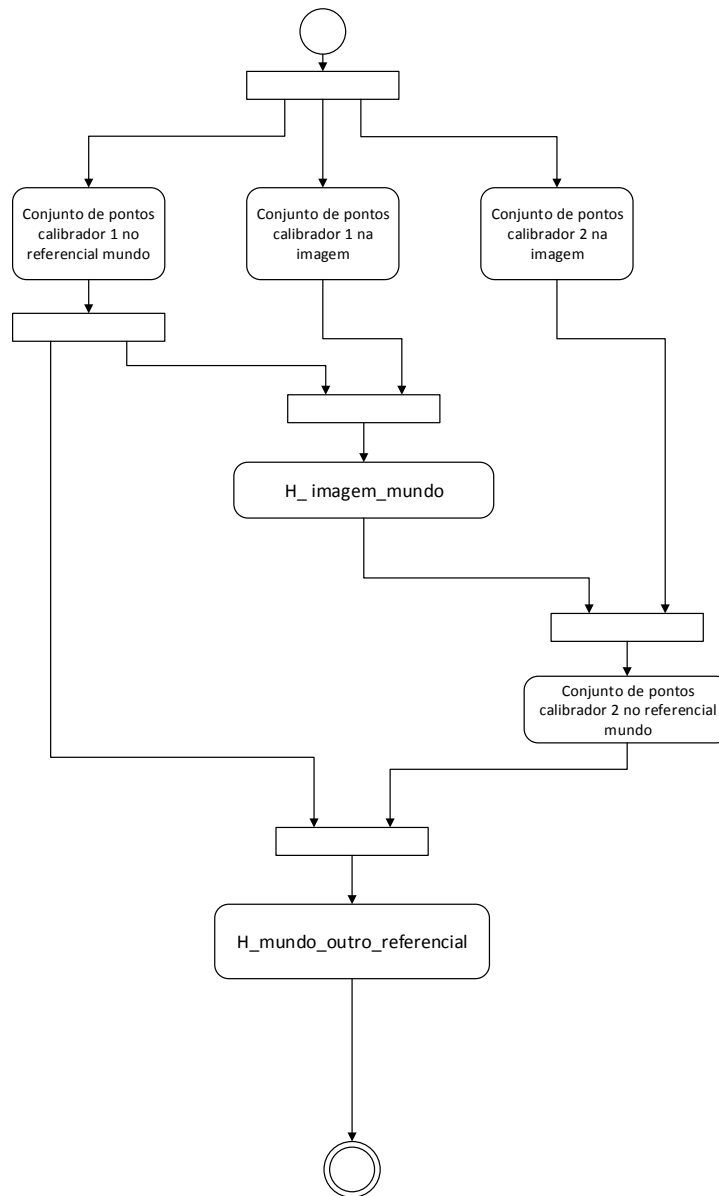
8.2.1 Cálculo dos parâmetros intrínsecos



UML 1: Cálculo dos parâmetros intrínsecos



8.2.2 Cálculo da relação entre referenciais



UML 2: Cálculo de relação entre referenciais