

PODER EXECUTIVO
MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO AMAZONAS
INSTITUTO DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

Alinhamento Espaço-Temporal em Sistemas Multissensoriais Heterogêneos

DIEGO DA SILVA FRONER

Manaus

2012

PODER EXECUTIVO
MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO AMAZONAS
INSTITUTO DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

DIEGO DA SILVA FRONER

Alinhamento Espaço-Temporal em Sistemas Multissensoriais Heterogêneos

Tese apresentada ao Programa de Pós-Graduação em Informática do Departamento de Ciência da Computação da Universidade Federal do Amazonas, como parte dos requisitos necessários para a obtenção do título de *Mestre em Informática*, área de concentração em Visão Computacional.

Orientador: Prof. Dr. José Luiz de Souza Pio

Manaus

2012

Agradecimentos

Agradeço a todos os meus familiares, em especial ao meu pai Hory Froner e falecida mãe M. do P. Socorro que condicionaram todas as minhas conquistas. A minha futura esposa Giselle Menezes que está junto comigo desde o início deste processo de obtenção de título de mestre, que sempre me apoiou e incentivou nos momentos difíceis. A todos os meus irmãos e irmãs, tios e tias, sobrinhos e sobrinhas, primos e primas, e demais familiares que são a base desta família maravilhosa a qual me orgulho em pertencer. Ao meu cãozinho de estimação, o Billy, que está sempre transmitindo felicidade ao lar em que resido, e ao meu time do coração Grêmio Football Porto Alegrense por existir, sendo assim a razão de minha paixão futebolística.

Aos meus amigos Kaio Wagner, Guilherme Monteiro, Rafael Leite, Felipe Hummel, Rodrigo Braga, William Albuquerque, Vinicius Secatti, Bruno Campos, Felipe Gomes, Marco Nobre, Petrina Kimura, Maísa Vidal, Rafael Rodolfo, entre outros, os quais conheci durante a época de graduação e mestrado.

Um agradecimento especial ao meu orientador José Pio, que sempre foi compreensivo com os problemas enfrentados por mim durante o desenvolvimento deste trabalho. Ao colega Ícaro Oliveira pelo auxílio em questões que envolveram o problema abordado, aos professores José Pinheiro e Eulanda Miranda pelo auxílio em minha qualificação, ao professor Horácio por gentilmente ceder equipamentos utilizados nos experimentos e a todos os professores do DCC-UFAM.

Por último, um agradecimento aos órgãos de fomento FAPEAM e CNPq por apoiarem o desenvolvimento deste trabalho.

Sumário

1	Introdução.....	1
1.1	Motivação	2
1.2	Justificativa	3
1.3	Objetivos.....	4
1.3.1	Objetivos Gerais.....	4
1.3.2	Objetivos Específicos	4
1.4	Organização do Texto	5
2	Trabalhos Relacionados	6
3	Metodologia	11
3.1	Aquisição de Dados	12
3.2	Cálculos de Variação de Velocidade e Alinhamento Temporal.....	14
3.3	Alinhamento Espacial	16
4	Aplicações Experimentais.....	19
4.1	Experimentos em Ambiente Simulado.....	20
4.2	Experimentos em Laboratório	28
4.3	Experimentos em Condições Reais	36
5	Conclusões e Trabalhos Futuros	43
5.1	Limitações	44
5.2	Extensões.....	44
5.3	Trabalhos Futuros.....	44
5.4	Considerações Finais.....	45
	Referências Bibliográficas	46

Lista de Figuras

- 1.1. Do lado esquerdo é mostrado um ambiente com o monitoramento de duas câmeras (A e B) e a limitação dos seus campos de cobertura. Do lado direito são adicionados sensores de posicionamento (C, D, E, F, G, H e I) que juntamente com as câmeras cobrem uma área mais ampla. O monitoramento poderia ser feito com a utilização de mais câmeras, porém, o custo de processamento seria muito maior devido ao excesso de informação visual presente em uma amostra. Já os sensores de localização provêm apenas as informações necessárias3
- 2.1. Ambiguidades no alinhamento entre duas imagens. (a) e (b) mostram frames correspondentes no tempo em duas sequências de vídeo diferentes mostrando o mesmo objeto (bola) em movimento. Vários alinhamentos são possíveis neste caso como mostrado em (c). (d) e (e) mostram duas sequências do movimento da bola. Com isto, agora existe apenas um alinhamento possível das duas trajetórias da bola. Esta correlação define o alinhamento tanto em tempo quanto em espaço entre as duas sequências de vídeo (f). (Caspi, 2002).6
- 2.2. Exemplo de imagem obtida por uma câmera catadióptrica. Este tipo de sistema é caracterizado por uma câmera anexada a um espelho convexo. (Santos, 2006).....7
- 2.3. Correspondência entre pontos versus correspondência entre trajetórias. (a) e (b) mostram dois frames retirados de duas sequências onde aparecem objetos (marcados como A, B, C, D e E). (c) e (d) mostra a trajetória destes objetos ao longo do tempo. Analisando apenas imagens estáticas dificulta a determinação de correspondência na imagem. No entanto, trajetórias de pontos contêm propriedades adicionais que simplificam o problema da correspondência entre duas sequências. (Caspi, 2002).8

2.4.	(a) e (b) mostram dois frames correspondentes a mesma ação que ocorre em momentos e lugares diferentes. (c) e (d) mostram os pontos da cena onde a ação foi detectada ((c) em magenta e (d) em verde). (e) mostra a dispersão dos movimentos e (f) a sobreposição dos mesmos. Em (g) e (h) podemos ver o antes e o depois do alinhamento. (Ukrainitz and Irani, 2006).	9
2.5.	Disposição dos sensores em um veículo automotivo utilizado para experimentação	10
3.1.	Fluxograma da metodologia proposta. As imagens e o posicionamento sensorial são adquiridos, em seguida são extraídas as variações de velocidade. Com os dados de aceleração determina-se o alinhamento temporal e, por último, o alinhamento espacial por uma matriz homografia.	11
3.2.	Os sensores A e B referem-se as câmeras enquanto que C aos campos de coberturas somados de todos os sensores de localização. Todos os sensores que pertencem ao sistema possuem campos de sobreposição entre si.....	12
3.3.	Em (a) é mostrada imagem real capturada de um veículo em movimento. Após o rastreamento dos centroides do veículo ao longo dos quadros que compõem a sequência, (b) mostra os pontos encontrados.....	13
4.1.	Esquematização da área de cobertura no modelo de simulação. O objeto monitorado movimenta-se pelo plano com probabilidades de seguir em cada direção. Todos os sensores possuem um campo de cobertura próprio	20
4.2.	Aceleração do ponto de vista de câmeras e do ponto de vista de sensores. Apesar de as câmeras capturarem a informação com uma frequência maior, seu comportamento é similar ao dos sensores ao longo do tempo por terem sido eliminadas as ambiguidades de observação.....	22
4.3.	Durante 50 iterações, a simulação gera um passeio aleatório com velocidade máxima de 10, aplica uma diferença entre o início de	

	aquisição de dados de cada sensor e verifica se o deslocamento encontrado foi similar ao aplicado. Repetindo 10 vezes este experimento, obteve-se uma média de 94% de acerto.	23
4.4.	Gráfico de comparação entre a variação da taxa de acertos com a variação de deslocamento. A medida que aumenta-se o valor máximo que o objeto pode deslocar-se, a precisão sofre um decréscimo.....	24
4.5.	O gráfico mostra a velocidade ao longo dos frames que compõem o vídeo em que foram obtidos os centroides de um objeto em deslocamento. A média de velocidade foi de 3.18.	25
4.6.	O valor de deslocamento que mais se aproxima da média obtida no teste de laboratório foi com valor 6, resultado obtido computando a velocidade média do centroide ao longo do testes simulados.	26
4.7.	A figura mostra como a taxa de acertos variou ao longo da repetição do experimento, oscilando em torno de 99% com os dados sendo computados a cada 100 simulações.....	27
4.8.	O eixo y contém a taxa de acertos e o eixo x a variação de γ , a variável de erros. O índice varia entre 98% e 99% até $\gamma=50$ onde, a partir deste ponto, a precisão sofre um declínio para aproximadamente 94%.....	28
4.9.	Visões do mesmo objeto em repouso em câmeras diferentes.....	29
4.10.	Sequência de frames que mostram a localização dos centroides detectados nos frames anteriores (os frames acima são de apenas uma câmera).....	29
4.11.	Último frame de cada vídeo com a marcação do caminho percorrido pelo objeto de acordo com seu centroide.	30
4.12.	Duas imagens geradas por câmeras diferentes alinhadas no tempo de acordo com o algoritmo proposto para alinhamento temporal.....	31
4.13.	As figuras a, b e c referem-se respectivamente aos combinados frames 339, 589 e 800 da câmera um e 200, 450 e 661 da câmera dois. Em cada uma são mostrados os pontos correspondentes encontrados pelo algoritmo RANSAC, as duas imagens sobrepostas alinhadas no espaço e a matriz homografia H com seus coeficientes.	33

- 4.14. A figura (a) mostra o frame 339 pertencente à Sequência um lado-a-lado com o frame 200 da Sequência dois sem acréscimo de ruído e o resultado encontrado. As figuras (b) e (c) mostram as mesmas imagens com acréscimo de ruído “sal e pimenta”, comparando as diferenças entre o resultado sem aplicação de filtro e com aplicação de filtro da mediana. As figuras (d) e (e) seguem as mesmas condições das (b) e (c), mas com aplicação de ruído gaussiano.36
- 4.15. As duas imagens mais acima mostram a perspectiva das câmeras dispostas lado-a-lado capturando o movimento de um veículo. As imagens abaixo mostram a rota percorrida de acordo com os dados capturados do GPS, utilizando como plano de fundo o Google Maps e Google Earth respectivamente.37
- 4.16. Comparativo entre as acelerações de S_1 , S_2 e S_3 . As amostras providas de câmeras possuem visualmente um comportamento similar, enquanto as providas de GPS são consideravelmente diferentes devido aos dados interpolados que, dependendo do intervalo entre a aquisição das posições, estão mais suscetíveis a erros.38
- 4.17. A Figura acima mostra respectivamente o frame 557 de S_1 e 540 de S_2 , os pontos correspondentes encontrados e as imagens sobrepostas alinhadas no espaço. Por último, a matriz homografia com seus coeficientes.40
- 4.18. As rotas sobrepostas mostram que o caminho percorrido ao longo de cada experimento foi similar. A variação ocorreu na sua taxa de amostragem.41
- 4.19. O eixo y mostra com quantos segundos após o início da aquisição dos dados cada posição foi adquirida. O eixo x contém a ordem (primeira, segunda, terceira, etc.) em que cada dado foi adquirido.42

Resumo

Este trabalho apresenta a utilização de diferentes sensores no aprimoramento da informação necessária para realizar o alinhamento espaço-temporal de imagens sequenciais. As propostas existentes chamadas *feature-based* utilizam-se da dinâmica da cena como maior indicador de que eventos estão ocorrendo simultaneamente no tempo, e assim posteriormente indicando suas posições relativas no espaço. Adicionando sensores de movimentação a um sistema com múltiplas câmeras de vídeo que possuam sobreposição de campos de cobertura, é possível adquirir informações de posicionamento dos objetos monitorados, servindo assim de auxílio para o alinhamento entre as imagens.

Palavras-chave: alinhamento espaço-temporal, integração de dados.

Abstract

This work presents the use of different sensors improving the information to perform spatio-temporal alignment of sequential images. The existing proposals called feature-based uses the dynamics of scenes as a major indicator that simultaneously events are occurring at the same time, and lately indicating their relative position in space. Adding motion sensors to a multiple video cameras system with overlapping fields of coverage, it's possible to acquire information about positions of the monitored objects that serves to aid the alignment between images.

Keywords: spatio-temporal alignment, data matching.

Capítulo 1

Introdução

O uso de múltiplas câmeras tem se difundido largamente nos últimos tempos devido ao rápido avanço da capacidade de armazenamento e de processamento de computadores. Já é possível encontrar vasta literatura especializada como trabalhos de Pádua (2005), Shakil (2005), Caspi and Irani (2000, 2001, 2002a, 2002b), Choi (2008), Ukrainitz (2008), Stein (1998), Wolf and Zomet (2002), Brito et al (2008), Vedula et al (2002), Hu et al (2004), Moeslund et al (2006), Zelnik-Manor and Irani (2001), Rao et AL (2003), Santos (2006), Munoz-Salinas (2007), Bertozzi et AL (2007) que são referências ao uso de múltiplas câmeras e obtenção de visão tridimensional através da stereoscopia. Contudo, agregando sensores heterogêneos a este tipo de sistema, são gerados diferentes tipos de dados, que, se trabalhados através de técnicas de fusão de dados (Nakamura (2005,2007), Loureiro (2002), Hall and Llinas (1997,2001)), podem ser utilizados como auxílio às informações visuais (Caspi (2001), Bertozzi (2007)).

Este trabalho apresenta o desenvolvimento de uma metodologia para o alinhamento espaço-temporal em uma rede heterogênea composta por câmeras e sensores. A questão que define o problema é a seguinte: Como sincronizar no tempo e no espaço informações de sensores visuais utilizando como referência dados obtidos por sensores não visuais? Tipicamente, o desalinhamento espacial ocorre devido às diversas posições em que os sensores estão localizados, nunca gerando a mesma informação respectiva ao ambiente monitorado, e, no caso dos sensores visuais, diferentes parâmetros com respeito à calibração de câmera e variáveis do ambiente. Já a falta de sincronismo temporal ocorre principalmente por duas razões: o tempo de resposta dos sensores (frames por segundo, delay no intervalo entre o envio de sinais, etc.) e as suas ativações que geralmente não ocorrem simultaneamente.

Os trabalhos existentes podem ser divididos em dois grupos principais: *feature-based* (Rao (2003), Brito (2008), Pádua (2005), Santos (2006), Wolf and Zomet (2002))

que consideram como principal característica de alinhamento a trajetória dos objetos, e *direct-based* (Caspi and Irani (2000,2001), Choi (2008), Ukrainitz (2006)) que extraem informações relativas a intensidade dos *pixels* pertencentes as regiões de sobreposição.

A metodologia deste trabalho se enquadra no grupo *feature-based*. Portanto, por considerar a trajetória dos objetos como principal característica, a hipótese principal é de que a utilização de outros sensores, além dos visuais, melhora a informação espaço-temporal.

Para alinhamento espacial, etapa que consiste em encontrar os campos de sobreposição das imagens, foi utilizado o algoritmo SIFT-RANSAC, método amplamente utilizado em diversos trabalhos como Mainberger (2008), Wang Wei (2008), Martínez-Gómez (2010), Constantinopoulos (2011), Ding (2007).

1.1 Motivação

A presença de sensores em acontecimentos cotidianos, principalmente visuais (câmeras), pode ser observada de forma mais comum com o passar dos tempos. A popularização de equipamentos como câmeras digitais, notebooks e computadores com *webcams* integradas, *smartphones* com diversos tipos de sensores acoplados, etc., tornou possível com que praticamente qualquer acontecimento esteja sendo gravado de alguma forma.

A principal motivação para este trabalho é o desenvolvimento de uma metodologia que possa auxiliar sistemas de segurança de acontecimentos com grande repercussão como atentados terroristas, acidentes automobilísticos, eventos esportivos e gravações cinematográficas. A reconstrução de *timeline*, que é uma técnica comumente utilizada empregada para análise de eventos criminais, pode ser bastante complexa, levando em conta a qualidade dos dados obtidos e sua taxa de amostragem, sendo necessário o uso de computação na maioria dos casos para resolver tal problema de forma precisa e confiável.

A Figura 1.1 mostra como seria a disposição dos sensores em um ambiente com e sem sensores visuais.

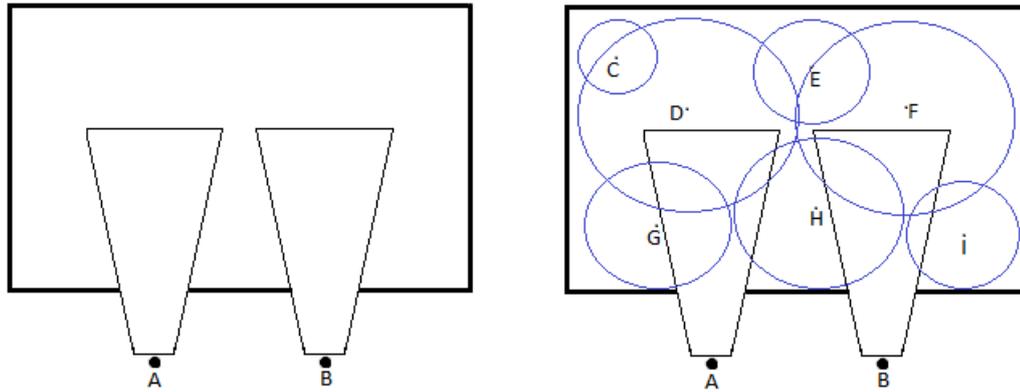


Figura 1.1: do lado esquerdo é mostrado um ambiente com o monitoramento de duas câmeras (A e B) e a limitação dos seus campos de cobertura. Do lado direito são adicionados sensores de posicionamento (C, D, E, F, G, H e I) que juntamente com as câmeras cobrem uma área mais ampla. O monitoramento poderia ser feito com a utilização de mais câmeras, porém, o custo de processamento seria muito maior devido ao excesso de informação visual presente em uma amostra. Já os sensores de localização provêm apenas as informações necessárias.

1.2 Justificativa

Diferentemente dos métodos já existentes para resolver tal problema esta abordagem utiliza um número maior de alternativas para melhorar a informação e controlar o ambiente. Para efeito de comparação, em um sistema composto unicamente por sensores visuais, é necessária a escolha dos dados providos por um destes para que seja usado como referencial aos demais. A idéia aqui desenvolvida é adicionar ao sistema um tipo diferente de sensor que tenha como principal objetivo obter dados a serem utilizados como referência.

Em relação ao ambiente monitorado, considera-se a disposição alguns dados como:

- Localização espacial conhecida (x,y,z) dos sensores não visuais.
- A frequência com que os dados são obtidos é constante;
- A resolução das imagens é conhecida;
- Existe sobreposição dos campos de cobertura;
- O tempo de ativação entre os sensores é desconhecido;
- Parâmetros intrínsecos e extrínsecos desconhecidos e invariáveis;

1.3 Objetivos

Os objetivos deste trabalho foram divididos em gerais e específicos.

Os objetivos gerais referem-se ao que será realizado para validar este trabalho e as expectativas quanto as suas aplicações. Já os objetivos específicos são os passos necessários para resolver cada etapa que envolve os objetivos gerais.

1.3.1 Objetivos Gerais

Desenvolver uma metodologia para alinhamento espaço-temporal de eventos em ambientes heterogêneos utilizando os sensores adicionais como referência para o alinhamento temporal entre câmeras e, conseqüentemente, o seu alinhamento espacial.

1.3.2 Objetivos Específicos

Como objetivos específicos, destacam-se os seguintes:

- Desenvolver uma metodologia capaz de utilizar como entrada dados heterogêneos;
- Montar um aparato experimental para testes controlados em laboratório;
- Analisar as variações que ocorrem nos dados obtidos por testes controlados quando aplicadas as técnicas de visão computacional utilizadas na metodologia;
- Simular situações com uma maior diversidade de sensores e perturbação variável nos dados;
- Validar o modelo por meio de um experimento real utilizando câmeras e um objeto rastreado por GPS.

1.4 Organização do Texto

O Capítulo dois apresenta o estado da arte, destacando as principais características dos trabalhos relacionados a este problema. A metodologia desenvolvida a partir deste estudo é descrita no Capítulo três, com os resultados dos experimentos sendo mostrados no Capítulo quatro. Por último, as conclusões são descritas no Capítulo cinco juntamente com sugestões de trabalhos futuros.

Capítulo 2

Trabalhos Relacionados

Em Shakil (2005) é feita uma revisão dos diversos métodos de alinhamento de vídeo existentes. Como simplificação do problema, o autor mostra que o problema se resume a alinhar frames correspondentes utilizando técnicas de alinhamento de imagens, mas, há casos em que apenas a correlação espacial não é informação suficiente como mostrado na Figura 2.1. A ideia de eliminar as ambigüidades espaciais em meio à utilização da movimentação de objetos na cena como fator determinante para que duas cenas distintas estejam acontecendo no mesmo instante, é utilizada na construção do modelo de alinhamento proposto por este trabalho.

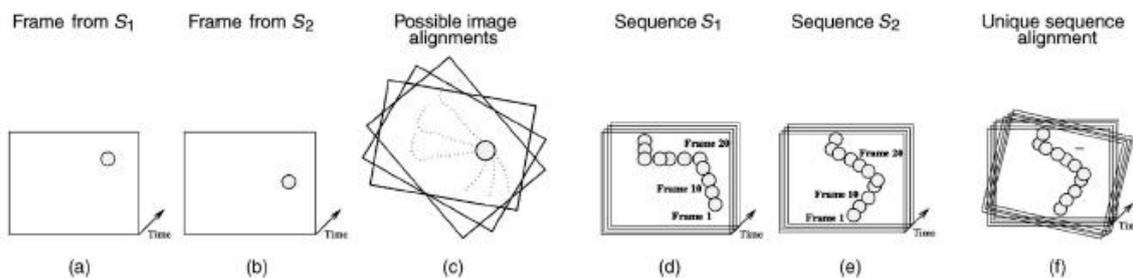


Figura 2.1: Ambigüidades no alinhamento entre duas imagens. (a) e (b) mostram frames correspondentes no tempo em duas seqüências de vídeo diferentes mostrando o mesmo objeto (bola) em movimento. Vários alinhamentos são possíveis neste caso como mostrado em (c). (d) e (e) mostram duas seqüências do movimento da bola. Com isto, agora existe apenas um alinhamento possível das duas trajetórias da bola. Esta correlação define o alinhamento tanto em tempo quanto em espaço entre as duas seqüências de vídeo (f). (Caspi, 2002).

Em Pádua (2005), o autor trabalha com casos em que diversas informações são desconhecidas como: taxas de quadros, geometria epipolar e parâmetros internos das câmeras. Nesta abordagem é feita uma procura em todo o espaço amostral para que sejam localizadas evidências de correlação espacial, o que exige um alto custo computacional. Santos (2006) faz uma extensão do trabalho de Pádua em que é feito o alinhamento espaço-temporal entre câmeras perspectivas e para-catadióptricas como

mostrado na Figura 2.2. Embora os trabalhos apresentem resultados satisfatórios, utiliza em sua abordagem apenas um tipo de sensor, o visual. A utilização de sensores heterogêneos permite que a área monitorada por estes sensores aumente consideravelmente, provendo assim uma solução com informações mais amplas a respeito do ambiente onde for aplicada esta metodologia.



Figura 2.2: exemplo de imagem obtida por uma câmera catadióptrica. Este tipo de sistema é caracterizado por uma câmera anexada a um espelho convexo. (Santos, 2006)

Em Caspi e Irani (2002), Stein (1998, 2000) e Rao (2003), o método proposto define que um par de seqüências não triviais (trajetórias em linha, por exemplo, podem ser ambíguas) é suficiente para encontrar o alinhamento de duas seqüências de imagens como mostrado na Figura 2.3. Contudo, é necessário fazer uma busca ao longo de toda a seqüência para a determinação das trajetórias que possuem a melhor correspondência. Tais trabalhos carecem de heurísticas que diminuam o processamento em conjuntos de dados muito extensos, pois a tendência em seus modelos é que o sistema perca desempenho consideravelmente a medida que cenas mais longas ou com maior taxa de captura sejam utilizadas.

Caspi (2001) considera o mesmo problema de desalinhamento, mas com câmeras que não possuem sobreposição. Em contrapartida, os seus centros de projeção devem ser aproximados e devem permanecer estáticas ao longo do

experimento tornando mais complexo o manuseio dos aparelhos utilizados em cena. Isto pode ser aprimorado utilizando-se diversos sensores heterogêneos onde, os visuais não possuem sobreposição entre si e os não visuais tratam de participar como fonte de correspondência entre estes. Sugerimos como extensão da proposta apresentada neste trabalho a experimentação em tais condições, partindo do pressuposto que a solução aqui apresentada pode ser também aplicada em sistemas com câmeras sem sobreposição.

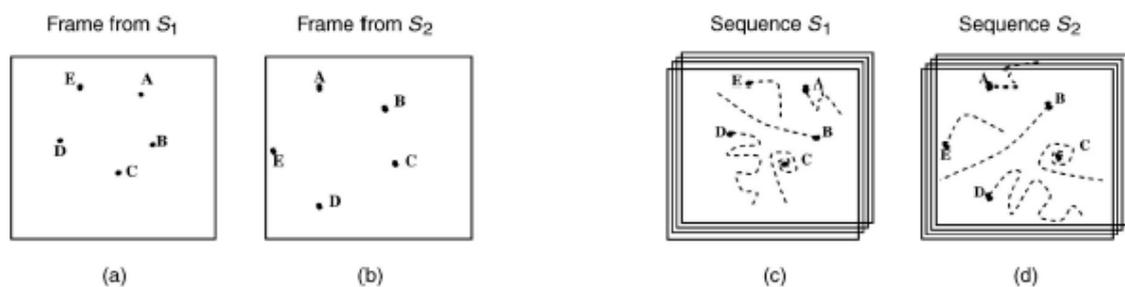


Figura 2.3: Correspondência entre pontos versus correspondência entre trajetórias. (a) e (b) mostram dois frames retirados de duas seqüências onde aparecem objetos (marcados como A, B, C, D e E). (c) e (d) mostra a trajetória destes objetos ao longo do tempo. Analisando apenas imagens estáticas dificulta a determinação de correspondência na imagem. No entanto, trajetórias de pontos contêm propriedades adicionais que simplificam o problema da correspondência entre duas seqüências. (Caspi, 2002).

Já Brito (2008), além de trabalhar com campos de visão disjuntos, considera que as câmeras não são necessariamente estacionárias ao longo do tempo, e sim, movem-se se necessário sendo ainda assim, por isso, suscetível a possíveis erros de movimentação caso ocorra este deslocamento. Além disso, o autor utiliza de um sensor móvel que se movimenta pela cena e em alguns momentos pelo campo de visão das câmeras, o que, apesar de sugerir a utilização de um sensor diferente além dos visuais, ainda assim é integralmente dependente destes. O sensor não visual utilizado não fornece dados diretos, sendo necessária a sua detecção através das câmeras.

Ukrainitz and Irani (2006) descreve um método utilizado para alinhamento de cenas capturadas em lugares diferentes e momentos diferentes, mas, com características semelhantes como mostrado na Figura 2.4. Porém, como outras informações passam a ser mais valiosas (fundo da cena, por exemplo) e não

unicamente a dinâmica dos objetos, o custo computacional também aumenta consideravelmente. Este trabalho reforça consideravelmente a ideia de que movimentos lineares tendem a causar ambiguidades, pois o trabalho propõe a solução de correlação entre cenas distintas onde a movimentação dos objetos presentes na cena são similares, apesar de ocorrerem em tempos distintos.

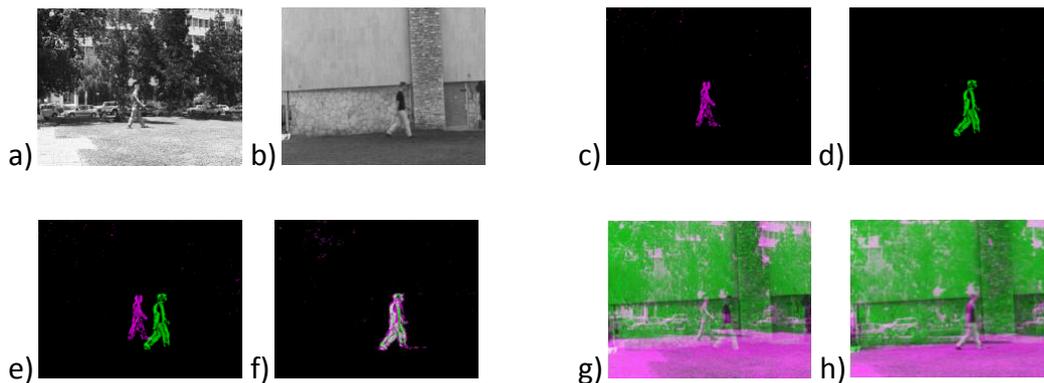


Figura 2.4: (a) e (b) mostram dois frames correspondentes a mesma ação que ocorre em momentos e lugares diferentes. (c) e (d) mostram os pontos da cena onde a ação foi detectada ((c) em magenta e (d) em verde). (e) mostra a dispersão dos movimentos e (f) a sobreposição dos mesmos. Em (g) e (h) podemos ver o antes e o depois do alinhamento. (Ukrainitz and Irani, 2006).

Em Mahlisch (2006), o autor destaca que em um veículo existem diversos sensores instalados, o que torna possível o desenvolvimento de sistemas de auxílio na direção. O problema neste caso encontra-se em alinhar estas informações, similar ao proposto neste trabalho em que sensores diferentes são utilizados, porém com a finalidade de gerar imagens sobrepostas originadas por câmeras diferentes. A Figura 2.5 mostra uma esquematização do aparato utilizado pelo autor para suas experimentações.



Figura 2.5: Disposição dos sensores em um veículo automotivo utilizado para experimentação.

Ledesma-Carbayo (2005) aborda um problema em alinhamento espaço-temporal de imagens obtidas a partir de ultrassons cardíacos, em que problemas no coração podem ser detectados a partir de seu comportamento no batimento cardíaco em diferentes regiões do miocárdio. Este problema utiliza exclusivamente imagens obtidas por sensores diferentes de câmeras para alinhamento espaço-temporal, diferente do método proposto nesta dissertação que é a utilização de informações conjuntas providas por pelo menos dois tipos de sensores.

Capítulo 3

Metodologia

A metodologia consiste em quatro etapas distintas como mostrada na Figura 3.1. Na primeira etapa é feita a aquisição de referenciais de informação providos pelos múltiplos sensores (de vídeo e sensoriamento) dispostos no ambiente de forma com que haja sobreposição em seus campos de cobertura. Na etapa seguinte, as variações de velocidade obtidas são extraídas de cada conjunto de dados. A terceira etapa consiste em determinar a equação de alinhamento temporal utilizando os dados de aceleração obtidos anteriormente. Por último, utiliza-se o algoritmo SIFT-RANSAC para determinar o alinhamento espacial resultante em uma matriz homografia.

O fluxograma mostrado na Figura 3.1 representa as etapas de resolução do problema. O problema pode ser dividido em cada etapa mostrada como sendo a sequência de subproblemas a serem resolvidos seguindo esta sequência para que ao final obtenha-se uma equação de alinhamento temporal e uma matriz homografia de alinhamento espacial.

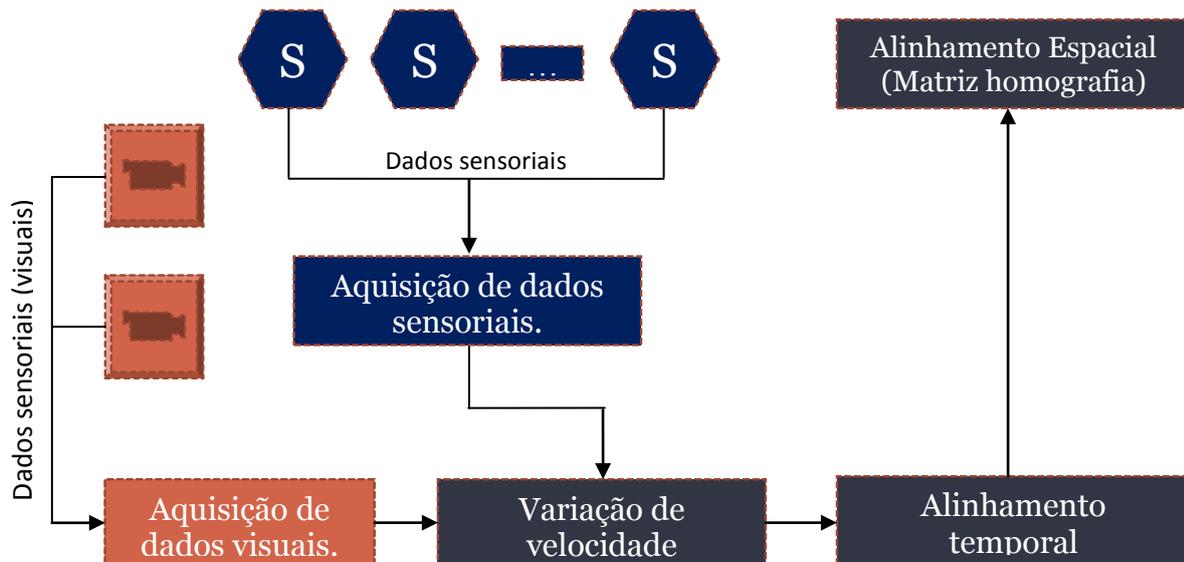


Figura 3.1: Fluxograma da metodologia proposta. As imagens e o posicionamento sensorial são adquiridos, em seguida são extraídas as variações de velocidade. Com os dados de aceleração determina-se o alinhamento temporal e, por último, o alinhamento espacial por uma matriz homografia.

3.1 Aquisição de Dados

A primeira etapa da metodologia consiste na aquisição de dados. Esses dados são adquiridos de duas fontes distintas: os providos por câmeras de vídeo e outro provido por sensores de posição. Utilizar sensores diferentes se dá devido à capacidade de adquirir informações em situações adversas como, por exemplo, uma câmera pode não localizar com precisão um objeto se as condições do ambiente estiverem desfavoráveis quanto à iluminação. Já um sensor de posição está sujeito a interferências e não fornece nenhuma característica a respeito do que está monitorando como cor, tamanho, etc. A ilustração desta disposição está mostrada na Figura 3.2.

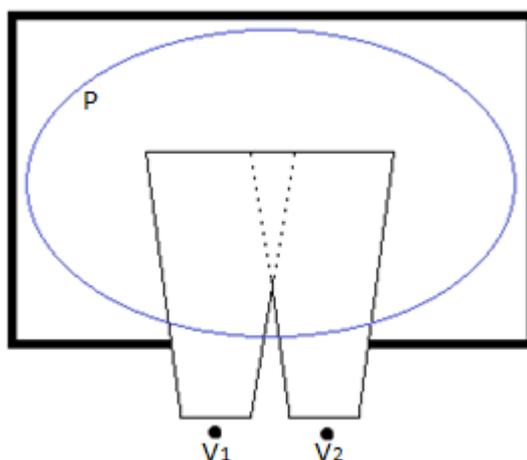


Figura 3.2: Os sensores V_1 e V_2 referem-se as câmeras enquanto que P aos campos de coberturas somados de todos os sensores de localização. Todos os sensores que pertencem ao sistema possuem campos de sobreposição entre si.

Considerando que as câmeras estão estáticas e convenientemente posicionadas, são chamadas de V_1 e V_2 as sequências de *frames* que compõem os vídeos obtidos pelas câmeras um e dois respectivamente, sendo necessária a sua conversão em sequências de *frames* em tons de cinza chamados respectivamente de P_1 e P_2 .

A conversão das sequências V_1 e V_2 em P_1 e P_2 é feita da seguinte forma: seja o frame $f_k(r_k, g_k, b_k)$ um k -ésimo *frame* pertencente a V_1 ou V_2 , os *pixels* dos *frames* que compõem P_1 e P_2 são calculados pela média aritmética de r_k , g_k e b_k .

$$P(k) = \frac{r(k)+g(k)+b(k)}{3} \text{ onde,}$$

$$r_k, g_k, b_k \in f_k.$$

A variação de movimento é obtida por meio da diferença entre imagens binarizadas. Para transformar as sequências de *frames* tons de cinza P_1 e P_2 em sequências de *frames* binarizados B_1 e B_2 , é necessário comparar o valor de cada *frame* p_k pertencente a P_1 ou P_2 com um limiar μ para determinar o valor de cada b_k pertencente a B_1 ou B_2 ,

$$\text{se } p_k < \mu \rightarrow b_k = 0,$$

$$\text{senão, } b_k = 1.$$

O ponto $c_k(x_k, y_k)$ equivale a localização do centroide do objeto em f_k , calculado através de uma estimativa a partir da subtração entre duas imagens binarizadas sequenciais $b_{k+1} - b_k$. Como a região onde o movimento foi detectado é composta por vários pontos, considera-se o mais centralizado como sendo o centroide. A Figura 3.3 ilustra como o conjunto dos pontos encontrados equivale a possível trajetória percorrida na cena.

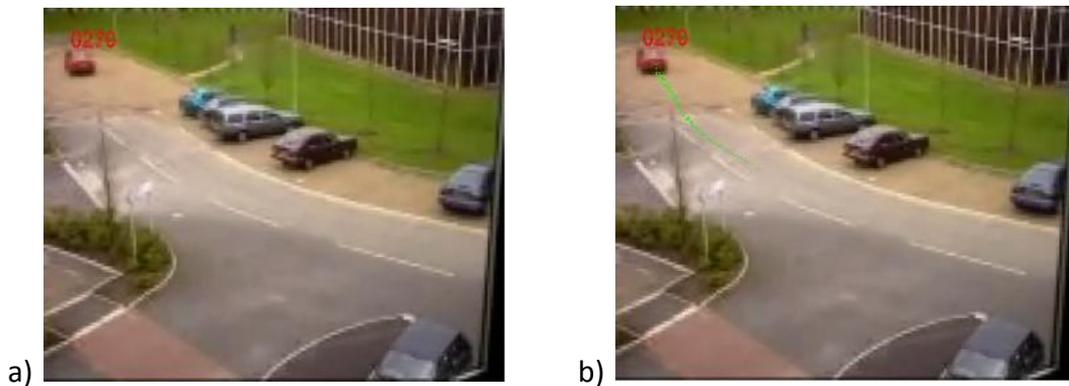


Figura 3.3: em (a) é mostrada imagem real capturada de um veículo em movimento. Após o rastreamento dos centroides do veículo ao longo dos quadros que compõem a sequência, (b) mostra os pontos encontrados. (Vídeos encontrados em pesquisas na *web*)

Os dados de posição, resultados da detecção realizada por um sensor de movimentos, são passados ao sistema no formato $P(x'_i, y'_i, t'_i)$ onde, x'_i e y'_i são

coordenadas de posicionamento e t'_i uma unidade de tempo no formato (hh:mm:ss). A diferença entre escalas de posição que neste sensor são dadas por valores de latitude e longitude, enquanto que nas câmeras são em coordenadas de pixel, são igualadas através do passo seguinte deste modelo, que é a estimativa de variação de velocidade (aceleração).

3.2 Cálculos de Variação de Velocidade e Alinhamento Temporal.

Logo depois de adquiridas, as informações obtidas pelos diferentes sensores estão defasadas quanto a possibilidade de qualquer comparação inicial que possa ser feita de acordo com o que foi observado e obtido. Enquanto os dados visuais estão em coordenadas de *pixels*, os dados sensoriais estão em latitude e longitude. Para que sejam comparados, faz-se necessário estabelecer uma medida padrão. Esta etapa envolve a transformação dos dados de espaço em dados de variação de velocidades, calculados primeiramente estimando a velocidade de deslocamento seguida pelo cálculo de aceleração.

Chamando de C_1 e C_2 as sequências de centroides compostas por elementos $c_k(x_k, y_k)$ detectados em V_1 e V_2 respectivamente, a velocidade v_k corresponde à distância Euclidiana entre os pontos c_k e c_{k+1} .

$$v_k^2 = (x_{(k+1)} - x_k)^2 + (y_{(k+1)} - y_k)^2,$$

Para cálculo de aceleração é utilizado o mesmo princípio do cálculo de velocidade. Cada diferença de velocidade em *frames*, dadas no formato v_{k+1} e v_k , é calculada sua aceleração através de subtração, sendo explícita no seguinte formato:

$$a_k = \frac{\Delta v}{\Delta t}.$$

$$\text{Sendo } \Delta v = v_{k+1} - v_k,$$

$$\text{e } \Delta t = k + 1 - k,$$

$$\text{logo } \Delta t = 1.$$

Portanto, para $a_n = \frac{v_{n+1} - v_n}{1}$, logo

$$a_n = v_{n+1} - v_n.$$

A utilização da aceleração como dado principal se dá pela necessidade de eliminar o efeito PARALAX, que ocorre quando diferentes observadores obtêm impressões diferentes de velocidade a respeito de um mesmo objeto em deslocamento. A velocidade aparente ao observador localizado mais próximo a cena é maior que aquela percebida pelo observador mais distante. Depois de determinadas as acelerações, o problema agora consiste em encontrar a equação de alinhamento temporal.

Considerando que as sequências de informações heterogêneas são chamadas de S_1, S_2, \dots, S_k onde k é o número de sensores (câmeras ou sensores de posicionamento), o problema em questão é relacionar as sequências entre si através de trajetórias comuns (relativas ao mesmo objeto monitorado) não triviais. Esta relação de tempo pode ser descrita pela seguinte equação:

$$S_j = \alpha \cdot S_i + \beta,$$

$$i, j \in \{1, 2, \dots, k\}, \quad i \neq j.$$

A constante α faz referência à diferença entre frequências de obtenção de dados, enquanto que β indica a diferença entre o tempo de início da gravação.

O cálculo das constantes de alinhamento temporal utilizando as acelerações é feito da seguinte forma:

1. Entre duas sequências S_i e S_j de informações de posicionamento, calcular as suas variações de velocidade. É assumido o intervalo de ativação (constante β) entre duas sequências não é maior que n segundos (n é um valor obtido de forma empírica).
2. O valor da constante α é calculado pela divisão entre as taxas de amostragem de S_i e S_j . Logo, $\alpha = \frac{S_i}{S_j}$.

3. Sendo $vet_i = [a_1, a_2, a_3, \dots, a_n]$ uma sequência de variações de velocidade (acelerações) da sequência S_i e $vet_j = [a_1, a_2, a_3, \dots, a_m]$ uma sequência de variações de velocidade (acelerações) da sequência S_j , buscar o valor de β que indique o menor valor de m_{ij} onde:

$$m_{ij} = \left(\sum_{k=1}^n |vet_i(a_{(k+\beta)}) - vet_j(a_k)| \right) / n.$$

4. Obter a equação de alinhamento temporal com os coeficientes α e β

$$S_j = \alpha \cdot S_i + \beta.$$

3.3 Alinhamento Espacial

Após a determinação da equação de alinhamento temporal, o alinhamento espacial corresponde em encontrar a região de sobreposição dos planos de projeção de câmeras diferentes. Possuindo imagens obtidas por câmeras diferentes alinhadas no tempo, é possível encontrar os seus pontos em comum utilizando o algoritmo SIFT para descrição de pontos, o algoritmo RANSAC para encontrar pontos correspondentes e, por último, determinar a matriz homografia que projeta cada *pixel* da câmera V_1 em V_2 . Na literatura são encontradas diversas referências para a utilização do método SIFT-RANSAC para determinação de pontos correspondentes e matriz homografia.

O Algoritmo *Scale Invariant Feature Transformation* (SIFT) consiste em detectar regiões da imagem com alto gradiente (onde há uma grande diferença de tonalidade entre pontos próximos) e, em seguida, descrever os *keypoints* destas regiões com um vetor de valores referentes a cálculos feitos sobre estes.

Já o algoritmo RANSAC recebe dois conjuntos de pontos onde entre estes há os *inliers* que correspondem aqueles que fazem parte da solução (no caso deste problema, os pontos correspondentes entre as imagens) e os *outliers*, aqueles considerados como ruído. O modelo consiste em procurar aleatoriamente supostos pontos correspondentes e assim avaliar o modelo aplicando aos demais pontos.

Depois de uma repetição n de vezes (quantidade determinada empiricamente), o modelo que obteve melhores resultados é utilizado como solução.

O alinhamento espacial resulta em uma matriz homografia que projeta os pontos pertencentes à imagem de um frame f_i em f_j alinhados no tempo. Para cálculo de seus coeficientes, é necessário encontrar os pontos correspondentes determinados pelas etapas anteriores. Para uma dada matriz homografia H , os pontos obedecem os seguintes valores:

$$p(x, y) \times H = p'(x', y').$$

Escrita de forma matricial, os coeficientes podem ser calculados de acordo com o seguinte cálculo:

$$[x \ y \ w] \times H = \begin{bmatrix} X' \\ Y' \\ W' \end{bmatrix} \text{ onde, } w \text{ e } W' \text{ são coordenadas homogêneas.}$$

Portanto, atribuindo o valor 1 à w , a dedução dos seus coeficientes pode ser feita da seguinte maneira:

$$x' = \frac{X'}{W'}, \quad y' = \frac{Y'}{W'}, \quad w = 1,$$

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix},$$

$$X' = x \times h_{11} + y \times h_{21} + h_{31},$$

$$Y' = x \times h_{12} + y \times h_{22} + h_{32},$$

$$W' = x \times h_{13} + y \times h_{23} + h_{33},$$

$$x' = \frac{x \times h_{11} + y \times h_{21} + h_{31}}{x \times h_{13} + y \times h_{23} + h_{33}},$$

$$y' = \frac{x \times h_{12} + y \times h_{22} + h_{32}}{x \times h_{13} + y \times h_{23} + h_{33}}.$$

O algoritmo a seguir descreve os passos necessários para o alinhamento espacial obtido através do alinhamento temporal calculado previamente e utilizando as técnicas descritas anteriormente:

1. Utilizar a equação de alinhamento temporal para obter os *frames* de V_1 e V_2 que estão alinhados no tempo.
2. Aplicar o algoritmo SIFT em cada par de *frames* alinhados temporalmente para obter os seus pontos descritores.
3. Aplicar o algoritmo RANSAC para encontrar os pontos correspondentes entre os pontos descritos no passo anterior.
4. Determinar os coeficientes da matriz homografia H .

O primeiro passo consiste em encontrar os coeficientes α e β da equação de alinhamento temporal através da aceleração dos objetos detectados. Após obtidas estas constantes, aplicar na equação os índices do *frames* de uma das sequências envolvidas para obter o *frame* correspondentes na sequência alinhada e assim, utilizar estas imagens como entrada para o algoritmo SIFT retornar pontos descritores. O terceiro passo envolve a utilização destes pontos no algoritmo RANSAC para determinar quais são os correspondentes para que por último, seja possível determinar os coeficientes da matriz homografia resolvendo o sistema que envolve o cálculo de seus coeficientes.

Capítulo 4

Aplicações Experimentais

Os experimentos foram divididos neste trabalho em três grupos principais: grupo 1 – Simulações, grupo 2 – Experimentação em ambiente controlado e grupo 3 – Experimentação em ambiente real.

Os experimentos do grupo 1 consistiram em criar uma simulação onde são escolhidas trajetórias aleatórias em um ambiente plano, que simbolizam o deslocamento de um objeto de um ponto qualquer p_i até um p_f onde, simbolizam respectivamente o ponto de partida e ponto de chegada. Após a criação de um caminho percorrido, são simuladas as detecções dos pontos que compõem a trajetória fictícia seguindo os princípios da detecção de centroides em vídeos e localização de objetos através da triangulação de dados de sensores de posicionamento, cuja informação fornecida é a distância do objeto monitorado para cada um destes. Espera-se que nesta etapa seja possível aplicar a metodologia desenvolvida em condições adversas como inclusão de erros na detecção de dados, diferença na frequência de detecção além do próprio modelo de alinhamento.

Nos experimentos do grupo 2, foi utilizado um objeto remotamente controlado em um ambiente fechado para que fossem minimizados os ruídos externos. Nesta etapa de testes, o objetivo é testar a funcionamento da metodologia proposta em dados reais, além de, observar o comportamento do erro na localização dos centroides ao longo da sua detecção.

Por último, os experimentos do grupo 3 foram realizados com câmeras posicionadas em uma posição elevada capturando a imagem de um objeto em movimento onde, as variações de luminosidade eram mais frequentes e outros objetos poderiam interferir na posterior captura de centroides. Além disso, foi acoplado a este

objeto um sensor de movimentação (GPS) para obterem-se informações de posicionamento além de informações visuais.

O modelo foi implementado e executado em ambiente MATLAB versão 2011a. O equipamento simulado também foi desenvolvido no ambiente MATLAB e os equipamentos utilizados durante os testes em ambiente controlado e ambiente real foram duas câmeras JVC GZ-MG330 e GARMIN GPSmap76CSx como sensor alternativo.

4.1 Experimentos em Ambiente Simulado

O objetivo destes experimentos é testar o modelo computacional com dados sintéticos e heterogêneos. Através de um modelo simulado, a inclusão de ruído fez-se de forma controlada validando a metodologia proposta.

A Figura 4.1 mostra uma esquematização do modelo de simulação onde as circunferências representam os campos de cobertura dos sensores, os retângulos pontilhados os campos de visão das câmeras e a área retangular em destaque como a área de sobreposição. Os pontos com bordas em destaque representam a localização de sensores e o ponto preenchido como o objeto monitorado.

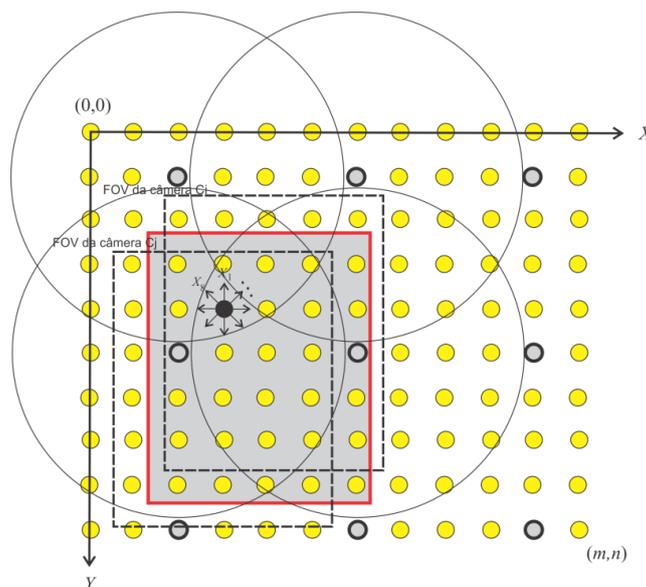


Figura 4.1: Esquematização da área de cobertura no modelo de simulação. O objeto monitorado movimenta-se pelo plano com probabilidades de seguir em cada direção. Todos os sensores possuem um campo de cobertura próprio.

Cada movimento do objeto no plano está associado a uma probabilidade previamente ajustada. Assim, sejam os eventos X_1 = movimento para frente, X_2 = movimento para os lados/frente, X_3 = movimento para os lados/atrás e X_4 = movimento de recuo, os valores de probabilidade são tais que $P(X_1) + P(X_2) + P(X_3) + P(X_4) = 1$. Com isso, os dados de saída são dispostos em um vetor de pontos \vec{v} onde cada posição $v[n]$ equivale a um ponto $p(x, y)$ no plano de simulação. Com estes dados disponíveis, é então possível simular a detecção do movimento sob as perspectivas de cada sensor. Empiricamente, foram determinados os seguintes valores de probabilidade:

$$P(X_1) = 0.35,$$

$$P(X_2) = 0.30,$$

$$P(X_3) = 0.25,$$

$$P(X_4) = 0.10.$$

No caso de um sensor visual, determina-se uma área no plano como sendo a área de cobertura da câmera e suas coordenadas de posicionamento como sendo a localização do *pixel* onde está localizado o centroide.

Já com sensores de posicionamento, atribuem-se pelo menos três coordenadas aleatórias distintas representando sua disposição no espaço. Em seguida, é calculada a distância Euclidiana de cada um destes sensores até a real localização do objeto em cada instante, além de uma inclusão de erro de precisão. Por último, utiliza-se da técnica de triangulação com todos os sensores disponíveis para que seja obtida uma sequência final de posicionamento, representando a trajetória detectada pelo conjunto de sensores. O sistema determinado seguinte demonstra como são computadas as distâncias para cada sensor e calculadas as coordenadas de posicionamento do objeto:

$$d_1^2 = (x_1 - x)^2 + (y_1 - y)^2$$

$$d_2^2 = (x_2 - x)^2 + (y_2 - y)^2$$

$$d_3^2 = (x_3 - x)^2 + (y_3 - y)^2$$

Em uma primeira simulação, foram gerados dados simulados de posicionamento ao longo de um plano e em seguida detectados estes movimentos. Este teste teve como objetivo analisar visualmente o comportamento de dados de aceleração vistos sob perspectivas diferentes, sendo estas a partir de um sensor de posicionamento e uma câmera. A Figura 4.2 mostra as acelerações ao longo do experimento para os dois tipos de sensores, ambos gerados através de um passeio aleatório sob um plano de tamanho 2500x2500. Apesar de a frequência com que as informações são adquiridas serem diferentes para cada tipo de sensor, seu comportamento é similar em ambos os casos, pois a escolha da variação de velocidade como dado principal para comparação, elimina o efeito PARALAX onde, objetos que estão sendo observados de distâncias diferentes podem aparentar não estar na mesma velocidade. Para efeitos de comparação entre os dados de aceleração em todos os experimentos, foi utilizada como parâmetro a aceleração em m/s^2 .

Número de sensores de posição: 3.

Latência dos sensores de posição: 10 posições por segundo.

Número de câmeras: 1.

Latência da câmera: 25 frames por segundo.

Comparação de variações de velocidade entre câmera e sensor.

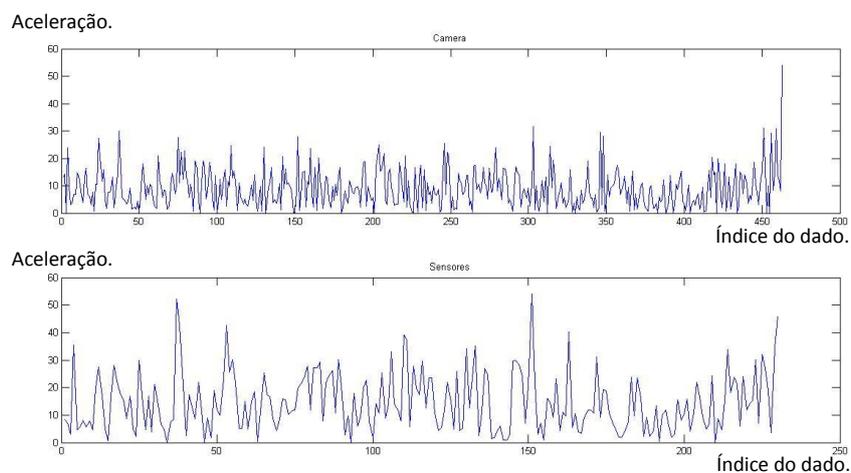


Figura 4.2: Acima, aceleração do ponto de vista de câmeras e, abaixo, do ponto de vista de sensores. Apesar de as câmeras capturarem a informação com uma frequência maior, seu comportamento é similar ao dos sensores ao longo do tempo por terem sido eliminadas as ambiguidades de observação.

A simulação seguinte teve testes sob a metodologia proposta para alinhamento temporal. O objetivo desta etapa foi calcular o índice de acerto do método sem a inclusão de ruídos semelhantes aos encontrados em situações práticas. O experimento simulado foi aplicado repetidamente para testar o comportamento da precisão de determinação dos coeficientes da equação de alinhamento temporal. Como o acréscimo de diferença no ponto de partida entre o início de captura dos dados de cada sensor pode ser determinado manualmente na simulação, é possível ao final de cada experimento avaliar se o algoritmo encontrou um resultado satisfatório. A Figura 4.3 mostra a porcentagem de acertos variando durante a quantidade de repetições do experimento simulado onde, a precisão é testada em 50 simulações para um deslocamento horizontal e vertical do objeto com valor máximo de 10, sem inclusão de erros.

Número de câmeras: 2.

Latência das câmeras: 25 *frames* por segundo.

Número de repetições: 10 vezes de 50 cada.

Índice de acertos do método ao longo da repetição do experimento simulado.

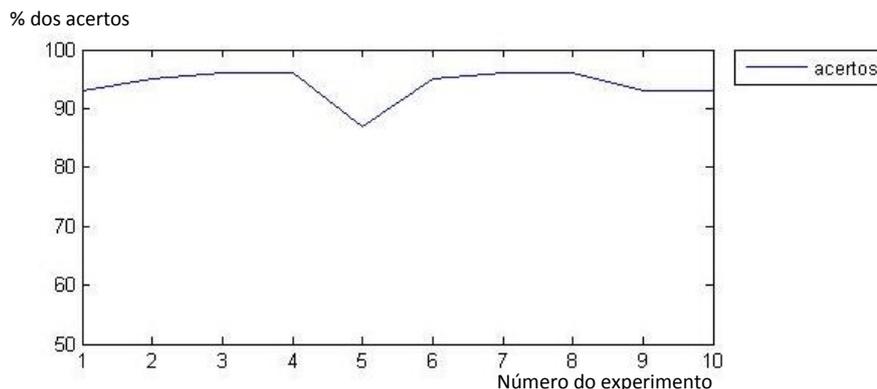


Figura 4.3: Durante 50 iterações, a simulação gera um passeio aleatório com velocidade máxima de 10, aplica uma diferença entre o início de aquisição de dados de cada sensor e verifica se o deslocamento encontrado foi similar ao aplicado. Repetindo 10 vezes este experimento, obteve-se uma média de 94% de acerto.

O modelo de simulação mostrou-se útil para trabalhar com um índice de deslocamento do objeto, valor este que caracteriza a distancia máxima percorrida pelo objeto em um passo. Realizando um próximo experimento, desta vez variando a distância máxima de deslocamento entre 4 e 12, a Figura 4.4 mostra um gráfico que representa a taxa de acertos com a variação de velocidade.

Número de câmeras: 2.

Latência das câmeras: 25 frames por segundo.

Repetições do experimento: 100 para cada índice de deslocamento de 4 a 12.

Índice de acertos de acordo com o índice da distância máxima percorrida pelo objeto.

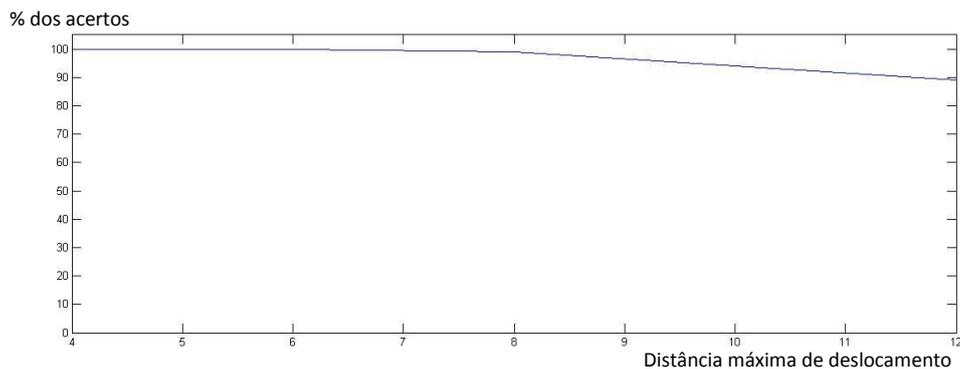


Figura 4.4: Gráfico de comparação entre a variação da taxa de acertos com a variação de deslocamento. A medida que aumenta-se o valor máximo que o objeto pode deslocar-se, a precisão sofre um decréscimo.

Como o modelo sofre uma perda considerável de precisão a medida que é considerado um maior deslocamento do objeto na cena simulada, tornou-se necessária a sua observação em um experimento com condições similares ao que é pretendido ser feito na experimentação real. Um dos experimentos em laboratório que serão abordados no próximo tópico deste capítulo foi utilizado para determinar o índice mais adequado.

Colocando os valores de velocidade (distância Euclidiana entre as coordenadas do centroide ao longo do experimento) obtidos neste experimento laboratorial em um

vetor, forma-se o gráfico mostrado na Figura 4.5. Conclui-se que, a média de velocidade do objeto ao longo do experimento possui valor de 3.18.

Câmeras utilizadas: 1

Taxa de *frames*: 25 por segundo.

Aceleração detectada pelo movimento de um objeto ao longo de um experimento em laboratório.

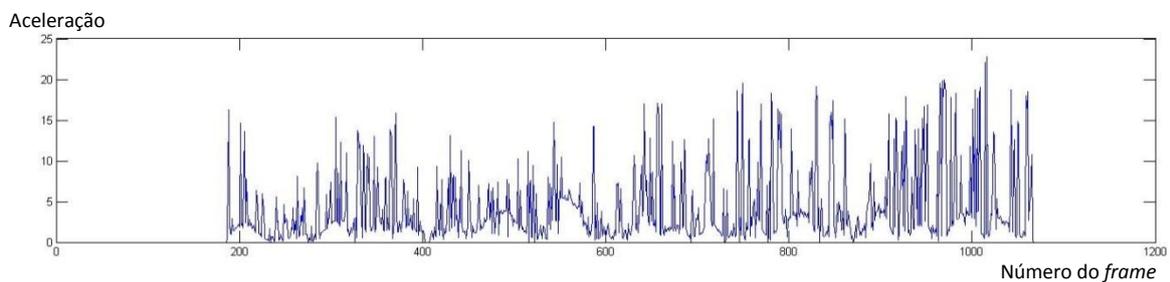


Figura 4.5: O gráfico mostra a velocidade ao longo dos frames que compõem o vídeo em que foram obtidos os centroides de um objeto em deslocamento. A média de velocidade foi de 3.18.

Para determinar qual valor da variável de deslocamento é mais adequada para os testes simulados, foram realizados 100 testes simulados para cada índice de deslocamento variando de 4 até 10 e computando a média de velocidade em cada. Desta forma foi possível concluir que o fator de deslocamento com valor igual a 6 foi o mais adequado, como mostra a Figura 4.6.

Número de câmeras: 1.

Taxa de *frames*: 25 por segundo.

Repetições do experimento: 100 para cada índice de 4 a 10.

Velocidade média do objeto em cada valor de deslocamento aplicado.

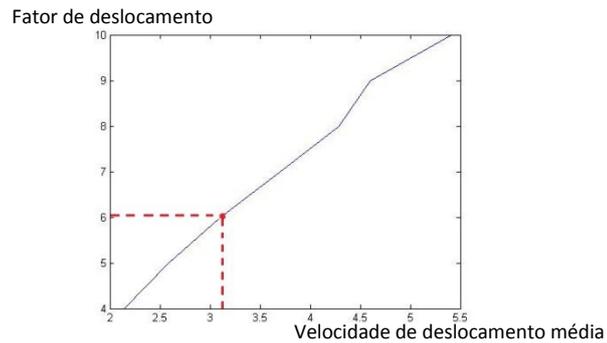


Figura 4.6: O valor de deslocamento que mais se aproxima da média obtida no teste de laboratório foi com valor 6, resultado obtido computando a velocidade média do centroide ao longo do testes simulados.

O modelo foi novamente testado aplicando o valor 6 como deslocamento máximo, a precisão de acertos teve melhora significativa, variando em torno de 99%, como mostra a Figura 4.7. Este resultado serviu de embasamento para que os experimentos em laboratórios continuassem, com a confiabilidade de que o modelo de alinhamento temporal obteve resultados satisfatórios durante a simulação, afirmando a ideia de utilizar a variação de velocidade (aceleração) como base para comparação.

Número de sensores de posição: 3.

Taxa de amostragem dos sensores: 10 posições por segundo.

Número de câmeras: 1.

Taxa de amostragem da câmera: 25 *frames* por segundo.

Índice de deslocamento do objeto: 6.

Número de repetições: 10 vezes de 100 cada.

Taxa de acertos do método.

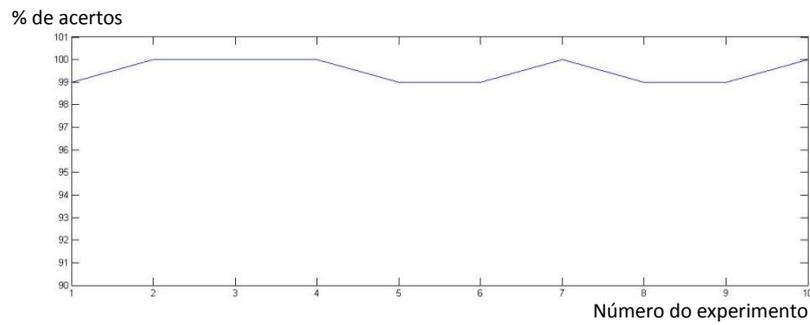


Figura 4.7: A figura mostra como a taxa de acertos variou ao longo da repetição do experimento, oscilando em torno de 99% com os dados sendo computados a cada 100 simulações.

De posse dos resultados obtidos anteriormente e verificando o comportamento satisfatório do modelo, foram realizados testes simulados com a inclusão de erro nas acelerações determinadas por cada sensor. Os dados de aceleração de uma sequência S_i estão armazenados em um vetor vet_i , onde cada posição $vet_i[a_n]$ armazena um valor. Um vetor de erros ver_i é criado com a mesma dimensão de vet_i contendo valores aleatórios ε onde, $-\gamma \leq \varepsilon \leq \gamma$, sendo que, γ é uma taxa de ruído. Assim é criado um vetor de perturbação nos dados de aceleração $vetEr_i$ definido da seguinte forma:

$$vetEr_i = vet_i + ver_i .$$

A Figura 4.8 mostra a taxa de acertos do experimento considerando como dados de acelerações aqueles armazenados no vetor $vetEr_k$ para cada sequência S_k . Os resultados continuaram satisfatórios, percebendo um declínio elevado no resultado a partir de uma variação onde $\gamma = 60$.

Número de sensores de posição: 3

Latência da detecção de posições: 10 por segundo.

Número de câmeras: 1.

Latência da câmera: 25 por segundo.

Repetições do experimento: 50 para cada valor de erro de 5 até 60.

Índice de acertos do método com inclusão de erros.

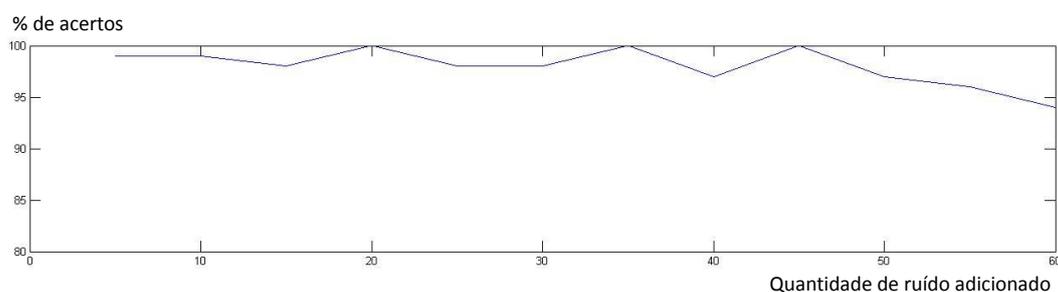


Figura 4.8: O eixo y contém a taxa de acertos e o eixo x a variação de γ , a variável de ruído. O índice varia entre 98% e 99% até $\gamma = 50$ onde, a partir deste ponto, a precisão sofre um declínio para aproximadamente 94%.

Com os testes a serem realizados em laboratório, espera-se obter a partir de dados reais quanto pode ocorrer de variação em um dado de aceleração entre duas sequências e feita a comparação com os resultados obtidos no teste simulado, a fim de determinar se a metodologia provê bons resultados mesmo com a perturbação encontrada em um experimento prático.

4.2 Experimentos em Laboratório

Após a realização de experimentos simulados, foram realizados testes em laboratório com o objetivo de validar em uma situação prática o modelo de alinhamento temporal e analisar os resultados do alinhamento espacial. Foram dispostas duas câmeras lado-a-lado com uma distância aproximada, rotações similares e campos de visão não disjuntos monitorando um objeto controlado remotamente reproduzindo movimentação aleatória como mostra a figura 4.9. A taxa de amostragem é de 25 *frames* por segundo e as demais configurações (compensação de cores, foco, etc.) são estritamente iguais permanecendo inalteradas durante o experimento.

Nesta experimentação não foi incluso o sensor de movimentação, pois, sua finalidade era testar a viabilidade da metodologia proposta e assim prosseguir com o experimento real.



Figura 4.9: Visões do mesmo objeto em repouso observado por câmeras diferentes.

Aplicando o detector de centroides com o algoritmo proposto na metodologia, foi possível perceber que houve um bom comportamento na detecção e que há uma pequena imprecisão na sua determinação, especialmente quando o objeto está se deslocando em baixa velocidade (de acordo com a perspectiva da câmera). Com os centroides de coordenadas $c(x, y)$ foi possível calcular a velocidade de deslocamento do objeto entre cada frame através da distância euclidiana dada em *pixels/frame*. A aceleração em seguida foi calculada como sendo a diferença entre as velocidades de dois frames consecutivos. A Figura 4.10 mostra alguns frames retirados do vídeo completo com a marcação dos centroides nos pixels já percorridos pelo objeto.

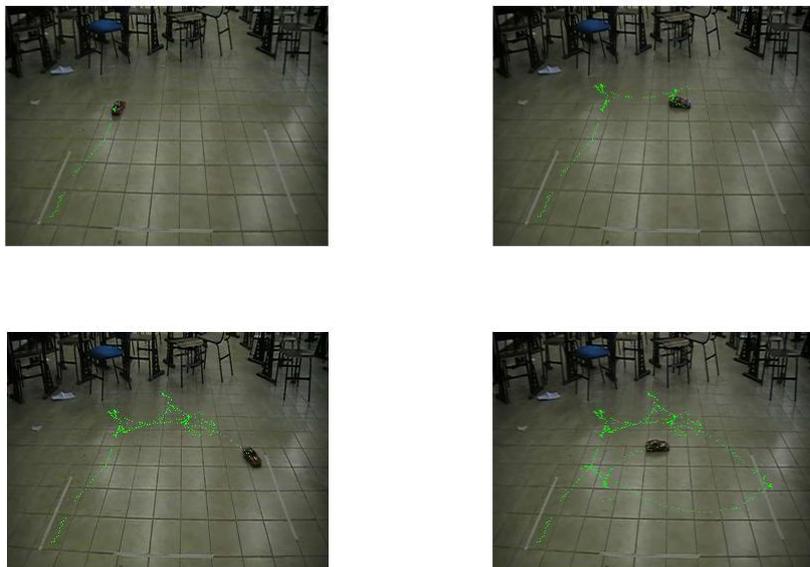


Figura 4.10: Sequência de frames que mostram a localização dos centroides detectados nos frames anteriores (os frames acima são de apenas uma câmera).

Depois de detectados os centroides em ambas as câmeras, foi possível observar primeiramente que o caminho percorrido em cada perspectiva possui características similares. Apesar disso, é sabido que há variações de espaço nestes casos causadas pelo efeito PARALAX que induz o observador a estimar uma velocidade abaixo da realmente percorrida pelo objeto. A Figura 4.11 mostra o último *frame* de cada câmera com as marcações anteriores de onde foi detectado o centroide.

Feitos os cálculos de acelerações, a metodologia proposta para encontrar o alinhamento temporal foi aplicada, gerando assim os primeiros resultados práticos do algoritmo proposto.



Figura 4.11: Último *frame* de cada vídeo com a marcação do caminho percorrido pelo objeto de acordo com seu centroide.

Para este experimento, os coeficientes da equação de alinhamento foram calculados da seguinte forma:

$$\alpha = \frac{\left(\frac{frames}{seg}\right)_1}{\left(\frac{frames}{seg}\right)_2} = \frac{25}{25} = 1, \beta = 139.$$

O coeficiente α faz referência à divisão entre a taxa de *frames* da câmera um pela taxa de *frames* da câmera dois, que possuem o mesmo valor, resultando em $\alpha = 1$. O Coeficiente β foi calculado como sendo o valor 139 de acordo com o algoritmo de alinhamento temporal. A equação de alinhamento temporal resultante neste experimento ficou da seguinte forma:

$$S_1 = S_2 + 139.$$

Aplicando esta equação para o *frame* 400 da câmera um, temos que o *frame* correspondente na câmera dois é o de índice 261. A Figura 4.12 mostra a comparação neste caso.



Figura 4.12: Duas imagens geradas por câmeras diferentes alinhadas no tempo de acordo com o algoritmo proposto para alinhamento temporal.

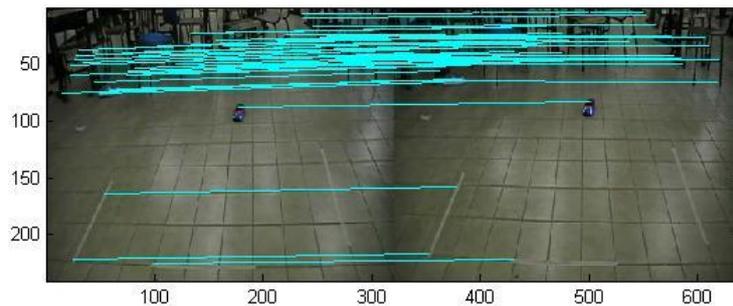
Com frames alinhados temporalmente, é possível então iniciar os passos necessário para obtenção do alinhamento espacial, cujo resultado será uma matriz homografia contendo os valores para projetar pontos de uma imagem de S_1 e S_2 .

O algoritmo SIFT é utilizado como descritor de pontos inicialmente para criar uma base comparativa. São escolhidos aleatoriamente três pares de imagens alinhadas no tempo previamente: as de índice 339, 589 e 800 em S_1 , que resultam nas de índice 200, 450 e 661 em S_2 respectivamente.

Depois de extraídos estes pontos o algoritmo RANSAC é aplicado para encontrar aqueles que são correspondentes, como mostrado na Figura 4.13. Esta etapa é de fundamental importância para que o modelo seja o mais preciso possível, sendo assim a sua avaliação validada para seguir com experimentos em condições reais. Apesar de ser comprovada a sua eficiência em diversos trabalhos anteriores de visão estereoscópica, em um problema de alinhamento espaço-temporal há o fator dependente do alinhamento temporal para que sua utilização seja também satisfatória nesse caso.

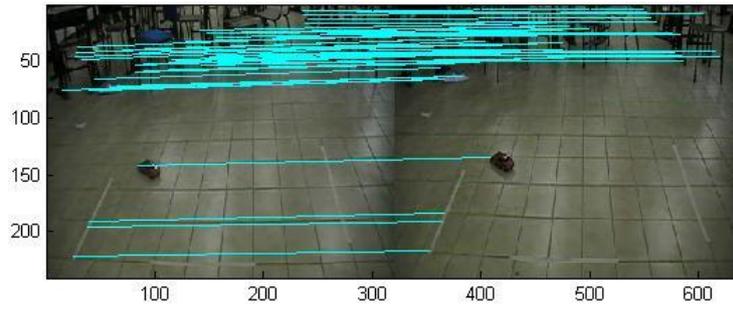
Com a determinação dos pontos correspondentes, o sistema sobre determinado para definir os coeficientes da matriz homografia H pode então ser concluído. Realizando multiplicações das coordenadas de um ponto $p(x, y) \in S_1(f_i)$ pela matriz H , o resultado obtido é seu ponto correspondente no plano de projeção da câmera que gerou as imagens de S_2 . As matrizes resultantes também são mostradas na Figura 4.13.

(a)



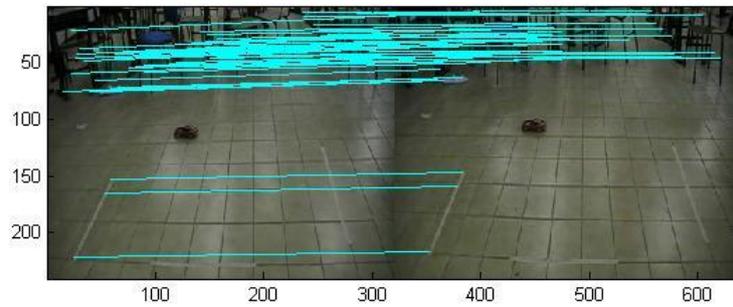
$$H = \begin{bmatrix} 0.9340 & -0.0745 & 10.5611 \\ -0.0558 & 0.9523 & 14.6610 \\ -0.0002 & -0.0000 & 1 \end{bmatrix}$$

(b)



$$H = \begin{bmatrix} 0.9311 & -0.1120 & 12.2796 \\ -0.0542 & 0.9471 & 14.4545 \\ -0.0002 & -0.0002 & 1 \end{bmatrix}$$

(c)

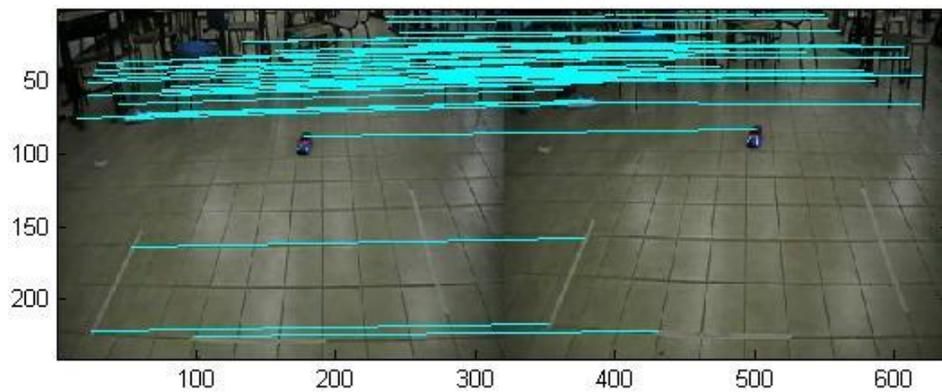


$$H = \begin{bmatrix} 0.9437 & -0.0729 & 9.9094 \\ -0.0535 & 0.9464 & 14.6555 \\ -0.0001 & -0.0001 & 1 \end{bmatrix}$$

Figura 4.13: As figuras a, b e c referem-se respectivamente aos combinados frames 339, 589 e 800 da câmera um e 200, 450 e 661 da câmera dois. Em cada uma são mostrados os pontos correspondentes encontrados pelo algoritmo RANSAC, as duas imagens sobrepostas alinhadas no espaço e a matriz homografia H com seus coeficientes.

Este teste de alinhamento espaço-temporal em laboratório utilizando somente vídeos foi também testado com aplicação de diferentes tipos de ruídos em imagens e, quando necessário, aplicado um filtro corretivo. A Figura 4.14 mostra novamente como o algoritmo encontrou o alinhamento espaço-temporal aplicando diferentes coeficientes de erro e seus respectivos resultados.

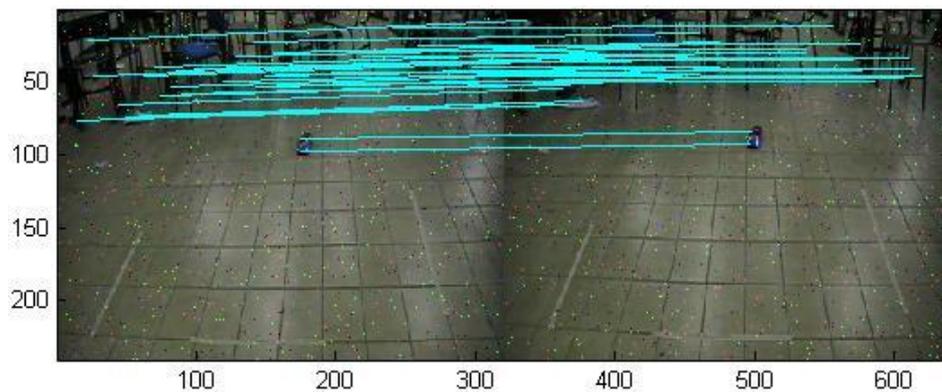
(a)



Pontos descritores encontrados: 315.

Pontos correspondentes encontrados: 60.

(b)



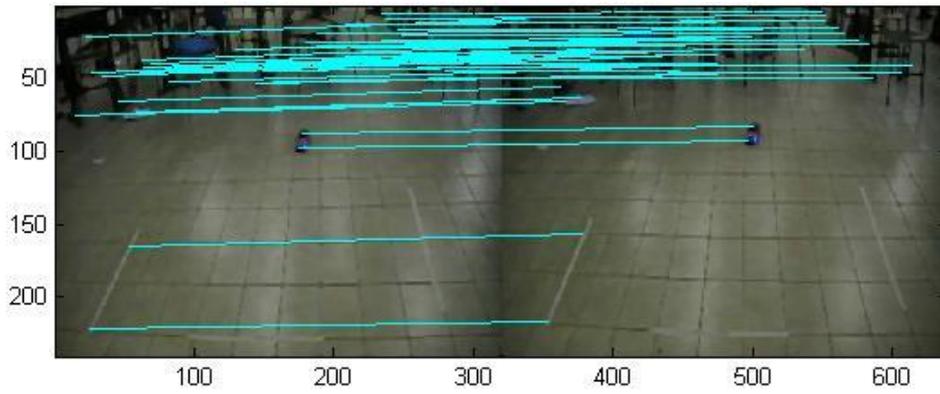
Pontos descritores encontrados: 329.

Pontos correspondentes encontrados: 45.

Tipo de ruído: "sal e pimenta".

Coefficiente de ruído aplicado: 0.02.

(c)

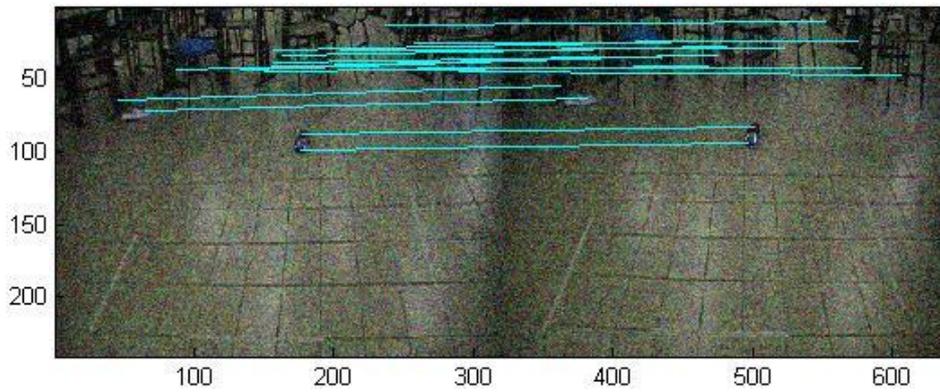


Pontos descritores encontrados: 313.

Pontos correspondentes encontrados: 51.

Filtro aplicado: mediana.

(d)



Pontos descritores encontrados: 319.

Pontos correspondentes encontrados: 19.

Tipo de ruído: gaussiano.

Coefficiente de ruído: 0.01.

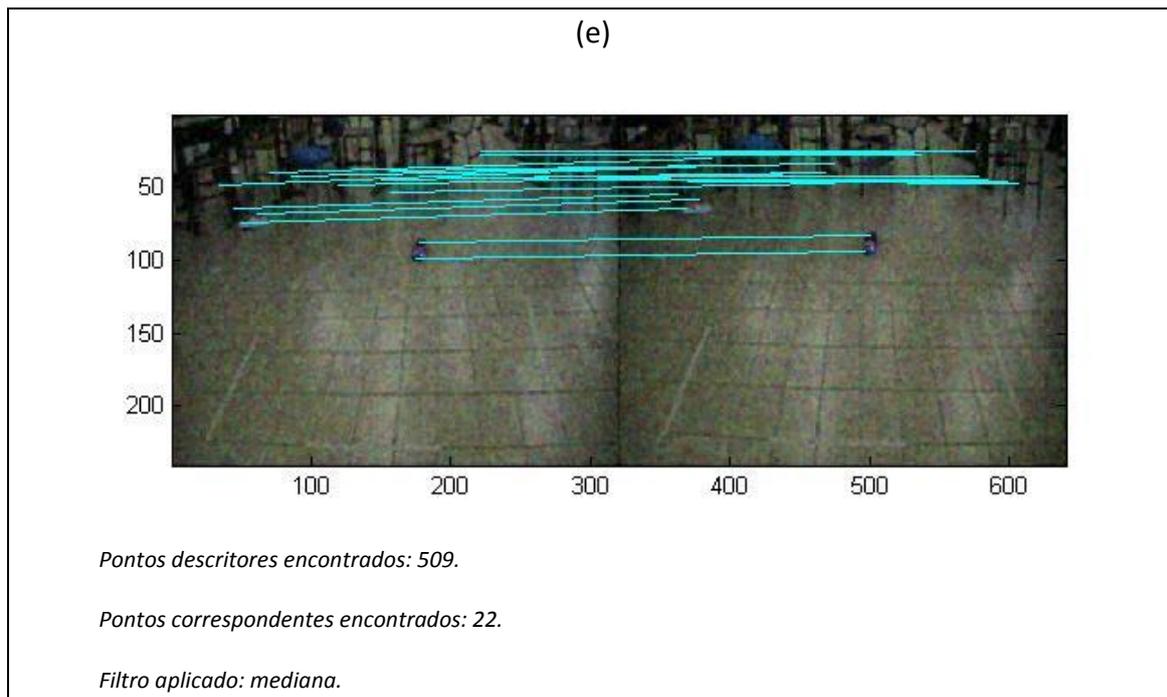


Figura 4.14: A figura (a) mostra o frame 339 pertencente à Sequência um lado-a-lado com o frame 200 da Sequência dois sem acréscimo de ruído e o resultado encontrado. As figuras (b) e (c) mostram as mesmas imagens com acréscimo de ruído “sal e pimenta”, comparando as diferenças entre o resultado sem aplicação de filtro e com aplicação de filtro da mediana. As figuras (d) e (e) seguem as mesmas condições das (b) e (c), mas com aplicação de ruído gaussiano.

Com esta expectativa o experimento real foi definido que será em um ambiente menos controlado, onde há a possibilidade de serem detectados diferentes tipos de movimento além do objeto, as condições de iluminação natural podem variar e um equipamento GPS será utilizado para prover os dados de posicionamento adquiridos de uma fonte alternativa à dos sensores visuais até então utilizados com exclusividade. Se o total de pontos correspondentes encontrados for muito baixo indicando presença de ruído, o filtro da mediana será aplicado para correção.

4.3 Experimentos em Condições Reais

A experimentação real foi conduzida utilizando duas câmeras alinhadas que monitoraram um veículo rastreado por GPS (representando o sensor não visual). Os dados de latitude e longitude com o horário de cada ponto foram utilizados para

determinação da aceleração neste caso. A figura 4.15 mostra uma parte do que foi obtido.

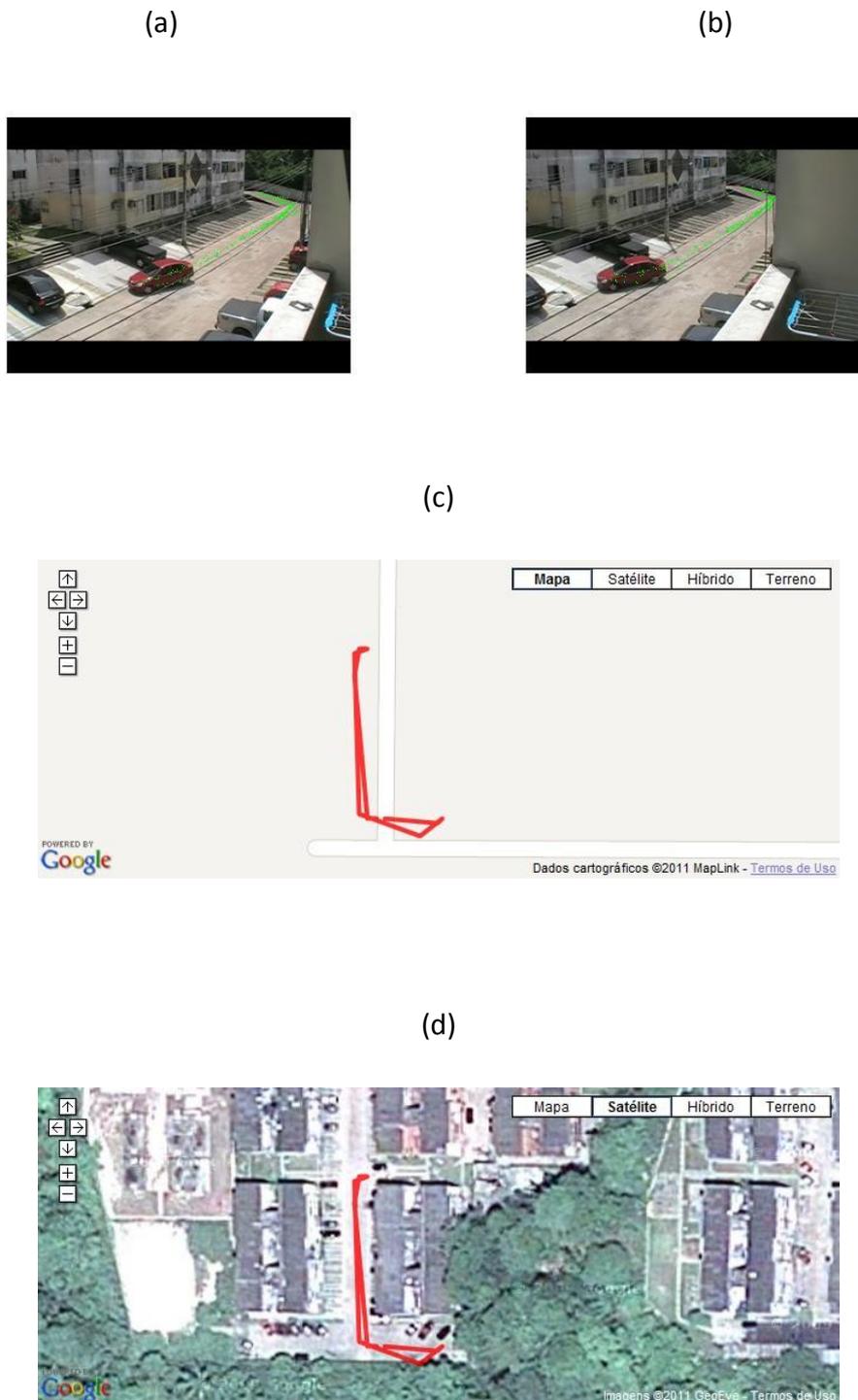


Figura 4.15: As imagens (a) e (b) mostram a perspectiva das câmeras dispostas lado-a-lado capturando o movimento de um veículo. As imagens (c) e (d) mostram a rota percorrida de acordo com os dados capturados do GPS, utilizando como plano de fundo o Google Maps e Google Earth respectivamente.

Os dados obtidos pelo GPS são mostrados com as coordenadas de latitude e longitude mais o horário em data, horas, minutos e segundos. Foi detectado que o tempo de captura não segue uma frequência padrão como encontrado, por exemplo, nas câmeras (Ex.: 25 *frames* por segundo). O tempo entre uma posição e outra pode variar de quatro a cinco segundos até dez e doze segundos. Para eliminar tal problema viu-se necessária fazer uma interpolação nos dados para que haja concordância no alinhamento entre as variações de velocidade.

A interpolação dos dados do GPS foi feita encontrando os pontos intermediários entre cada coordenada. Estes pontos preenchem intervalos de 1 segundo entre cada coordenada detectada. Para isto, foram calculados os k pontos entre uma coordenada $C_n(t)$ e $C_{n-1}(t')$ onde $k = t' - t$ com distancias equivalentes entre si. Para que isto não tornasse uma ambiguidade, pois, a distância supostamente percorrida em cada segundo seria a mesma durante a interpolação fazendo com que a aceleração fosse igual a zero, foram acrescidos valores aleatórios em cada coordenada interpolada simbolizando um eventual erro de detecção de posicionamento. Nos experimentos foram consideradas as amostras das câmeras colocadas à esquerda e direita como sendo respectivamente S_1 e S_2 . As relativas ao GPS foram consideradas S_3 .

Aplicando no primeiro experimento os algoritmos desenvolvidos de acordo com a metodologia, obtiveram-se as variações de velocidades mostradas na Figura 4.16 comparando cada amostra já transformada em escalas de vel/seg^2 .

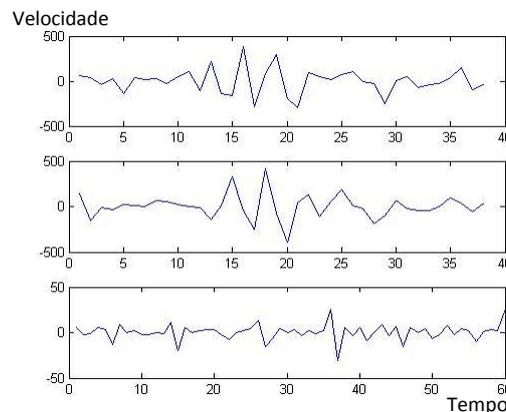


Figura 4.16: Comparativo entre as acelerações de S_1 , S_2 e S_3 . As amostras providas de câmeras possuem visualmente um comportamento similar, enquanto as providas de GPS são consideravelmente diferentes devido aos dados interpolados que, dependendo do intervalo entre a aquisição das posições, estão mais suscetíveis a erros.

Abaixo as equações de alinhamento relacionando cada sensor:

$$S_1 \cong S_3 \times 25 + 25,$$

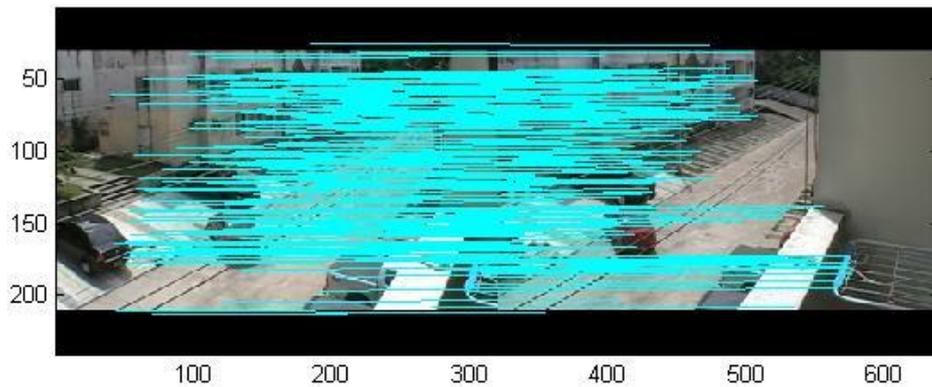
$$S_2 \cong S_3 \times 25 + 0,$$

$$S_1 = S_2 + 17.$$

Com estes valores pôde-se concluir principalmente que:

- a) S_2 está atrasado 17 *frames* em relação a S_1 , valor este que foi atribuído ao coeficiente β da equação de alinhamento temporal que relaciona estas duas sequências.
- b) O desalinhamento temporal de S_3 com as câmeras não é possível obter com exata precisão devido à defasagem de amostras, porém, conclui-se que está entre 0 e 1 segundos.
- c) Testando os possíveis valores para β que atendem ao desalinhamento temporal entre S_1 e S_2 , conclui-se que o erro para a subtração dos vetores que contém os dados de aceleração é de aproximadamente 5.7 para uma média de 10 entre os demais valores testados. Isto indica que por se tratar do valor correto de acordo com a metodologia, a margem de erro está bem abaixo da média, aumentando consideravelmente a segurança de que este valor é confiável.
- d) Os valores de β para o desalinhamento entre S_3 e os outros sensores é dado por uma margem de erro com valores de 59.7 em uma média de 61 e de 91.2, em uma média de 93.1 para os testes com S_1 e S_2 respectivamente. Conclui-se que o erro está com uma variação muito próxima da média, o que se remete a falta de segurança em concluir com precisão o real desalinhamento temporal, problema causado pela defasagem na captura de dados.

O alinhamento espacial resultante é mostrado na Figura 4.17 juntamente com a decomposição de suas etapas, utilizando a equação de alinhamento temporal determinada anteriormente.



$$H = \begin{bmatrix} 0.9435 & -0.0244 & 39.3410 \\ 0.0027 & 0.9598 & 4.6954 \\ -0.0001 & 0 & 1 \end{bmatrix}$$

Figura 4.17: A Figura acima mostra respectivamente o frame 557 de S_1 e 540 de S_2 , os pontos correspondentes encontrados e as imagens sobrepostas alinhadas no espaço. Por último, a matriz homografia com seus coeficientes.

Realizado o alinhamento espacial, pôde-se concluir que os resultados foram bastante satisfatórios, visto que foram encontrados 192 pontos correspondentes (muito acima do encontrado em laboratório, provavelmente devido ao maior número

de detalhes existentes na imagem, o que causa maior diferenciação nos pontos) e a imagem de alinhamento temporal resultante foi muito similar às imagens originais.

Para as cinco demais experimentações que seguiram o mesmo processo do primeiro experimento, a localização de pontos correspondentes variou entre 192 e 188, indicando que o alinhamento espacial continuou seguindo corretamente em decorrência da boa precisão seguida do alinhamento temporal. A sincronização com os dados do GPS falharam em duas experimentações em decorrência da defasagem que chegou por vários instantes a ser em torno de 10 segundos, criando assim uma interpolação imprecisa.

A Figura 4.18 mostra as rotas percorridas sobrepostas no mapa de acordo com os dados do GPS nos pontos adquiridos em todos os experimentos. Isto mostra que a sua precisão de localização foi satisfatória, com problemas apenas na sua taxa de amostragem para uma eventual comparação utilizando dados de aceleração.



Figura 4.18: As rotas sobrepostas mostram que o caminho percorrido ao longo de cada experimento foi similar. A variação ocorreu na sua taxa de amostragem.

Para demonstrar a irregularidade com que os dados foram adquiridos pelo GPS, a Figura 4.19 demonstra em um gráfico a variação do intervalo entre os dados adquiridos. Para uma taxa de aquisição ideal, a forma do gráfico gerado deveria estar

próxima de uma reta sendo possível observar uma maior taxa de acertos nos casos onde o GPS forneceu informações em que o seu intervalo de aquisição foi menor.

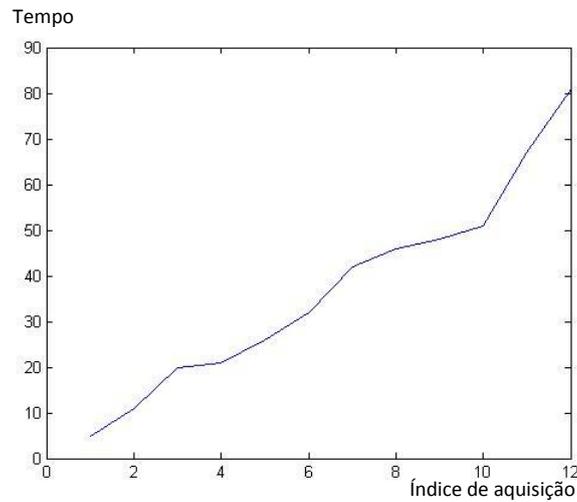


Figura 4.19: O eixo y mostra com quantos segundos após o início da aquisição dos dados cada posição foi adquirida. O eixo x contém a ordem (primeira, segunda, terceira, etc.) em que cada dado foi adquirido.

O índice total de acertos foram de 5 em 7 experimentos para os problemas de alinhamento temporal, sendo que a utilização de sensores de movimento são para este fim, enquanto que o algoritmo SIFT-RANSAC mostrou-se eficaz em todos os *frames* utilizados como teste, gerando imagens sobrepostas fiéis as originais. Disto foi possível concluir que, nos casos onde houve falha de alinhamento temporal entre câmera e sensor, o intervalo de aquisição dos dados no GPS chegou a ser de 11 segundos, enquanto que a câmera sempre obtém 25 *frames* por segundo, causando uma defasagem que não foi possível de contornar mesmo com a interpolação dos dados.

Capítulo 5

Conclusões e Trabalhos Futuros

Neste trabalho foi estudado o problema do alinhamento espaço-temporal, uma abordagem que envolve alinhar informações com origem em fontes diferentes, sendo neste caso, imagens de vídeos e dados de posicionamento por sensores. Foi estabelecida a hipótese de que, com a utilização de sensores adicionais ao sistema, o problema poderia ser igualmente solucionado aos casos onde há apenas o uso de câmeras.

Para a comprovação da metodologia proposta, foram realizadas experimentações simuladas, experimentos laboratoriais e uma experimentação real. Na simulação desenvolvida, foram obtidos dados a serem alinhados no tempo utilizando o método de alinhamento temporal. Nos experimentos em laboratório, o método de alinhamento temporal foi utilizado para que o alinhamento espacial fosse possível de ser testado, no qual se utilizou o algoritmo SIFT-RANSAC. Por último, uma experimentação real foi realizada para comprovar o alinhamento espaço-temporal utilizando um equipamento GPS em conjunto com a informação visual.

Os resultados obtidos nas simulações e testes laboratoriais foram satisfatórios, testando ambos os métodos de alinhamento temporal e espacial separadamente, obtendo resultados de até 98% de precisão. Já os experimentos reais foram obtidos com um sucesso de 5 em 7 aplicações, sendo que as falhas ocorreram principalmente devido a limitações na utilização do GPS.

5.1 Limitações

As principais falhas ocorridas no experimento real aconteceram devido à falta de uma frequência maior de aquisição por parte dos dados de GPS. Nestes casos, o intervalo entre os dados de posição chegou a ser até de 11 segundos, intervalo onde podem ocorrer muitas variações aleatórias no movimento do objeto, sendo impossível de prever realizando apenas a interpolação dos dados.

A falta de um ambiente amplo para que fosse possível realizar testes com GPS também foi considerada uma limitação. Devido a este motivo, os movimentos aleatórios gerados pelo objeto monitorado estavam limitados a poucas direções e ocorreram frequentes interrupções devido a presença de outros movimentos detectados não desejados.

5.2 Extensões

Os mesmos experimentos realizados neste trabalho podem ser estendidos para outros ambientes e com novos equipamentos. A utilização de uma rede de sensores especializada em detectar movimentos pode substituir o GPS por utilizar dados semelhantes. Diferentes sensores que obtenham informações com a possibilidade de serem tratadas e utilizadas para detectar movimento também podem ser utilizados.

Casos práticos como a aplicação em sistemas de segurança e geração de imagens cinematográficas são sugestões de aplicabilidade, além da utilização do resultado de alinhamento espaço-temporal para a reconstrução de *time-lapses* e geração de imagens tridimensionais quando houver câmeras calibradas no aparato.

5.3 Trabalhos futuros

A realização de um trabalho com a utilização de diferentes sensores sugere o estudo de diferentes problemas similares ao desta abordagem.

A utilização de câmeras sem sobreposição que se baseiam nas informações de um sensor de posicionamento capaz de monitorar uma ampla área de cobertura é um bom exemplo em que este trabalho pode ser de grande auxílio, principalmente com alinhamento temporal.

Câmeras esparsas, posicionadas com uma diferença considerável entre si, também dão origem a um novo problema a ser estudado, pois certamente a metodologia a ser aplicada na resolução de um problema com estas características necessitará de adaptações.

A reconstrução tridimensional de movimentos, problema a ser estudado envolvendo técnicas de computação gráfica, também pode ser estudada a partir da utilização da metodologia proposta, onde será necessário previamente alinhar os vídeos no tempo e espaço para que estes sejam utilizados como modelo.

5.4 Considerações Finais

Esta dissertação objetivou o estudo da integração de informações visuais e não visuais em um contexto de um ambiente monitorado por uma rede de sensores heterogêneos.

Mostrou-se que é possível alinhar no tempo e espaço informações visuais, utilizando como base o alinhamento temporal entre câmeras e sensores, o que estende a aplicação deste modelo a diferentes casos e problemas.

Por último, espera-se que sejam desenvolvidos novos trabalhos e geradas novas aplicações com a utilização de sensores não visuais em conjunto com câmeras por estes serem hoje mais acessíveis e estarem disponibilizados para sua utilização no dia a dia.

Referências Bibliográficas

PÁDUA, Flávio Luis Cardeal. *Alinhamento Espaço-Temporal de Seqüências de Vídeo Capturadas a Partir de Múltiplos Pontos de Vista*, Tese de Doutorado, Universidade Federal de Minas Gerais, 2005.

SHAKIL, O. *Video Aligment, Literature Survey*. In Proc. Of Citeseer, The University of Texas at Austin, 2005.

CASPI, Y. and Irani, M. (2000). *A Step Towards Sequence-to-Sequence Alignment*. In Proc. IEEE Conference on Computer Vision and Pattern Recognition, volume 2, pages 682–689, Hilton Head Island, South Carolina, 2000.

CASPI, Y. and Irani, M. *Alignment of Non-Overlapping Sequences*. In Proc. IEEE International Conference on Computer Vision, volume 2, Vancouver, Canada, 2001.

CASPI, Y., Simakov, D., and Irani, M. *Feature-Based Sequenceto-Sequence Matching*. In VAMODS (Vision and Modelling of Dynamic Scenes) workshop with ECCV, Copenhagen, Denmark, 2002a.

CASPI, Y. and M. Irani. *Spatio-temporal alignment of sequences*. IEEE Trans. Pattern Anal. Machine Intell, 2002b

UKRAINITZ, Y. and M. Irani. *Aligning sequences and actions by maximizing space-time correlations*. In European Conference on Computer Vision (ECCV), Graz, 2006.

STEIN, G. *Tracking from Multiple View Points: Self-Calibration of Space and Time*. In DARPA Image Understanding Workshop, pages 521–527, Monterey, Canada, 1998.

WOLF, L. and Zomer, A. *Correspondence-Free Synchronization and Reconstruction in a Non-Rigid Scene*. In Workshop on Vision and Modelling of Dynamic Scenes, Copenhagen, Denmark, 2002.

BRITO, D., Pádua, F., Pereira, G., Carceroni, R. *Synchronizing video cameras with non-overlapping fields of view*. In: Brazilian Symp. on Computer Graphics and Image Processing. pp. 37–44, 2008.

VEDULA, S., Baker, S., and Kanade, T. *Spatio-Temporal View Interpolation*. In Proc. Eurographics Workshop on Rendering, pages 65–76, 2002.

HU, W., Tan T., *Fellow, IEEE*, Wang L., Maybank, S. *A Survey on Visual Surveillance of Object Motion and Behaviors*. In Proc. Of IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews, VOL. 34, N 3, 2004.

MOESLUND, T. B., Hilton A., Kruger, V. *A survey of advances in vision-based human motion capture and analysis*. Computer Vision and Image Understanding 104 (2006) 90–126, 2006.

MAHLISCH, M.; Schweiger, R.; Ritter, W.; Dietmayer, K.; , *Sensorfusion Using Spatio-Temporal Aligned Video and Lidar for Improved Vehicle Detection*. Intelligent Vehicles Symposium, 2006 IEEE.

LEDESMA-CARBAYO, M.J.; Kybic, J.; Desco, M.; Santos, A.; Suhling, M.; Hunziker, P.; Unser, M. *Spatio-temporal nonrigid registration for ultrasound cardiac motion estimation*. Medical Imaging, 2005 IEEE.

ZELNIK-MANOR, L. and Irani, M. *Event-Based Analysis of Video*. In Proc. IEEE Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii USA, 2001.

RAO, C., Gritai, A., Shah, M., and Syeda-Mahmood, T. *View-invariant Alignment and Matching of Video Sequences*. In Proc. of IEEE International Conference on Computer Vision, volume 2, pages 939–945, Nice, France, 2003.

SANTOS, G. *Alinhamento Temporal de Seqüências de Vídeo Adquiridas por Câmeras Perspectivas e Catadióptricas*. Tese de Pós-Graduação por Universidade Federal de Minas Gerais (UFMG), 2006.

MUNOZ-SALINAS, R., Aguirre, E., Garcia-Silvente, M. *People detection and tracking using stereo vision and color*. Image and Vision Computing, Volume 25, Issue 6, 1, Pages 995-1007, ISSN 0262-8856, 2007.

CHOI, J., W. J. Jeon, & S.-C. Lee. *Spatio-temporal pyramid matching for sports videos*. In MIT '08: Proceedings of the 1st ACM international conference on Multimedia information retrieval (pp. 291-297): New York, NY: ACM, 2008.

BERTOZZI, M., A. Broggi, C. Caraffi, M. Del Rose, M. Felisa, G. Vezioni. *Pedestrian detection by means of far-infrared stereo vision*, Computer Vision and Image Understanding, Volume 106, Issues 2-3, Special issue on Advances in Vision Algorithms and Systems beyond the Visible Spectrum, Pages 194-204, ISSN 1077-3142, 2007

NAKAMURA, E. F., Nakamura, F. G., Figueiredo, C. M., and Loureiro, A. A. *Detecção de falhas em redes de sensores sem fio baseada na medição do tráfego e em técnicas de fusão de dados*. In 23o. Simpósio Brasileiro de Redes de Computadores (SBRC 2005), pages 579–592, Fortaleza, Brazil. SBC, 2005.

NAKAMURA, E. *Fusão de Dados em Redes de Sensores Sem fio*. Tese de Doutorado, UFMG, 2007.

LOUREIRO, Antonio A., José Marcos S. Nogueira, Linnyer B. Ruiz, and Raquel A. Mini. *Rede de sensores sem fio*. pages 193-234. XXI Jornada de Atualização em Informática do Congresso da Sociedade Brasileira de Computação, 2002.

HALL, David L. and James Llinas. *An introduction to multi-sensor data fusion*. *Proceedings of the IEEE*, 1997.

HALL, David L. and J. Llinas, editors. *Handbook of Multisensor Data Fusion*. CRC Press, 2001.

CONCI, Aura. et all. *Computação Gráfica, Teoria e Prática volume 2*. 2008, Elsevier Editora Ltda.

MAINBERGER, Markus et all. *Image Analysis and Recognition*, Lecture Notes in Computer Science. 2008, Springer Berlin / Heidelberg.

WANG WEI, Hong Jun, Tang Yiping. *Image Matching for Geomorphic Measurement Based on SIFT and RANSAC Methods*. Computer Science and Software Engineering, 2008 International Conference on.

MARTÍNEZ-GÓMEZ, Jesús et all. *Recognizing Patterns in Signals, Speech, Images and Videos*. 2010, Springer Berlin / Heidelberg.

CONSTANTINOPOULOS, C. Meinhardt-Llopis, E. Liu, Y. Caselles, V. , *A robust pipeline for logo detection*. Multimedia and Expo (ICME), 2011 IEEE International Conference on.

DING, Yuanyuan. Yu, Jingyi. *Computer Vision – ACCV 2007*. 2007, Springer Berlin / Heidelberg.