

# UMA METODOLOGIA DE ODOMETRIA VISUAL/INERCIAL E SLAM 3D COM UM VANT

LUIZ G. B. MIRISOLA\*<sup>†</sup> JORGE DIAS<sup>†</sup> SAMUEL SIQUEIRA BUENO\*

\**Divisão de Robótica e Visão Computacional - DRVC*  
*CTI - Centro de Tecnologia da Informação Renato Archer*  
*Campinas, SP, Brasil*

<sup>†</sup>*Instituto de Sistemas e Robótica*  
*Univ. de Coimbra, Portugal*

Email: {lgm,jorge}@isr.uc.pt, samuel.bueno@cti.gov.br

**Abstract**— This article shows a new methodology for visual-inertial odometry and 3D SLAM in aerial robotics based on a on-board monocular camera rigidly mounted with a *Attitude Heading Reference System* (AHRS) which measures orientation. A joint calibration routine permits to use the AHRS to compensate the rotational motion, generating virtual images that simulate pure translational movement. Then, the translational motion is recovered by direct registration in the scene coordinate system, facilitating vision based navigation over horizontal flat terrain, and obtaining more accuracy than image only methods. This methodology was implemented in a tele-operated airship, and the experimental results of visual-inertial odometry are shown and compared with GPS data. Moreover, by recognizing images of already visited regions, and by utilizing a open source SLAM package based on 3D constraint graph optimization, the odometry drift is corrected, generating a visual-inertial SLAM methodology and showing its results. The conclusions highlight the current limitations and the needed improvements to implement aerial SLAM in real-time.

**Keywords**— UAV, Visual Odometry, AHRS, SLAM

**Resumo**— Este artigo apresenta uma nova metodologia de odometria visual-inercial e SLAM 3D em robótica aérea baseada no uso embarcado de uma camera monocular rigidamente acoplada a um *Attitude Heading Reference System* (AHRS) que fornece medidas de orientação. A calibração conjunta do sistema permite que o AHRS seja usado para compensar o movimento rotacional, gerando imagens virtuais que simulam movimento puramente translacional. Dessa forma, o movimento translacional entre duas imagens é recuperado por registro direto no sistema de coordenadas da cena, facilitando a navegação baseada em visão sobre terreno plano horizontal, e alcançando maior precisão do que métodos que utilizam apenas as imagens. Essa metodologia foi implantada em um dirigível tele-operado, e os resultados experimentais de odometria visual-inercial são apresentados e comparados com dados de GPS. Além disso, o reconhecimento de imagens de regiões já visitadas, e a utilização de um pacote de código aberto para SLAM, baseado em otimização em grafos de restrições 3D, permite corrigir a deriva da odometria, gerando uma metodologia de SLAM visual-inercial cujos resultados são, também, apresentados. As conclusões destacam as limitações atuais e as melhorias necessárias para implementar SLAM aéreo em tempo real.

**Palavras-chave**— VANT, Odometria Visual, AHRS, SLAM

## 1 Introdução

É crescente o interesse em Veículos Aéreos Não Tripulados (VANTs) pelas suas diferentes aplicações civis e militares. Em qualquer veículo robótico, um aspecto crucial consiste na determinação da posição do veículo, e no caso aéreo o movimento tridimensional dificulta esta tarefa. Usualmente são utilizados sistemas de navegação baseados em receptores *Global Positioning System* (GPS) e um sistema inercial, mas há risco de oclusão dos satélites, e a incerteza do GPS é significativa em alturas muito baixas ou durante decolagens e pousos. Além disso estes sistemas não utilizam ou geram informação sobre o ambiente. Sistemas de visão apresentam grande potencial de percepção do ambiente, e a utilização de visão para a determinação de posição e orientação 3D ainda é um tema de pesquisa.

Neste contexto, o presente artigo combina técnicas de visão computacional com medidas de orientação de um *Attitude Heading Reference System*

(AHRS) para determinar a pose relativa entre vistas sucessivas tomadas por um observador em movimento sobre um terreno plano, como um VANT.

Um AHRS moderno fornece medidas de orientação geo-referenciadas, combinando acelerômetros que medem a direção da gravidade e magnetômetros que medem o campo magnético da Terra. A câmera é montada rigidamente com o AHRS, e, para determinar a sua orientação, a transformação entre os sistemas de coordenadas da câmera e do AHRS deve ser determinada. Esta transformação é estimada por uma rotina de calibração (Lobo and Dias, 2007), eliminando a necessidade de uma montagem mecânica precisa.

Para uma seqüência de imagens de uma área plana, a transformação que relaciona as coordenadas de *pixels* correspondentes em duas imagens é uma homografia. Com a rotação compensada, os *pixels* correspondentes podem ser projetados no plano do chão, resultando em conjuntos de pontos correspondentes neste plano, que são registrados calculando-se a transformação rígida que registra

estes conjuntos de pontos diretamente no sistema de coordenadas da cena.

Depois de recuperar-se outras restrições de pose relativa para “fechar o *loop*”, uma técnica de SLAM (*Simultaneous Localization and Mapping*) baseada em otimização em grafos 3D é aplicada para mitigar a deriva da odometria visual. O artigo inclui recuperação de trajetórias do dirigível tele-operado mostrado na figura 1, que carrega uma câmera rigidamente montada com um AHRS, e um receptor GPS. Embora apenas a altura relativa possa ser recuperada a partir de imagens e uma medida da altura inicial seja utilizada para recuperar a escala métrica, estes métodos não necessitam de medidas adicionais de altura para recuperar a trajetória 3D.



Figura 1: O dirigível tele-operado com o AHRS, câmera e GPS montados na sua gôndola.

### 1.1 Trabalhos Relacionados

Pode-se construir um mapa 3D do ambiente a partir de imagens estéreo tomadas por um dirigível teleoperado (Hygounenc et al., 2004), mantendo-se estimativas da posição da câmera e de marcos detectados automaticamente nas imagens. A trajetória de um VANT também pode ser recuperada registrando-se conjuntos sucessivos de pontos 3D triangulados para cada imagem estéreo (Kelly et al., 2007), embora o tamanho da linha base estéreo limite a altura do VANT a alguns metros.

Medidas de sensores inerciais e observações de marcos artificiais podem ser fundidas, realizando SLAM e compensando a imprecisão das medidas de altitude do GPS (Kim, 2004).

A trajetória de um veículo aéreo pode ser recuperada a partir de imagens de uma superfície plana horizontal usando casamento de pontos de interesse, com o auxílio de um altímetro para resolver a escala (Conte and Doherty, 2008). Variações de algoritmos de pontos de interesse baseadas em blobs ou clusters também foram utilizadas com imagens aéreas, construindo um mosaico de imagens que por sua vez é usado para navegação (Caballero et al., 2006).

O método proposto de odometria visual usa apenas medidas de orientação de um AHRS e *pixels* correspondentes fornecidos pelo algoritmo SURF (Bay et al., 2006) detectados em imagens de uma câmera monocular, portanto não é limi-

tado por uma linha base estéreo e não depende de marcos artificiais. O movimento da câmera é livre, desde que a sua orientação seja medida. O GPS é utilizado apenas para comparação.

A reprojeção das imagens em um plano virtual é bem conhecida (Gracias, 2002; Brown and Sullivan, 2002), mas o modelo de homografias continua sendo usado. Modelos de translação pura parecem ser mais adequados para estimar o movimento vertical. Além disso, a determinação da translação (exceto a escala) tem uma solução única. Em contraste, com o modelo de homografias, a matriz recuperada deve ser decomposta, resultando em quatro soluções das quais apenas uma é correta (Caballero et al., 2006). Adicionalmente, o passo de otimização incluído na estimação de homografias é eliminado, diminuindo o custo computacional em cenários realistas embora a remoção de correspondências aberrantes seja mais demorada. Embora existam resultados preliminares com imagens aéreas e um modelo de translação (Michaelsen et al., 2004), resultados extensivos ainda não existem na literatura.

O restante do artigo está organizado da seguinte forma: A próxima seção apresenta o método de odometria visual-inercial, implementado na plataforma experimental apresentada na seção 3 com resultados apresentados na seção 4. A seção 5 apresenta o método de SLAM e o seu resultado, e finalmente a seção 6 apresenta as conclusões.

## 2 Método de Odometria Visual-Inercial

### 2.1 O conceito do plano virtual horizontal

A medição da orientação da câmera permite que a imagem seja reprojeta em entidades definidas em um sistema de coordenadas geo-referenciado, como um *plano horizontal virtual* (com normal paralela à gravidade), à uma distância  $f$  abaixo do centro da câmera, como mostrado na figura 2. Esta reprojeção é calculada por uma homografia infinita (Ma et al., 2004), e corresponde à imagem de uma câmera virtual com o mesmo centro óptico da câmera original, mas com eixo óptico coincidente com o vetor gravidade. A reprojeção compensa diferenças devidas à orientação e ponto de vista, mas não altura, resultando em diferenças de escala nas imagens.

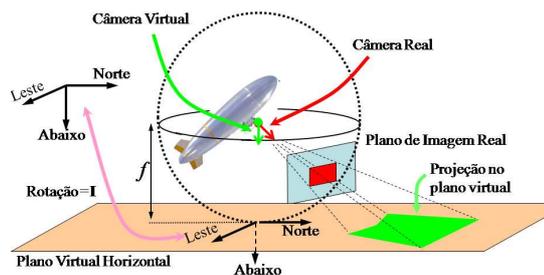


Figura 2: O conceito do plano virtual horizontal.

## 2.2 Registrando conjuntos de pontos no plano

Tome uma seqüência de imagens aéreas de uma área horizontal plana, que foram reprojetaadas no plano virtual horizontal. As câmeras virtuais têm planos de imagem horizontais e paralelos ao plano do chão. Dadas duas vistas sucessivas, tomadas nos instantes  $i$  e  $i + 1$ , e coordenadas de *pixels* correspondentes entre elas, a pose relativa entre as duas posições de câmera será recuperada como uma translação  $\mathbf{t} = (t_x, t_y, t_z)^T$ .

Cada *pixel* entre as correspondências é projetado no plano do chão gerando um ponto 3D, como mostrado na figura 3(a). Dois conjuntos de pontos são gerados para as duas vistas sucessivas, e estes conjuntos são registrados diretamente no sistema de coordenadas 2D da cena por uma rotina de Procrustes (Gower and Dijksterhuis, 2004), utilizando RANSAC (Fischler and Bolles, 1981) para remover correspondências aberrantes. A figura 3(b) mostra um diagrama deste processo.

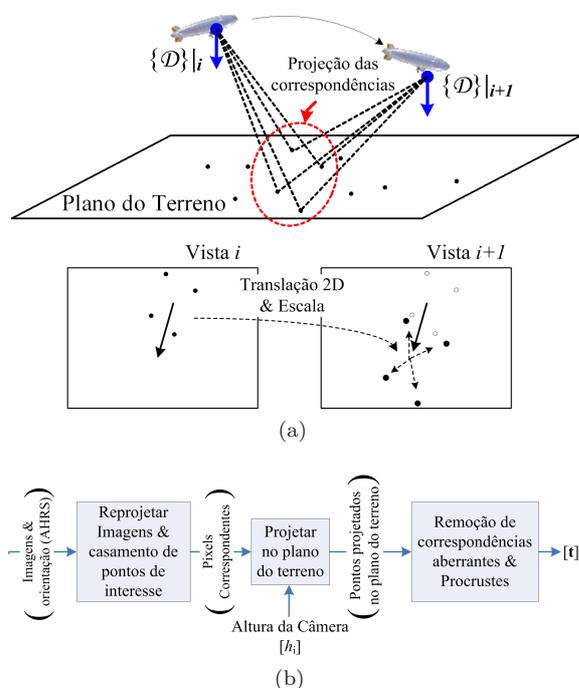


Figura 3: Calculando a translação entre posições de câmera sucessivas por registro de pontos de interesse projetados no plano do chão.

A rotina de Procrustes encontra a translação 2D e a escala que registram os dois conjuntos de pontos, resultando em  $t_x$  e  $t_y$  e na escala  $s$  que representa a altura relativa entre a primeira e segunda câmeras. O componente  $t_z$  é calculado pela equação  $s = h_{i+1}/h_i = (h_i + t_z)/h_i$ , onde  $h$  representa a altura do VANT no instante indicado pelo subscrito. Se as restrições de ter os planos imagem e do chão horizontais são satisfeitas, com correspondências aberrantes removidas, e com igual peso à todas as coordenadas dos *pixels*, então pode-se provar que a rotina encontra a

solução que minimiza o erro quadrado médio.

## 3 Plataforma Experimental e Calibração

O GPS é um Garmin GPS35. O AHRS é um Xsens MTi (XSens Tech., 2007). O fabricante declara que o desvio padrão das medidas de orientação é  $1^\circ$  se o sensor está parado. Este erro deve ser maior no veículo em movimento. A gôndola do dirigível possui motores de combustão e estrutura de alumínio, logo não há interferência magnética sobre a bússola do AHRS.

A câmera é uma Point Grey Flea (Point Grey Inc., 2007), mostrada na figura 1 montada rigidamente com o AHRS. Os seus parâmetros intrínsecos e de distorção das lentes são calibrados primeiramente (Bouguet, 2006). Outra calibração subsequente (Lobo and Dias, 2007), *off-line*, determina a rotação entre os sistemas de coordenadas da câmera e AHRS.

## 4 Odometria Visual para um VANT

As próximas experiências utilizam imagens tomadas pelo dirigível tele-operado da figura 1, que carrega o sistema AHRS-câmera e o receptor GPS, voando sobre uma área plana no Aeródromo de Coimbra. Dados e vídeos dos ensaios estão disponíveis (Mirisola and Dias, 2008).

Objetos no chão com dimensões conhecidas estão visíveis na primeira imagem. As suas dimensões, reais e na imagem, foram medidas e a partir delas a altura inicial do dirigível foi calculada como  $h_1 = 25m$ .

A trajetória recuperada é mostrada em 2D e 3D na figura 4. A trajetória medida pelo GPS é indicada pelos pequenos círculos, com círculos maiores indicando um desvio padrão (por volta de  $8m$ ). O GPS indica um deslocamento de  $543m$ , com velocidade média de  $9m/s$ . Um filtro de Kalman com um modelo de aceleração de processo de Wiener discreto (Bar-Shalom et al., 2001) foi utilizado para filtrar os vetores de translação para todos os métodos.

Os quadrados na figura 4 mostram a trajetória recuperada ao se usar os vetores de translação 3D estimados pela rotina de Procrustes como entrada para o filtro de Kalman. Os asteriscos indicam a trajetória recuperada pelo modelo de homografias usando as imagens originais, e não utilizando dados de AHRS, imagens reprojetaadas, ou a rotina Procrustes. Como apenas a razão  $\mathbf{t}/h$  é recuperada pelas homografias, a escala é recuperada multiplicando-se o vetor  $\mathbf{t}/h$  pela altura estimada do dirigível. A posição do dirigível incluindo a altura é estimada pelo filtro de Kalman usando estes vetores de translação como entrada. As tabelas 1 e 2 mostram os erros e tempo de execução da odometria visual.

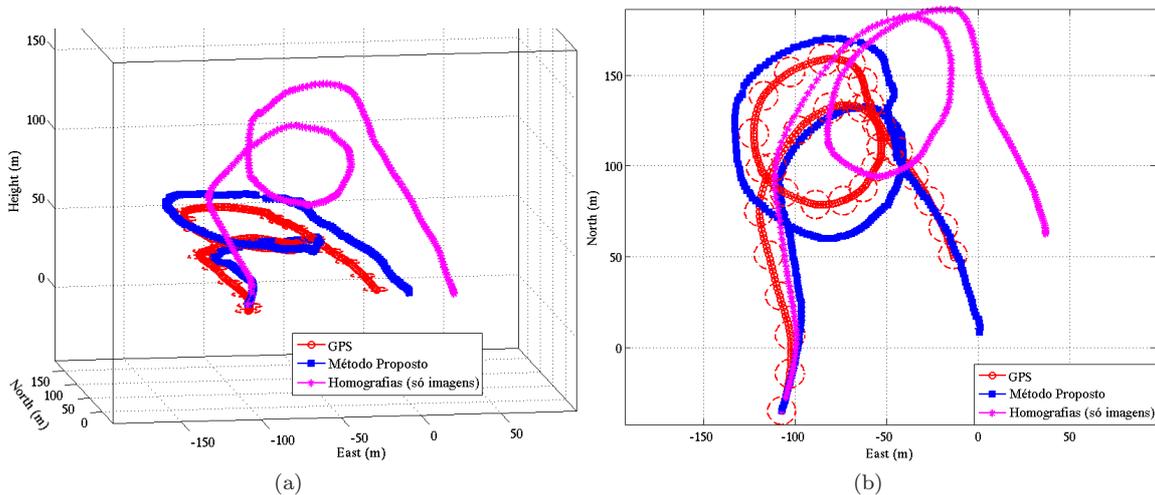


Figura 4: Gráficos 3D (a) and 2D (b) da trajetória recuperada pela odometria visual-inercial comparadas com dados de GPS. Os círculos indicam o erro esperado fornecido pelo GPS (um desvio padrão).

Unidade (s)	RANSAC		Otimiz.		Total
	Méd.	std.	Méd.	std.	
Homografia	0.05	0.02	0.43	0.19	0.48
Odometria Visual com AHRS	0.18	0.08	-	-	0.18

Tabela 1: Tempo de execução médio das técnicas de odometria visual, em segundos.

## 5 SLAM: “fechando o loop”

Quando uma área é vista por uma segunda vez pela camera embarcada, medidas de pose relativa entre vistas não-consecutivas são usadas para “fechar o loop”. Um grafo é construído onde os vértices representam as posições 3D da câmera e as arestas as restrições de pose relativa (na forma de vetores de translação). Otimizando as posições dos vértices para minimizar o erro nas restrições, pode-se mitigar a deriva da odometria visual. A figura 5 mostra a mesma trajetória da figura 4 com as restrições não-consecutivas indicadas pelas setas verdes. Cada seta representa uma medida de pose relativa entre o vértice na sua base e o vértice conectado à sua ponta por uma linha colorida. Portanto as linhas coloridas representam o erro relacionado à esta restrição, que deve ser minimizado.

Para detectar as restrições não-consecutivas, o ponto principal de cada imagem é projetado no plano do chão e armazenado. Então, para cada imagem, a projeção do seu ponto principal é comparada com a base de dados, procurando por uma imagem anterior da mesma área. A imagem correspondendo ao ponto principal mais próximo armazenado é escolhida como uma candidata para casamento de pontos de interesse, excluindo da procura os últimos 5 s do voo e pontos principais que estão muito longe. Então se o casamento de

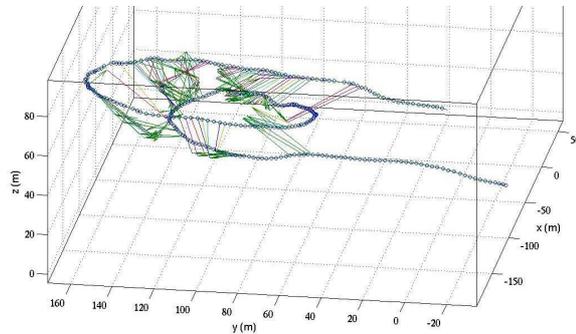


Figura 5: A trajetória do dirigível com as restrições não-consecutivas indicadas pelas setas. As linhas representam o erro a ser minimizado.

pontos de interesse entre este par de imagens tem sucesso a pose relativa é recuperada e uma nova aresta adicionada ao grafo.

Para otimizar o grafo de restrições, o pacote de código aberto TORO (Grisetti et al., 2007) foi utilizado com algumas modificações: (1) Como a pose relativa é um vetor de translação cuja magnitude depende da estimativa da altura da câmera, as covariâncias de todas as restrições são definidas como proporcionais à altura da câmera. (2) Pela mesma razão, depois de cada iteração cada vetor de translação é recalculado levando em conta a nova altura do vértice correspondente. (3) As covariâncias das restrições não-consecutivas foram multiplicadas por quatro: o erro na orientação fornecida pelo AHRS tende a ser correlacionado entre instantes sucessivos, e como no caso das restrições não consecutivas isto não ocorre a sua incerteza se torna maior.

A figura 6 mostra a trajetória do dirigível recuperada pela odometria visual-inercial antes e depois da correção. A trajetória do GPS é fornecida para comparação. Algumas poucas restrições

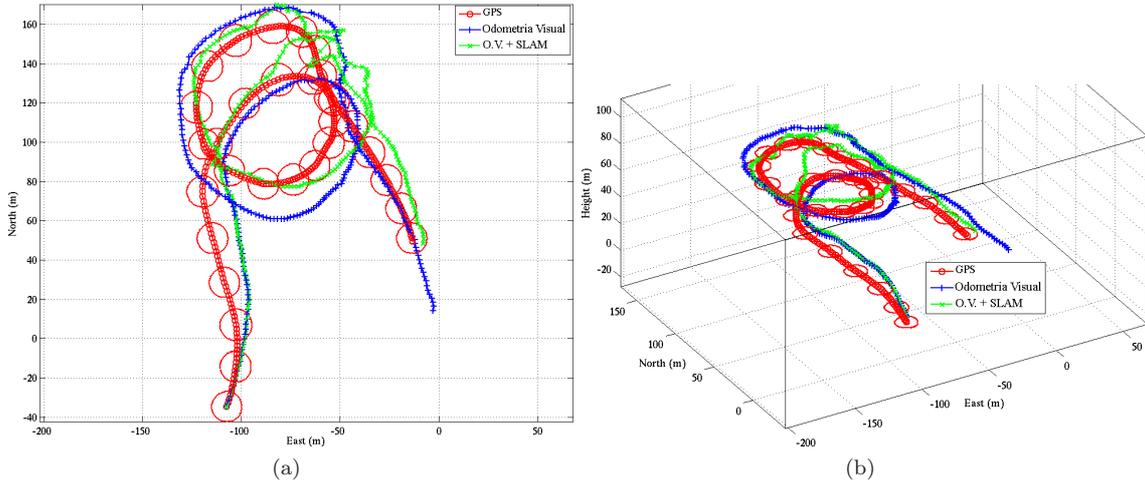


Figura 6: A trajetória do dirigível recuperada pela odometria visual antes e depois da correção pelo pacote de otimização em grafos.

Unidade: metros	Erro de posição 3D			Erro de posição 2D			Erro no comprimento de $t$		
	Médio.	Max.	Final	Médio.	Max.	Final	RMS	Médio.	Max.
Homografias (apenas imagens)	57.5	98.2	52.3	43.1	67.6	52.0	0.88	-0.63	1.24
Odometria Visual c/ AHRS	16.5	43.6	43.6	15.3	43.3	43.3	0.73	-0.46	0.59
Odometria Visual c/ AHRS depois da otimização do grafo	13.5	25.2	5.8	11.7	23.5	5.2	1.08	-0.57	1.1

Tabela 2: Comparação dos métodos de odometria visual (o GPS é a referência). Valores em metros.

não-consecutivas erradas distorceram o lado nordeste das duas voltas, mas as próximas restrições corrigiram a trajetória, aproximando-a da trajetória de referência. A deriva da odometria visual foi mitigada. A Tabela 2 compara ambas as trajetórias. O erro final em 3D da trajetória recuperada após a correção está dentro da margem de erro do GPS. Necessitou-se de  $0.17s$  para executar 100 iterações da otimização no grafo completo.

## 6 Conclusões

Neste artigo, um modelo de translação pura é usado para recuperar a trajetória de um VANT usando medidas de orientação de um AHRS e imagens aéreas do terreno horizontal. Um pacote de otimização em grafos para SLAM é usado para mitigar a deriva da odometria.

O modelo proposto apresenta melhores resultados na recuperação do movimento vertical do que o método que utiliza apenas as imagens. O movimento vertical é mais crítico porque erros na altura estimada se propagam não apenas no eixo vertical, mas também como um erro na escala dos componentes horizontais. Este fato é inerente à recuperação de pose relativa a partir de imagens. Como a incerteza do GPS é maior no eixo vertical do que nos horizontais, a estimação da altura é ainda mais crítica, especialmente em alturas baixas, como durante aterrisagens e decolagens.

Uma trajetória de centenas de metros com

movimento vertical significativo foi recuperada, mas o limite da técnica depende da precisão e confiabilidade das medidas de orientação e da detecção de restrições não-consecutivas em condições de voo. Os resultados experimentais indicam que as estimativas de orientação de um AHRS permitem estimar a translação com precisão suficiente.

A otimização de grafos 3D deve considerar que a magnitude da translação entre duas posições da câmera depende da altura da câmera, e atualizar as restrições adequadamente quando uma posição da câmera é recalculada, caso contrário não é possível corrigir o comprimento da trajetória. Isto pode levar a divergência quando um vértice é atualizado em uma posição mais alta, levando a um vetor de translação maior, que por sua vez eleva a altura de outro vértice, e assim até o grafo divergir. É necessário estudar quais configurações do grafo garantem a convergência da otimização.

Trabalhos futuros podem explorar modelos de erro mais elaborados, por exemplo estimar a incerteza das coordenadas de cada pixel, e propagar esta incerteza para a pose relativa. Variantes do problema Procrustes com diversos modelos de incerteza foram resolvidos, embora alguns casos necessitem soluções iterativas (Gower and Dijkstra, 2004). Outros métodos podem ser empregados para detectar restrições e “fechar o loop” com maior confiabilidade. Por exemplo, bases de dados de descritores de pontos de interesse podem ser usadas para encontrar imagens anteriores

da mesma área mesmo que a trajetória estimada tenha derivado se afastando muito da posição verdadeira. Estruturas de dados como árvores k-d podem garantir que descritores sejam encontrados rapidamente.

O registro de Procrustes e o filtro de Kalman são rotinas algébricas rápidas. A otimização do grafo pode ser executada em paralelo quando uma restrição não-consecutiva é encontrada, ou pode ser intercalada executando-se algumas iterações após processar cada imagem. Estas operações poderiam ser executadas em tempo real. A única operação necessariamente custosa computacionalmente é o casamento de pontos de interesse, incluindo a detecção de correspondências aberrantes e de restrições não consecutivas. Usar as imagens reprojadas acelera o casamento de pontos de interesse, mas não o suficiente. O mínimo número de pontos de interesse necessário deve ser estabelecido, e parâmetros adaptativos nos algoritmos poderiam procurar alcançar este número. Além disso, métodos de registro de imagens que não utilizam pontos de interesse podem ser explorados.

## 7 Agradecimentos

O projeto recebeu financiamento da FCT/Portugal (processo BD/19209/2004).

## Referências

- Bar-Shalom, Y., Li, X. R. and Kirubarajan, T. (2001). *Estimation with Applications to Tracking and Navigation*, J. Willey & Sons Inc.
- Bay, H., Tuytelaars, T. and van Gool, L. (2006). SURF Speeded Up Robust Features, *European Conf. on Comp. Vision*, Graz, Austria.
- Bouguet, J. (2006). Camera calibration toolbox for matlab, [http://www.vision.caltech.edu/~bouguetj/calib\\_doc/index.html](http://www.vision.caltech.edu/~bouguetj/calib_doc/index.html).
- Brown, A. and Sullivan, D. (2002). Precision kinematic alignment using a low-cost GPS/INS system, *ION GPS*, Portland, OR, USA.
- Caballero, F., Merino, L., Ferruz, J. and Ollero, A. (2006). Improving vision-based planar motion estimation for unmanned aerial vehicles through online mosaicing, *IEEE Int. Conf. on Robotics and Automation (ICRA06)*, Orlando, FL, USA, pp. 2860–2865.
- Conte, G. and Doherty, P. (2008). An integrated UAV navigation system based on aerial image matching, *IEEE Aerospace Conference*, Big Sky, MT, USA.
- Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, *Comm. of the ACM* **24**: 381–395.
- Gower, J. C. and Dijksterhuis, G. B. (2004). *Procrustes Problems*, Oxford Statistical Science Series, Oxford University Press.
- Gracias, N. (2002). *Mosaic-based Visual Navigation for AUV*, PhD thesis, Instituto Superior Técnico, Lisbon, Portugal.
- Grisetti, G., Grzonka, S., Stachniss, C., Pfaff, P. and Burgard, W. (2007). Efficient estimation of accurate maximum likelihood maps in 3D, *IEEE Int. Conf. on Intelligent Robots and Systems (IROS)*, San Diego, CA, USA.
- Hygounenc, E., Jung, I.-K., Soueres, P. and Lacroix, S. (2004). The Autonomous Blimp Project at LAAS/CNRS: Achievements in Flight Control and Terrain Mapping, *Int. J. of Robotics Research* **23**(4/5): 473–512.
- Kelly, J., Saripalli, S. and Sukhatme, G. S. (2007). Combined visual and inertial navigation for an UAV, *6th Int. Conf. Field and Service Robotics (FSR'07)*, Chamonix, France.
- Kim, J. (2004). *Autonomous Navigation for Airborne Applications*, PhD thesis, The University of Sydney, Sydney, NSW, Australia.
- Lobo, J. and Dias, J. (2007). Relative pose calibration between visual and inertial sensors, *Int. J. of Robotics Research* **26**(6): 561–575.
- Ma, Y., Soatto, S., Kosecka, J. and Sastry, S. (2004). *An Invitation to 3D Vision*, Springer.
- Michaelsen, E., Kirchhof, M. and Stilla, U. (2004). Sensor pose inference from airborne videos by decomposing homography estimates, *XXth ISPRS Congress*, Istanbul, Turkey.
- Mirisola, L. G. B. and Dias, J. (2008). Datasets and Complementary results, <http://paloma.isr.uc.pt/~lgm/>.
- Point Grey Inc. (2007). [www.ptgrey.com](http://www.ptgrey.com).
- XSens Tech. (2007). [www.xsens.com](http://www.xsens.com).