Reconstrução de cenas e objetos em 3D a partir imagens

Marcio Calixto Cabral

Junho 25, 2002

Relatório de Estudo para Tópicos em Ciência da Computacao
MAC5701

Referências Bibliográficas

[1] Aloimonos, John et al. , "A Multi-frame Approach to Visual Motion Perception" *Internacional Journal of Computer Vision, 6:3, 245-255*, Received June 26, 1989.

[2] Shashua, A "Projective Structure from Uncalibrated Images: Structure-from-Motion and Recognition" *"PAMI"16, 8,778-790* August, 1994

[3] Marc Pollefey, "Tutorial on 3D Modeling from Images" *In conjunction with ECCV2000* 26 June 2000, Dublin, Ireland

[4] Shashua, A et al. "Trilinearity of Three Perspective Views and its Associated Tensors" *"IEEE"920-925* 1995

[5] O.D. Faugeras. "Stratification of three-dimensional vision: projective, affine and metric representations" *Journal of the Optical Society of America, 12(3):465–484* 1995.

[6] Schweitzer, H et al. "Structure from Multiple 2D Affine Correspondences without Camera Calibration" *"IEEE" 258-263* 1996

[7] J.-Y. Bouguet and P. Perona, "3D Photography on your Desk". *Proc. 6th Int. Conf. Computer Vision, Bombay, India, pages 43-50* January 1998.

[8] P. Debevec, C. Taylor and J. Malik, "Modeling and Rendering Architecture from Photographs: A Hybrid Geometry- and Image-Based Approach" *Siggraph* 1996

Sumário

Capítulo 1

Introdução

A necessidade de representação coerente e consistente para modelos 3D é requerida por um número grande de aplicações hoje em dia como modelagem, simulação, visualização e CAD/CAM. O hardware para visualização desses modelos tem evoluído rapidamente o que causa a necessidade de modelos com alta complexidade poligonal e mais realistícos. Mesmo com a disponibilidade de aplicativos de modelagem eficientes, o tempo necessário para criar e construir tais modelos e cenas é grande, e com isso, os custos envolvidos tornam-se grandes e inibitivos.

Quase que a maioria dos objetos modelos atualmente são representações virtuais de objetos reais que fazem parte da nossa vida cotidiana. Se fosse possível extrair automáticamente a informação 3D contida intrinsicamente em uma cena particular através de fotografias dela, muito tempo e trabalho custosos seriam poupados.

Extrair informações de profundidade através de múltiplas imagens (fotografias) não é uma tarefa fácil. Calibração de camera, correspondência entre imagens, reconstrução 3D da cena são alguns dos problemas enfrentados pelos pesquisadores.

Duas técnicas tem sido mais usadas em detrimento de outras: Aquisição passiva e ativa. A primeira é mais flexível e não tão cara quanto a segunda, mas é computacionalmente mais cara. A última é baseada em ambientes controlados de luz como luz direta ou feixe de laser. Com isso, a aplicabilidade é restrita mas simplifica o problema.

Há uma demanda importante para flexibilidade na aquisição. Calibração e ambientes restritos não devem ser usados, ou então restritos para o mínimo possível. Muitas aplicações requerem sistemas de baixo custo robustos que estimulam o uso de cameras não-profissionais para aquisição de imagens.

Capítulo 2

Breve Histórico

A extração de informação de profundidade através de imagens tem sido investigada a alguns anos. Um dos primeiros trabalhos relacionados a extração de profundidade para reconstrução de modelos foi introduzida por John Aloimonos [1] no ano de 1989. O método proposto foi o de extração de características particulares de uma sequência de imagens para estimar o movimento de camera.

Como este paper mesmo sugere, no princípio os pesquisadores estavam mais voltados no uso de informação 3D adquirida através de imagens para visão robótica. Inspeção visual é necessária para neste caso para que um robô possa decidir se ele é capaz de alcançar um objeto do mundo e consequentemente problemas relacionados a pertinência aparecem. Atualmente o foco está mudando. Representações de modelos 3D mais realistas de objetos reais são necessárias em computação gráfica e realidade virtual. Enquanto qualidade visual era deixada de lado no passado, hoje é a principal preocupação.

Outra técnica utilizada baseia-se nas informações contidas nas sombras dos objetos a serem modelados. Este método foi desenvolvido por Bouget and Perona [7]. Proesmans et al. propôs outro método baseado no uso de malhas pré-definidas que são projetadas nas cenas e objetos a serem recuperados. Estes dois métodos são considerados estado-da-arte usando-se construção ativa.

Usando-se métodos baseados em construção passiva, o objetivo é aumentar a automação nas técnicas de reconstrução. O resultado final seria uma cena ou objeto totalmente reconstruídos simplesmente movendo uma camera em volta dele.

Um dos sistemas usando estas técnicas é o *Facade* proposed by Debevec et al. [8]. Este método usa um modelo inicial aproximado para reconstruir a cena.

Outro processo desenvolvido recentemente foi o introduzido por Marc Pollefey [3]. Usando cameras comuns encontradas facilmente no mercado, o seu método consiste em tirar fotos ao redor do objeto a ser modelado. Nenhuma calibração é necessária, o que faz com que este sistema um dos mais fáceis e portáveis de se usar atualmente.

Capítulo 3

Plano de Trabalho

Este o plano de trabalho para a matéria *MAC5701 - Tópicos em Ciência da Computação*. Este é baseado principalmente em artigos científicos, teses e tutoriais publicados ao longo dos últimos anos no assunto. A maioria dos artigos foram selecionadas e indicadas pelo professor Carlos Hitoshi Morimoto e pelo professor Antonio Elias Fabri. Alguns deles foram selecionados por mim e aprovados pelos professores mencionados. Estas referências formam a base de conhecimento para o desenvolvimento de minha tese de mestrado.

As referências selecionados para leitura durante este semestre são apresentadas a seguir:

1. A Multi-frame Approach to Visual Motion Perception by Aloimonos [1]

2. Projective Structure from Uncalibrated Images: Structure from Motion and Recognition [2]

3. Tutorial on 3D Modeling from Images [3]

4. Trilinearity of Three Perspective Views and its Associated Tensors [4]

5. Stratification of three-dimensional vision: projective, affine and metric representations[5]

6. Structure from Multiple 2D Affine Correspondences without Camera Calibration[6]

$Semana$	$3 - Abril$	$4 - Abril$	$1 - Maio$	$2 - Maio$	$3 - Maio$
Ref	$Ref1$	$Ref2$	$Ref3$	$Ref4$	$Ref5$

$Week$	$4 - Maio$	$1 - Junho$	$2 - Junho$	$3 - Junho$
Ref	$Ref6$	$Elabor.$	$Elabor.$	$Elabor.$

Capítulo 4

Resumo dos Artigos

1. `A Multi-frame Approach to Visual Motion Perception by Aloimonos [1]`

Este paper foi publicado em 1991 no *International Journal of Computer Vision* mas ja havia sido escrito por Aloimonos em 1989.

Um dos primeiros trabalhos na área de reconstrução de objetos em 3D a partir de imagens baseava-se na estimação do movimento de uma camera, dadas imagens que retratavam este movimento. Aloimonos introduziu com este artigo uma técnica para uso de várias vistas. Até este momento, as técnicas utilizavam apenas duas vistas ou entao muitos valores sobre a estimação do movimento ou sobre o formato da cena eram fixados no início do algoritmo. Esta técnica é considerada um grande avanço com relação aos algoritmos de duas vistas, além de ser considerada matematicamente elegante

---

Resumo do Artigo Percepção de movimento, até pouco tempo atrás, não se sabia ao certo se existia uma solução ou não. Após a descoberta que soluções existem, muitos pesquisadores começaram a trabalhar pesquisando como encontrá-la e como provar sua unicidade. Um fator crucial que envolve todas as provas apresentadas até este artigo é a sensitividade a ruídos que os algoritmos apresentam devido a quantidade de dados utilizada pelos algoritmos, que é a mínima possível. Algoritmos baseados nestas técnica estão condenados a uma sensitividade muito maior na prática do que a descrita na teoria. Para que isso seja superado pelos algoritmos, mais informação redundante sobre a cena deve ser usada (mais correspondências e mais vistas) e esta deve ser usada otimamente. O algoritmo descrito neste artigo não faz nenhuma assumpção além de rigidez, além de ser simples e robusto.

A rigidez assumida significa que a camera se locomove em um ambiente estacionário e que a movimentação da camera pode ser expressa entre os frames i e $i+1$ através de uma matriz de rotação $R_{i,i+1}^T$ e um vetor de translação $-T_{i,i+1}$. Estes são os parâmetros de movimento de camera que o algoritmo se propõe a recuperar, e consequentemente, com estes parâmetros recuperados, a estrutura do objeto ou da cena.

O algoritmo proposto recebe como entrada um conjunto de pontos de imagem $p_{ij} = [p_{ijx}, p_{ijy}, p_{ijz}]^T$ onde i é o número da vista e j a um ponto desta vista. Como entrada também são fornecidos as correspondências de todos os pontos em todas as vistas (para as vistas onde correspondências existam, isto é, onde não haja oclusão).

A saída do algoritmo é exatamente o conjunto de parâmetros de movimento de camera e o conjunto de pontos 3D que, quando projetados em vistas sucessivas movendo-se de acordo com os parâmetros de movimentação de camera, produz imagens que são as mais próximas das imagens recebidas como entrada.

As vantagens do algoritmo apresentado no artigo são as seguintes:

(a) **Assumpções mínimas** Nenhum algoritmo apresentado até a publicação deste artigo supunha como restrição apenas a rigidez. A movimentação pode ser qualquer não se restringindo a movimentos do tipo *suave, fixado em algum eixo ou etc.*

(b) **Conhecimentos a priori** *Nenhum conhecimento a priori sobre a estrutura de cena ou sobre a movimentação da camera precisa ser dado como entrada. Os únicos dados de entrada são os pontos das imagens e suas correspondências. Outros métodos precisam deste tipo de informação como entrada.*

(c) **Flexibilidade** *Não há restrição no número de vistas a ser usado*

pelo algoritmo ou no número de pontos de imagens dado como entrada (deve-se levar em conta apenas restrições de tamanho de memória para não ocorrer estouro de pilha). Também não restrição com relação as correspondências entre os pontos dados. Ou seja, porblemas de oclusão podem ocorrer, pois os pontos podem estar tanto em uma vista como em todas.

(d) **Elegância Matemática** *O algoritmo não se utiliza de métodos estatísticos como heurísticas para o cálculo dos parâmetros de movimentação de camera.*

(e) **Eficiência** *O algoritmo apresentado converge de maneira rápida e eficiente.*

O algoritmo foi testado em um conjunto de imagens sintéticas. Comparando-se com algoritmos que usam duas vistas, há um aumento da ordem de magnitude em quase duas vezes com relação ao erro ao se estimar os parâmetros de movimentação da camera. O algoritmo ainda é muito suscetível ao ruído da entrada. Mas computacionalmente o algoritmo apresentado é eficiente e a melhora de tempo aumenta polinomialmente da ordem de 4.

2. `Projective Structure from Uncalibrated Images: Structure-from-Motion and Recognition [2]`

Este paper foi publicado no *"PAMI"* em agosto de 1994.

O problema da reconstrução 3D através de duas ou mais vistas, e gerar artificialmente novas vistas é discutido neste artigo. Uma invariante para descrever a estrutura de cena é proposta baseando-se em correspondências de características entre vistas.

Resumo do Artigo Duas grandes áreas em visão computacional têm se beneficiado do tratamento analítico de 3D para 2D: reconhecimento visual e reconstrução através de múltiplas vistas. O caminho clássico trilhado consiste em encontrar uma representação métrica de um mundo 3D através de suas projeções. Muitos problemas estão envolvidos na recuperação desta estrutura. Um deles é a estimação dos parâmetros internos de calibração da camera, principalmente porque muitas vezes os parâmetros estimados apresentam discrepâncua muito grande comparados com os parâmetros fornecidos pelo fabricante. Ao invés de representar o mundo 3D através de descrições métricas, o autor sugere uma representação projetiva, que é apresentada ao longo do artigo e é chamada "projective depth". Para estabelecer uma restrição entre os planos de projeção o autor mostra que duas homografias são necessárias (as tranformações projetivas) entre os dois planos de imagens. A restrição é dada pela seguinte fórmula: $p' = Ap + kEp$, onde p e p' são pontos correspondentes entre duas imagens, A e E são matrizes $3x3$ representando as homografias e k é o invariante. As vantagens desta representação são as seguintes:

- A representação (restrição que mapeia um ponto de uma imagem na outra) depende dos epipolos apenas para extrair as homografias. Isto representa um aumento na robustez pois extrair informação sobre geometria epipolar geralmente é muito suscetível a erro.
- A restrição é representada por um número apenas ao invés de coordenadas.

O algoritmo proposto funciona da seguinte forma:

(a) Recuperar os epipolos v e v'. (Para isso, correspondências adicionais entre as imagens são necessárias)

(b) Recuperar a homografia A que satisfaz $Ap = p'$.

(c) Recuperar a homografia E através das relações $Ep = p'$ e $Ev = v'$.

(d) Escalar E (ou A) para satisfazer $p' = Ap + kEp$.

(e) Para cada correspondência p,p', achar a constante k de "projective depth".

A teoria apresentada por Shashua pode ser usada para dois propósitos: reconstrução e reprojeção. As características importantes deste trabalho são a nova forma introduzida para se trabalhar que é usando um plano de trabalho projetivo e a nova equação para representar movimentação de camera, substituindo a equação clássica.

Os testes foram divididos em dois casos: quando uma configuração coplanar de pontos poderia ser identificada (neste caso, o cálculo da geometria epipolar não é necessário) e um caso onde tal assumpção não pode ser feita. No primeiro caso, a reconstrução da invariante k é razoavelmente acurada comparando-se a reprojeção feita com este k (erro máximo de 1.53 pixels). No segundo caso a acuidade do cálculo depende diretamente do cálculo dos epipolos. A acuidade é razoável mesmo assim, mas não tão boa quanto no primeiro caso. A reprojeção tem erro máximo de 17.8 pixels, mas mesmo assim ainda é melhor do que reprojeção usando intersecção de linhas epipolares. De acordo com o autor, a simplicidade ganha por aumentar o espaço de imagens e de objetos por espaços projetivos, e consequentemente aumentando o grupo de transformações para transformações projetivas, se traduz em aumento de robustez e aplicações práticas.

3. `Tutorial on 3D Modeling from Images [3]`

Este tutorial foi lecionado no ano de 2000 no *ECCV* na Irlanda.

Neste tutorial, Marc Pollefeys explica como um modelo de uma superfície 3D pode ser obtida através de uma sequência de imagens tiradas com cameras convencionais disponíveis comercialmente. O usuário captura essas imagens movendo a camera livremente em volta do objeto e tirando-se as fotos. Nenhum parâmetro de calibração precisa ser conhecido, bem como nenhum conhecimento sobre a movimentação da camera. O sistema proposto e explicado usa cameras perspectivas e nao requer modelos iniciais como entrada.

Visao geral do Tutorial Atualmente a complexidade computacional que cartões gráficos conseguem lidar é muito grande. Isto faz com que modelos que representam objetos reais sejam mais complexos. Mas complexidade alta em modelos significa muito trabalho e consequentemente custo elevado. Dessa forma, aquisição automática de modelos através de objetos reais se tornou algo necessário.

Algumas técnicas baseadas em laser foram desenvolvidas para aquisição desses modelos. Técnicas como essa exigem hardware especializado, calibração do sistema e ambientes específicos e não resolvem o problema de custo elevado para construção de modelos. Novas aplicações baseadas em outras tecnologias são necessárias para redução de custo e automação da tarefa. A técnica apresentada neste tutorial utiliza cameras comuns para obter o modelo 3D do objeto e não requer calibração.

A introdução matemática do tutorial nos uma visão da termologia adotada durante o curso bem como introduz noções de projeção perspectiva, modelo de câmera, parâmetros de calibração intrínsicos e extrínsicos de câmera, transformações projetivas, geometria epipolar, geometria de duas vistas. Com estas informações preliminares, o leitor pode prosseguir a leitura do tutorial.

De acordo com o tutorial temos os seguintes problemas para se resolver após a aquisição das images até a recuperação do modelo:

- Estabelecer correspondências entre as vistas;
- Estimar a calibração intrínsica e extrínsica da câmera;
- Traçar as linhas de projeção para obtenção do modelo 3D.

Para se estabelecer todas as correspondências entre as imagens preciso dos parâmetros de calibração da câmera que são estimados a partir de algumas correspondências, no caso do tutorial oito correspondências iniciais para estimação da *Fundamental Matrix* e a partir daí utilização do *RANSAC* para melhora desta matriz.

A partir então da *Fundamental Matrix* é que se calcula todas as correspondências entre as imagens.

Vários algoritmos são usados para a busca automática dos primeiros pontos de correspondências. A idéia apresentada é basear-se em informações ao redor do pixel e não apenas o pixel em si, extraindo-se características que mapeiam únicamente cada um dos pixels. A busca de características usada no tutorial é baseada em detecção de cantos. O algoritmo usado é o algoritmo de *Harris*.

Esta características não é a única a ser levada em conta, pois em uma imagem existem muitos cantos. Outras técnicas são então aplicadas em conjunto. A próxima técnica é baseada em uma medida de similaridade entre os cantos, que é calculada usando-se a intensidade de cor um uma janela de tamanho pré-definido em volta dos pixels detectados como cantos. Dessa forma, algumas correspondências são encontradas. Estas correspondências são usadas como entrada para o algoritmo de RANSAC para estimação da *Fundamental Matrix*.

A busca de todas as correspondências entre as imagens pode ser então efetuada usando-se a *Fundamental Matrix* como uma guia. Traçando raios de projeção que passam pelos centros de projeção das imagens e por cada um dos pixels correspondentes nas imagens é possível encontrar o ponto de intersecção dessas duas retas. Este ponto de intersecção é o ponto 3D reconstruído que estamos procurando.

Devido ao erro envolvido no cálculo e estimação da *Fundamental Matrix*, o ponto de intersecção entre as duas retas é inexistente. Dessa forma, um erro é admitido e se as retas estiverem próximas até um certo δ, o ponto de intersecção, e logo, o ponto em 3D procurado é definido.

O processo é repetido para todos os pixels correspondentes entre as imagens. E uma estrutura 3D inicial é encontrada. A partir dessa estrutura o processo é repetido para cada vista adicional, sendo que na reconstrução, os novos pontos 3D encontrados são adicionados a estrutura.

A metodologia proposta por Pollefeys se apresenta de forma eficiente para cenas com diferentes graus de escala. Os modelos encontrados ainda são muito simples e não fidedignos aos objetos originais, mas a captura de imagens é feita sem qualquer restrição e a reconstrução é feita sem intervenção interativa alguma.

4. `Trilinearity of Three Perspective Views and its Associated Tensors [4]`

Este artigo foi publicado no ano 1995 nos proceedings da *IEEE*

Neste artigo Shashua mostra que uma permutação dos coeficientes do tensor que estabelece restrições entre três vistas produz três matrizes de homografias entre três planos de vistas, o que significa que a estrutura de restrição, as invariantes entre as três vistas pode ser recuperada diretamente apenas com um rearranjo dos coeficientes tensoriais.

O proprio Shashua já havia definido anteriormente as restrições trilineares entre três imagens através de um tensor com 27 coeficientes. As vantagens das constantes trilineares (sobre três vistas) sobre as constantes bilineares (sobre duas vistas) são que o tensor trilinear pode ser recuperado através de 7 correspondências entre as imagens quando a *Fundamental Matrix* precisa de 8 correspondências para ser recuperada. Além disso, a concatenação da geometria epipolar entre três vistas não torna sempre possível a recuperação das constantes bilineares (*Fundamental Matrix*) ao invés da constantes trilineares (*Trilinear Tensor*), que não falha nos casos em que a primeira falha. E, por último, as constantes trilineares utilizam-se das três vistas em conjunto, ao contrário das constantes bilineares que as usam em pares, ganhando com isso estabilidade numérica.

A principal idéia deste artigo é a descoberta da existência de uma permutação entre os coeficientes trilineares (*Tensor*) que produz três matrizes homográficas implícitas. Os planos correspondentes a estas homografias estão associados ao frame de coordenadas da terceira vista e provê uma base de referência para a reconstrução dos invariantes.

Outra importante derivação do artigo foi a forma para se calcular a *Fundamental Matrix*. A partir de agora é possível estabelecer uma relação entre as três vistas involvendo-se a *Fundamental Matrix*. Dessa forma o cálculo da mesma com a introdução da terceira vista será potencialmente mais estabilizada nos cálculos numéricos.

No campo das motivações práticas, o cálculo envolvido com três vistas era o de se achar 3 *Fundamental Matrix* distintas para cada par de vistas. Como estas vistas eram decompostas duas a duas, tínhamos três conjuntos diferentes, o que na prática se traduzia para duas vistas de cada vez apenas. Com o *Tensor*, apenas 7 correspondências entre as imagens são necessárias e as equações não podem ser decompostas linearmente, implicando que as três vistas são realmente usadas de uma só vez.

5. `Stratification of three-dimensional vision: projective, affine and metric representations [5]`

Este artigo foi publicado no ano de 1995 no *Journal of the Optical Society of America* por Olivier D. Faugeras.

Neste artigo Faugeras introduz conceitos matemáticos sobre a estratificação das formas de representação das relações 3D contidas no mundo vistas por um sistema de cameras. Geometria de câmera também é apresentada e discutida por ele além de formas e métodos pra recuperação das três formas de representação.

De acordo com o Faugeras, o mundo é geralmente visto como sendo Euclidiano, ou seja, um lugar onde faz sentido termos medidas como ângulos e distâncias. Olhando-se o espaço com câmeras, primeiro as medidas são feitas por imagens tiradas a partir destas câmeras e depois disso tenta-se estabelecer relações de medidas ao mundo.

É muito difícil estabelecermos relações métricas a partir de imagens, por que uma câmera não é um equipamento de medidas, a não ser que esteja completa e cuidadosamente calibrada para isso, o que é extremamente complicado. Neste sentido é importante saber como extrair invariantes da cena (imagem) observada. Um invariante é a propriedade de configuração geométrica que não varia quando uma transformação é aplicada a esta configuração.

Mais especificamente, se um invariante é calculado no stratum projetivo (que é o mais geral), então estes invariantes vão permanecer no próximo stratum, que são os afim e euclidiano.

O espaço tridimensional é dividido nas seguintes camadas de acordo com Faugeras:

- **Espaços tridimensionais como Espaços Projetivos** O uso de espaços projetivos se tornou comum no uso de aplicações em computação gráfica porque fornece uma forma compacta de representação de mudança de coordenadas usando matrizes 4x4 ao invés de matrizes de rotação e vetores de translação

- **Espaços tridimensionais como Espaço Afim** A segunda camada pode ser considerada como sendo encaixada em um espaço projetivo correspondente. Nesta camada existe uma relação de um pra um entre o espaço afim usual e o espaço projetivo menos o plano no infinito.
- **Espaços tridimensionais como Espaços Euclidianos** Especializando ainda mais o conjunto de transformações afins, temos que além de preservarmos pontos, linhas e planos no infinito existem dois pontos especias chamados pontos absolutos.

Com estas representações cada camada corresponde a uma estrutura geométrica específica que podem ser ordenadas hierarquicamente desde a mais geral (Espaços Projetivos) até a mais especializada (Espaços Euclidianos). Para cada camada estão associadas um grupo de transformações. Estas transformações podem ser incluídas umas nas outras de forma que a menos geral é um subgrupo da mais geral e cada uma delas deixa alguma invariante: o cross ratio é a mais importante para o grupo projetivo, a razão de dois vetores paralelos é o mais importante para o grupo afim e os ângulos e razões de distâncias são as mais importantes para o grupo euclidiano.

O modelo de câmera pinhole é uma excelente aproximação do mundo real para a maioria das aplicações práticas e por isso é o modelo adotado ao longo do artigo e na maioria das aplicações desenvolvidas em computação gráfica e visão computacional.

A câmera mapeia espaços projetivos tridimensionais em planos projetivos bidimensionais através de uma projeção perspectiva do centro de projeção C em cima de um plano R não que não contém C. A matroz P 3x4 define essa projeção de um ponto 3D no espaço para este plano de projeção e é chamada de matriz projetiva da câmera.

Com duas câmeras temos que a linha que passa pelos dois centros de projeção das câmeras C e C' intersecionam os planos de projeção R e R' em dois pontos chamados de epipolos e e e'. Por construção, qualquer plano contendo a linha (C, C'), chamado de plano epipolar intersecciona R e R' em uma linha ao longo do epipolo e (e'). Essas duas linhas são chamadas de linhas epipolares.

A propriedade fundamental desta construção geométrica é que a correspondência natural entre os dois lápis de linhas epipolares é projetivamente linear, isto é, é uma homografia entre as linhas epipolares. A correspondência natural é associar a cada linha epipolar do primeiro plano projetivo a uma linha no segundo plano projetivo.

Esta correspondência epipolar tem grande importância e já foi apresentada na literatura relacionada por muitas vezes. A matriz que define essa correspondência é chamada de *Fundamental Matrix* e expressa algebraicamente a correspondência existente entre dois planos projetivos através de uma matriz $3x3$. Isto significa mais especificamente que dado um ponto em um plano projetivo, a *Fundamental Matrix* nos devolve a linha no outro plano projetivo onde a imagem deste ponto deve estar.

6. `Structure from Multiple 2D Affine Correspondences without Camera Calibration [6]`

Este artigo foi publicado nos proceedings do *IEEE* no ano de 1996 por H. Schweitzer.

Neste artigo o autor apresenta como extrair informações 3D através de parâmetros afins sem calibração de câmera com as assumpções de o objeto ser rígido, ser localmente planar e seu movimento 3D inclui apenas translação. Com essas assumpções, a estrutura 3D pode ser reconstruída totalmente mas a movimentação não é possível ser estimada.

A estrutura 3D pode ser extraída diretamente através dos parâmetros afins, mas a recuperação da movimentação de câmera não é possível ser estimada sem calibração. Os resultados apresentados são baseados em assumpções muito mais restritivas do que outros trabalhos na mesma área, mas mesmo assim o autor argumenta que estas restrições fazem sentido localmente para o algoritmo.

Com essas restrições mais fortes, o método proposto consegue extrair muito mais informações sobre a estrutura 3D mesmo sem calibração mesmo que com isso se restrinje a atuar apenas em pequenas regiões.

Capítulo 5

Conclusão

A área de reconhecimento de imagens para reconstrução automática de modelos 3D tem uma gama de aplicações práticas na sociedade atualmente. Incluem-se entre estas as áreas de visão robótica para controle automático de equipamentos, piloto automático para veículos e detecção de colisão; reconstrução de modelos representativos de objetos reais para aplicações em passeios virtuais e modelagem precisa de ambientes reais.

A caracterização e construção de tais estruturas não é uma tarefa trivial e tem sido objeto de estudo de vários pesquisadores ao longo dos anos. Apesar de ser uma área recente dentro da computação, avanços significativos foram atingidos. Exemplos de resultados obtidos podem ser observados através programas como o de Marc Pollefeys [3] e de Paul Debevec [8].

Problemas intrinsicamente ligados a reconstrução atualmente são a calibração de câmera e problemas matemáticos de ordem numérica. Atualmente o conceito de sistema automático têm sido procurado pelos pesquisadores da área, onde a calibração é feita de forma automática, sem que seja necessário ajustes finos e construção de padrões, facilitando o processo de captura e automatizando o sistema. O aprimoramento do processo de calibração implica diretamente na qualidade da estimação numérica, tornando assim o sistema mais robusto e menos suscetível a erro.

A partir das soluções e resultados obtidos ao longo dos últimos anos, pretendemos implementar alguns desses algoritmos e testar seus potenciais efetivos. Variações de alguns algoritmos para o caso de detecção de características como cantos correspondentes em imagens serão estudadas. Entre essas variações o uso da cor como fator de detecção de cantos deverá ser proposto. Formas coerentes para representação das estruturas e formas reconstruídas também deverão ser estudadas.