

NAVEGAÇÃO 3D A PARTIR DE FOTOS DE EDIFICAÇÕES E MODELOS CAD

Pablo C. Elias

Asla M. Sá

Alberto Raposo

Marcelo Gattass

Tecgraf - Grupo de Tecnologia em Computação Gráfica, PUC-Rio

R. Marquês de S. Vicente, 225, Gávea, 22453-900, Rio de Janeiro - RJ - Brasil

{*pelias, asla, abraposo, mgattass*}@tecgraf.puc-rio.br

Paulo Cezar Carvalho

IMPA - Instituto de Matemática Pura e Aplicada

Estrada Dona Castorina, 110, 22460-320, Rio de Janeiro - RJ - Brasil

pcezar@impa.br

Resumo. *Antes de serem construídas, a maioria das edificações atuais tem seu modelo geométrico totalmente definido por modelos CAD. Uma vez construída a edificação, tais modelos podem ser explorados de diferentes formas por fornecerem informações métricas precisas que podem ser associadas a dados reais úteis para a manutenção e inspeção das respectivas edificações. Um exemplo prático é a utilização de fotografias de uma edificação que podem ser associadas ao modelo virtual para fins de navegação e comparação. Nesse contexto, propomos um sistema de Realidade Aumentada que permite associar fotografias diretamente ao modelo CAD utilizando técnicas de Visão Computacional. Através da sobreposição semi-automática da foto no modelo recuperamos a posição e orientação da câmera real, o que possibilita a comparação entre dados da foto e do modelo. Uma vez que as posições das cameras foram recuperadas, o sistema permite a navegação tridimensional entre as fotos utilizando o modelo como referência. Esta funcionalidade permite a catalogação 3D das fotos da edificação.*

Keywords: *Calibração de câmera, Realidade Aumentada, Navegação 3D por fotos*

1. INTRODUÇÃO

A reconstrução de câmera é um dos problemas fundamentais da visão computacional. Sistemas de software desta área utilizam modelos de câmeras virtuais para interpretar, reconstruir e visualizar a estrutura tridimensional de uma cena real a partir de fotografias e vídeos digitais no intuito de produzir imagens sintéticas com aparência realista. As técnicas de reconstrução de câmera da Visão Computacional podem ser aplicadas em conjunto com técnicas de Realidade Virtual dando origem a aplicações na área de Realidade Aumentada. Dentre os diversos usos destes tipos de aplicação, neste trabalho estamos interessados em proporcionar *visitas aumentadas* a edificações, isto é, proporcionar passeios em modelos virtuais aumentados por informações de fotografias reais da edificação.

Outra funcionalidade de grande interesse diz respeito a levantamentos fotográficos úteis para a realização de inventários e manutenção de uma dada edificação. Em geral é comum o fato de os nomes dos arquivos de fotos do levantamento sugerirem pouca ou nenhuma informação a respeito de seu conteúdo específico, o que dificulta a busca e o armazenamento de fotos por conteúdo. No caso de edificações, para procuramos uma foto tirada a partir de um ponto de vista pre-estabelecido se faz necessária uma busca linear em todos os arquivos para encontrá-la.

Com a possibilidade de registro 3D, muitas funcionalidades de busca por uma foto a partir de uma região da cena virtual podem ser desenvolvidas para facilitar o gerenciamento de acervos fotográficos. Por exemplo, no caso de construções inacabadas ou em ruínas, há a possibilidade de utilizar um modelo tridimensional da construção completada de acordo com o projeto ou registros históricos (no caso de ruínas), criando uma aplicação de Realidade Aumentada, onde as fotos reais contrastam com o modelo virtual completado.

Do ponto de vista técnico, a reconstrução de câmera é um dos problemas fundamentais a serem resolvidos para desenvolver um sistema de navegação 3D em acervos fotográficos. A reconstrução de câmera consiste em recuperar parâmetros que compõem um modelo virtual de câmera a partir da imagem de uma cena real. Neste trabalho é abordado o problema de recuperar o posicionamento de câmeras a partir de uma única foto, sem uso de marcadores aplicados ao objeto real, e seu posicionamento aproximado relativo ao modelo computacional da edificação retratada.

O método proposto utiliza técnicas integradas que auxiliam o usuário a criar um conjunto de correspondências entre imagens digitais da edificação e seu modelo virtual. A partir dessas associações, a posição da câmera real que originou a fotografia relativa ao modelo é recuperada possibilitando a navegação 3D do conjunto de imagens fornecidas.

Apesar de propormos um método semi-automático temos como premissa demandar pouca interação do usuário, minimizando seu esforço. O conceito de interface proposto permite que o usuário manipule apenas a imagem para que as associações com pontos do mundo sejam detectadas, dispensando associações diretas entre segmentos da imagem e do modelo. Contudo, a reconstrução de câmera utilizada pode funcionar também com um conjunto de associações imagem-modelo fornecidas explicitamente pelo usuário.

A aplicação de exemplo implementada provê solução para a recuperação e o registro de câmeras a partir do casamento semi-automático entre modelos CAD e fotos. A aplicação disponibiliza ferramentas para compará-los e provê uma nova experiência de navegação tridimensional entre as diversas fotos registradas.

Os principais requisitos funcionais da aplicação proposta são: reconstrução e posicionamento de câmeras virtuais a partir de um conjunto esparsa de fotos, modelos computacionais e marcações auxiliares feitas pelo usuário; capacidade de registrar as diversas câmeras reconstruídas; navegação tridimensional entre as câmeras registradas. A técnica proposta pressupõe que um modelo geométrico da construção de interesse seja conhecido e propõe técnicas origi-

nais para o posicionamento de fotografias em relação ao modelo.

O caso de estudo abordado envolve um conjunto de fotos e um modelo do convento de São Boaventurado do Século XVII, localizado em Itaboraí, RJ, atualmente em ruínas. Contudo, como será apresentado, não há restrições em relação ao tipo de modelo tratado, de modo que é possível utilizar também modelos de objetos cuja estrutura é conhecida com mais precisão, como modelos de engenharia ou modelos CAD de edificações modernas.

O artigo está organizado da seguinte forma: a próxima seção cobre as principais referências utilizadas na conceituação deste trabalho e enfatiza as diferenças da abordagem proposta em relação aos trabalhos propostos na literatura. Na Seção 3 as etapas de pré-processamento dos dados é apresentada. Na Seção 4 a técnica de recuperação da posição de câmera é apresentada. Na Seção 5 descrevemos a aplicação desenvolvida e o sistema de navegação 3D de um conjunto de fotos. Alguns resultados são apresentados e discutidos na Seção 6 e finalmente conclusões e trabalhos futuros são discutidos na última seção.

2. TRABALHOS RELACIONADOS

Dois sistemas fotogramétricos serviram de inspiração para este trabalho: o Photosynth da Microsoft¹, que teve origem no trabalho de Noah Snavely Snavely et al. (2006), e o Façade, proposto por Paul E. Debevec Debevec (1996).

O Photosynth parte de um conjunto denso de fotos de um objeto, com regiões de sobreposição entre elas, para recuperar nuvens de pontos tridimensionais e modelos de câmeras dessas fotos a partir apenas das correspondências entre elas, calculadas utilizando os algoritmos RANSAC Fischer and Bolles (1981) e SIFT Lowe (2004). A partir das correspondências encontradas automaticamente, a geometria epipolar é utilizada para computar a matriz fundamental e recuperar pontos do mundo, que é refinada de acordo com o número de correspondências encontradas. Embora seja uma técnica bem-sucedida e cujo resultado é bastante eficaz, principalmente em relação à recuperação de câmeras, ela possui uma condição de funcionamento que não é desejada neste trabalho: pressupõe a disponibilidade de um conjunto denso de fotos que tenham regiões de sobreposição entre si e demanda longo tempo de processamento.

Já o Façade adota uma abordagem semi-automática em que um conjunto esparsa de fotos é utilizado e a geometria do modelo é parcialmente conhecida. O modelo de câmera é reconstruído com o objetivo de recuperar proporções de um modelo paramétrico previamente criado pelo usuário que o associa a diversas marcações em diversas imagens. Debevec baseia-se principalmente em Taylor and Kriegman (1995) para reconstruir a estrutura tridimensional de uma cena a partir de múltiplas fotos. Ele propõe também um novo método para estimar uma rotação inicial de câmera usando direções conhecidas na cena. A técnica de ajuste de câmera e de casamento entre modelos e imagens que propomos inspirou-se nestes trabalhos.

Na fase inicial deste trabalho, a técnica proposta por Debevec foi implementada e concluiu-se que ela demandava intensa interação do usuário, o que diverge dos objetivos aqui traçados. Além disso, o resultado final desta técnica é tão bom quanto a qualidade do modelo paramétrico criado e das marcações feitas nas imagens. No entanto, Debevec utiliza tal técnica com o objetivo principal de reconstruir proporções de um modelo rústico que é usado posteriormente como entrada para outras técnicas integradas, tal como View-Dependent Texture Mapping Debevec (1996), geração de mapas de profundidade usando o modelo como restrição e outras técnicas que aumentam a qualidade da renderização final, mas fogem ao objetivo aqui proposto. O sistema Façade criado por Debevec, deu origem a outros sistemas comerciais e é utilizado com sucesso para recuperar proporções de modelos e posição de câmeras com eficácia.

¹<http://labs.live.com>

O fato de assumirmos que a geometria do modelo é conhecida caracteriza uma diferença substancial em relação ao objetivo do Photosynth e do Façade e permite uma simplificação do processo de reconstrução das câmeras. No entanto, apesar das diferentes abordagens, certos conceitos da aplicação final de Snavely foram utilizados aqui para dar origem à aplicação de exemplo, como o sistema de calibração utilizando EXIF tags e o conceito geral de álbum de fotos tridimensional.

3. PRÉ-PROCESSAMENTO

Durante a fase de pré-processamento, duas tarefas são realizadas independentemente e concluídas antes da fase principal da aplicação, são elas: a recuperação dos parâmetros intrínsecos da câmera que originou a fotografia real da edificação e o carregamento e processamento da geometria do modelo CAD da edificação.

3.1 Recuperação dos parâmetros intrínsecos da câmera

Uma câmera calibrada é aquela na qual se pode relacionar um ponto x qualquer em seu plano de projeção ao respectivo raio d que liga seu centro óptico a x . Formalmente, isso equivale a encontrar uma matriz K de calibração que representa uma transformação entre esse ponto x e a direção do raio $d = K^{-1}x$ Hartley and Zisserman (2003)

Neste trabalho adotamos o modelo pinhole CCD de câmera, que é um modelo similar ao modelo pinhole clássico, mas que leva em consideração que os pixels do plano de projeção podem não ser quadrados. Neste caso a matriz K é dada por:

$$K = \begin{bmatrix} \alpha_x & 0 & p_x \\ 0 & \alpha_y & p_y \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

Nos referiremos à reconstrução de parâmetros intrínsecos da câmera como a etapa de calibração de câmera. Tais parâmetros são calculados para cada imagem fornecida de entrada. Se essa imagem contiver informações EXIF apropriadas, a geometria da câmera pode ser diretamente extraída dos meta-dados do arquivo da imagem. Caso tais informações não estejam disponíveis, a calibração é feita a partir dos pontos de fuga da imagem.

Em Rother (1999), Schaffalitzky and Zisserman (2000), McLean and Kotturi (1995) e Gamba et al. (1996) são desenvolvidas técnicas para detecção automática de pontos de fuga, mas funcionam apenas em casos em que há pouco ruído e na presença de bons segmentos de reta detectáveis na imagem. Além disso, a probabilidade de falso positivo é grande, ou seja, freqüentemente segmentos que não pertencem a nenhuma das três direções são detectados uma vez que retas perpendiculares do mundo podem ser projetados quase de forma paralela na imagem.

Optamos por uma abordagem semi-automática onde as direções principais da cena são fornecidas pelo usuário, como ilustrado na Figura 1. A foto foi tirada na década de 1980, de modo que não existem as informações EXIF e há bastante ruído devido ao processo de digitalização, constituindo um caso bastante difícil para a abordagem automática.

O método implementado para a calibração da câmera é o método sugerido em Hartley and Zisserman (2003). Embora a calibração por pontos de fuga forneça uma boa estimativa da orientação da câmera, apenas seus parâmetros intrínsecos são calculados nesta etapa, a orientação e a translação da câmera serão obtidos em outra etapa do processo.



Figura 1: Exemplo de calibração por pontos de fuga. As linhas selecionadas pelo usuário são indicadas por setas na figura.

3.2 Extração das Arestas Estruturais

Frequentemente, malhas geométricas geradas em aplicações de modelagem exibem muitas arestas que não têm chance de corresponder a segmentos da imagem de entrada, seja porque não são visíveis ou por serem decorrentes de processos de representação da malha visando rendering ao invés de priorizarem uma descrição estrutural sucinta do modelo.

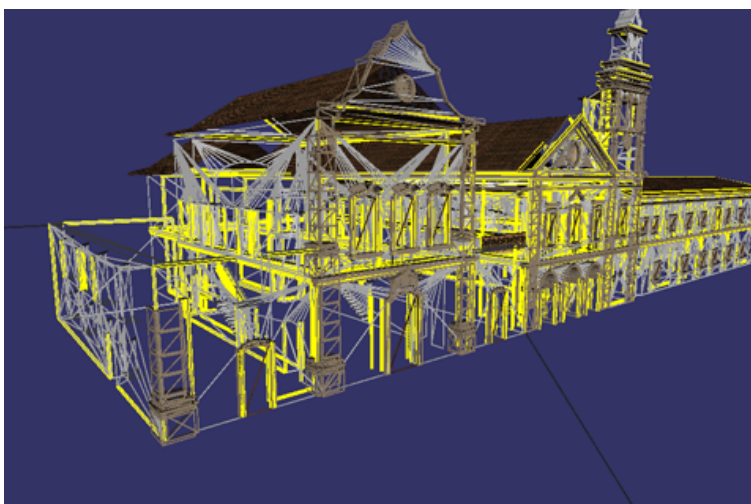


Figura 2: Parte da malha original de um modelo exibida em wireframe.

A etapa de pré processamento do modelo virtual é constituída de uma sequência de operações simples que resulta em uma nova lista de arestas chamadas arestas estruturais do modelo utilizadas para facilitar a etapa de localização de correspondências entre modelo e imagem. As operações de pré processamento do modelo original são as seguintes:

- Descarte de arestas duplicadas;
- Descarte de arestas coplanares;

- Agregação de sequências de arestas colineares;
- Normalização da malha e subsequente descarte de arestas curtas;
- Possibilidade de descarte manual de arestas.



Figura 3: Algumas das arestas estruturais visíveis selecionadas marcadas em amarelo por cima do modelo.

Os testes do processo de simplificação em um modelo com mais de 1.000.000 de arestas resultou em uma lista com pouco mais de 3.000 arestas (0.3 %), das quais a maioria será descartada durante o teste de oclusão mostrado adiante. De fato, no modelo ilustrado nas Figuras 2 e 3, pouco mais de 90 arestas visíveis foram selecionadas, em média, para a seleção das relações entre modelo e imagem.

Quando a malha de um modelo é carregada pela primeira vez, ela é processada e a lista de arestas estruturais é armazenada em um arquivo auxiliar (.simp). A geometria original do modelo torna a ser utilizada para decoração da cena e para efetuar a etapa do teste de oclusão, descrita adiante.

4. RECUPERAÇÃO DA POSIÇÃO DA CÂMERA

Para a recuperação do ponto de vista de uma determinada câmera é fundamental a identificação de características correspondentes entre fotografia e modelo virtual. O modelo virtual aproximado da cena retratada nas imagens faz o papel de padrão de calibração para as câmeras. Isso exige que se conheça suas proporções reais e que as fotografias fornecidas sejam processadas para que características correspondentes entre elas e o modelo virtual possam ser identificadas.

Como o caso de estudo se aplica a edificações, a identificação de arestas foi preferida ao invés de pontos por se mostrarem melhor rastreáveis em fotografias de edificações. A abordagem aqui adotada estabelece que o usuário tem um papel importante no gerenciamento das associações foto-modelo, principalmente em casos onde, por fatores inerentes à qualidade da imagem de entrada, não é possível detectar uma quantidade suficiente de arestas de boa qualidade.

O método proposto visa auxiliar ao máximo as ações do usuário de forma que, em casos simples em que há pouco ruído nas imagens, ele possivelmente não precise fornecer marcações

extras para que a aplicação recupere a câmera. Nesses casos, o usuário precisará apenas fornecer um posicionamento inicial de câmera apropriado.

4.1 Solução Inicial

É pressuposto que a posição inicial do modelo virtual que corresponde aproximadamente ao ponto de vista da fotografia cuja camera deve ser recuperada seja fornecida pelo usuário. Para posicionar o modelo, o usuário fornece manualmente, através de um manipulador tipo *trackball*, uma translação e uma rotação que são aplicados à câmera virtual. O objetivo é que as arestas estruturais do modelo alinhem-se aproximadamente com as arestas correspondentes da fotografia que está mapeada no plano de projeção da câmera. Do ponto de vista do observador, a imagem está sempre parada no espaço da tela e o modelo se move.

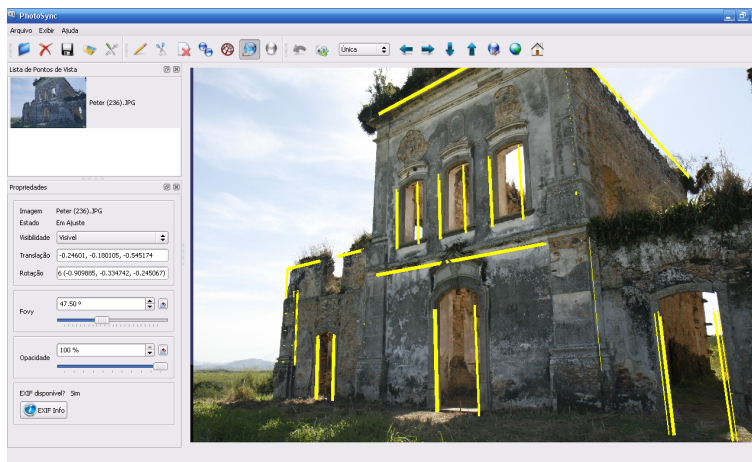


Figura 4: Modelo sobreposto por uma imagem com 100% de opacidade. Mesmo com opacidade máxima as linhas-guia ainda são visíveis e o modelo pode ser manipulado.

Utilizando a posição inicial fornecida pelo usuário, as arestas estruturais do modelo virtual passam por um teste de oclusão que leva em consideração o modelo original e cuja finalidade é descartar as arestas estruturais não visíveis a partir do ponto de vista da posição inicial.

Para implementar o teste de oclusão utiliza-se a funcionalidade de Occlusion Query do OpenGL. O procedimento é simples: o modelo original é renderizado; em seguida cada uma das arestas estruturais é renderizada, seguida da verificação da Occlusion Query. Se a aresta renderizada gerou pixels visíveis, ela é visível e deve ser preservada. Caso contrário a aresta é descartada. O resultado final consiste de um conjunto de arestas estruturais visíveis V .

4.2 Correspondência entre Arestas Estruturais e da Imagem

O método semi-automático proposto para resolver o problema de associação de características entre a fotografia e o modelo virtual consiste de 3 etapas: 1) detecção de arestas na fotografia; 2) interpretação da correspondência entre as arestas detectadas e as arestas estruturais visíveis do modelo virtual; 3) intervenção e validação das correspondências pelo usuário.

Detecção de Arestas na Imagem Para localizar segmentos de reta as fotografias são submetidas à aplicação do filtro Canny para o realce das arestas Canny (1986) e em seguida algoritmos específicos para detecção de arestas são aplicados. Neste trabalho avaliamos a utilização de dois algoritmos: a Transformada de Hough Padrão (SHT - Standard Hough Transform) Trucco and Verri (1998) e a Transformada de Hough Probabilística (PPHT) Matas et al. (2004). A

biblioteca do OpenCV² foi utilizada para a implementação de ambos os métodos.

Ambos os métodos foram testados e observou-se que cada um apresenta capacidades diferentes dependendo da qualidade da imagem de entrada e, por isso, não é possível afirmar definitivamente qual o melhor. O PPHT é mais eficiente em imagens que possuem bastante ruído, mas é também menos preciso. A escolha por algum dos dois métodos depende diretamente da natureza da imagem de entrada. O número de segmentos de qualidade detectados pelo PPHT foi maior do que o teste feito com o HT e o tempo de execução foi consideravelmente menor em imagens de alta resolução.

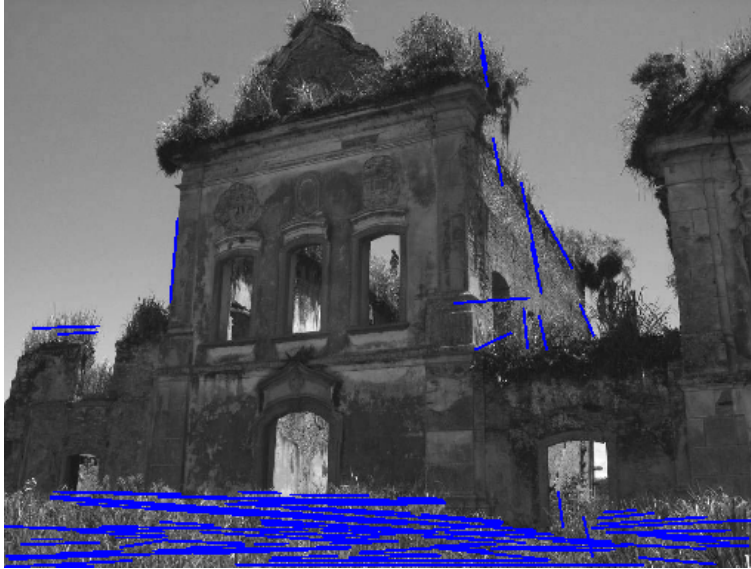


Figura 5: Caso difícil. Mais de 400 arestas detectadas em sua maioria na região da vegetação.

Como resultado desta etapa obtemos um conjunto $U = \{u_1, \dots, u_p\}$ de arestas da imagem.

Correspondência entre Arestas As arestas estruturais do conjunto $V = \{v_1, \dots, v_n\}$ (que passam no teste de oclusão) são então projetadas no espaço da imagem, utilizando os parâmetros intrínsecos de câmera obtidos na etapa de pré-processamento, criando assim um conjunto de segmentos de reta $V' = \{v'_1, \dots, v'_n\}$ dados pela projeção das arestas do conjunto V no espaço da imagem.

Utilizaremos o modelo para restringir a área de busca por segmentos correspondentes na imagem criando uma janela em volta de cada segmento v'_s , delimitando assim uma região de interesse $ROI(v'_i)$ relativa a cada aresta. Cada ROI funciona como uma sub-imagem que restringe a área de busca por arestas do conjunto U candidatas a serem correspondentes à aresta estrutural v'_i .

Para eleger uma aresta em U que seja correspondente a uma aresta v'_i é razoável analisar apenas os elementos u_i contidos na $ROI(v'_i)$, dando preferência aos elementos cuja inclinação, tamanho e extremidades sejam comparáveis às de v'_i , uma vez que partimos do princípio que a posição inicial do modelo fornecida pelo usuário está aproximadamente ajustada com a foto.

Para classificar objetivamente as arestas de U candidatas a correspondentes, isto é, que estão dentro da região efetiva de um ROI, adaptamos a função de erro proposta em Taylor and Kriegman (1995). Assim cada aresta de U que está na $ROI(v'_i)$ recebe uma classificação de acordo com a seguinte função:

²<http://www.intel.com/technology/computing/opencv>

$$rank(u_i) = \frac{l}{h_1^2 + h_1 h_2 + h_2^2} \quad (2)$$

onde l é o tamanho do segmento u_i e h_1 e h_2 são as distâncias dos pontos extremos do segmento u_i à reta L_i , construída a partir dos pontos extremos de um segmento projetado v_i .

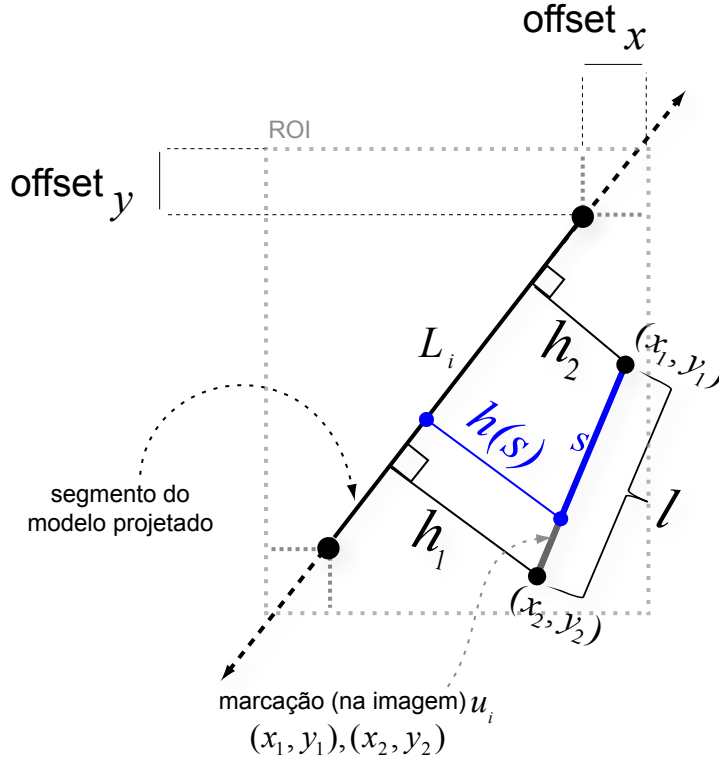


Figura 6: Classificação de uma aresta em um ROI.

A aresta u_i melhor classificada segundo a função $rank$ é escolhida como sendo a correspondente na imagem à aresta estrutural v_i do modelo virtual.

Observa-se que a classificação dada pela função $rank$ prioriza arestas maiores sobre arestas menores, mas arestas menores que estão mais próximas e mais alinhadas têm prioridade sobre arestas maiores e distantes ou menos alinhadas. A qualidade da classificação obtida é dependente da posição do modelo virtual fornecida pelo usuário.

Por se tratar de um método guloso de busca local, se houver intersecções entre ROIs pode ocorrer da mesma aresta u_i da imagem ser classificada como correspondente a duas ou mais arestas estruturais distintas do modelo, o que é uma inconsistência. Para contornar este problema o critério de desempate adotado para eleger a aresta estrutural que melhor corresponde à aresta u_i da imagem consiste de comparar o valor do $rank(u_i)$ em relação às diversas arestas v_i classificadas como correspondentes ao mesmo u_i e novamente adotar uma estratégia gulosa.

Validação e Edição Manual de Correspondências As correspondências entre modelo e fotografia obtidas automaticamente são finalmente exibidas para o usuário através de marcações visuais, para que ele possa então descartá-las, aprová-las ou complementá-las. A complementação realizada pelo usuário é composta de novas marcações feitas apenas na imagem. A partir

destas marcações e da posição inicial de câmera fornecida pelo usuário, classifica-se a melhor candidata a arestas correspondentes do modelo. Somente em casos onde não há possibilidade de se detectar correspondências por proximidade (devido à baixa qualidade da imagem) é que se faz necessária a associação manual direta entre modelo e imagem.

A complementação é necessária em casos onde não é possível detectar arestas nas três direções principais da cena, o que caracteriza um requisito fundamental para a etapa seguinte de ajuste de câmera (minimização). Tal fato pode ocorrer principalmente devido à ruído ou por oclusão. Assim, o usuário deve complementar o conjunto de arestas detectadas marcando novas arestas em regiões em que nenhum ou poucos segmentos foram encontrados.

Em resumo, a saída da etapa de correspondência de arestas é um conjunto de pares $\{(v_1, u_1), (v_2, u_2), \dots, (v_t, u_t)\}$ a ser usado como entrada da etapa de minimização, descrita a seguir.

Revisado até aqui

4.3 Otimização da posição inicial da Câmera

Utilizando o conjunto de associações entre modelo e imagem já validado pelo usuário e o posicionamento inicial de câmera, o posicionamento da câmera é ajustado de modo a maximizar o alinhamento dos segmentos correspondentes da imagem e do modelo.

Para tanto, é necessário mensurar o erro entre a projeção de arestas do modelo da imagem. A partir de tais medidas pode-se recuperar a posição e orientação de uma câmera calibrada minimizando uma função-objetivo que depende dos parâmetros externos da câmera e das correspondências obtidas.

A função-objetivo que é minimizada para a obtenção da posição e da orientação finais da câmera é não linear. Para tanto, é utilizada a função de erro proposta por Taylor and Kriegman (1995) e já apresentada na Seção anterior. Essa função mede o erro total entre as arestas do modelo e da imagem em função dos parâmetros externos da câmera, que são recuperados a partir de um processo de minimização para encontrar os valores ótimos dos parâmetros.

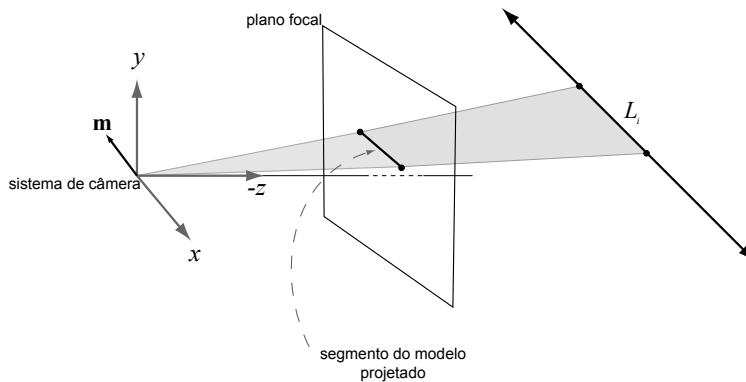


Figura 7: Representação da linha L_i pelo vetor normal \triangleright

5. NAVEGAÇÃO 3D

Uma aplicação de exemplo foi criada para demonstrar o método proposto que possibilita a recuperação semi-automática do posicionamento de câmeras de um número arbitrário de fo-

tografias de uma edificação em relação ao seu modelo virtual. Cada uma das câmeras recuperadas pode ser registrada permitindo uma catalogação e gerenciamento das fotografias no espaço 3D.

Além disso o sistema possibilita uma nova experiência de navegação tridimensional entre as fotografias permitindo a comparação de detalhes entre foto e modelo através de diversas ferramentas como a capacidade de zoom e controle de transparência.

Outra funcionalidade de grande utilidade no contexto de produção de inventários sobre o bem cultural é a possibilidade de averiguar se o conjunto de fotografias fornecido oferece uma cobertura visual completa da edificação em questão. Para implementar tal funcionalidade basta projetarmos as fotografias registradas no modelo a partir de uma transformação projetiva inversa, usando a matriz K^{-1} , e observarmos os espaços não cobertos do modelo.

A navegação tridimensional recebe como entrada uma direção específica e o conjunto de câmeras virtuais recuperadas e produz uma saída vazia ou uma determinada câmera do conjunto fornecido. Dado uma posição atual de câmera, pode-se navegar até uma outra mais próxima em uma determinada direção escolhida utilizando-se o critério abaixo (Figura 8):

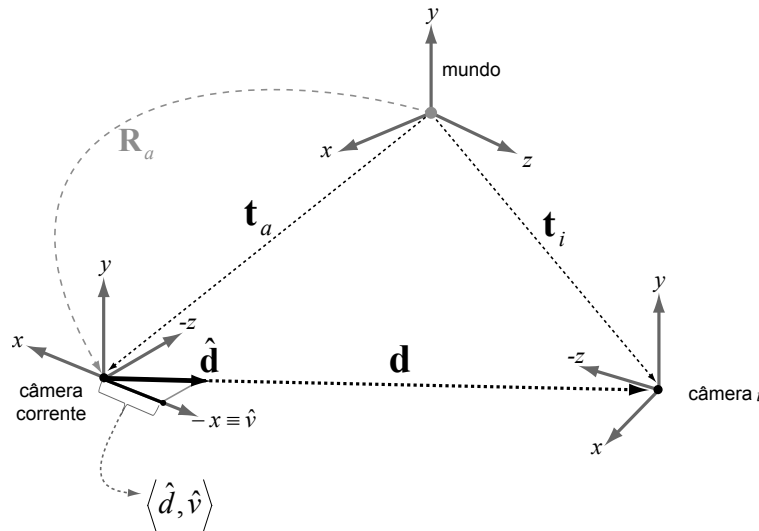


Figura 8: Uma estratégia simples para navegação entre as câmeras recuperadas.

6. RESULTADOS

O conjunto de fotos utilizado como caso de estudo foi tirado da ruína do Convento de São Boaventura - Itaboraí (RJ) - Brazil, construído por volta de 1660. O modelo do convento foi criado com base em documentos históricos e em registros fotográficos. O modelo não é totalmente preciso em relação a estrutura da ruína mas é suficiente para ilustrar e testar a aplicação e o método proposto.

As fotos utilizadas, contém muita vegetação e ruído, o que configura um caso de estudo bastante complexo. Ainda assim um conjunto 10 de câmeras foram recuperadas com sucesso a partir das respectivas fotografias em menos de 10 minutos.

7. Conclusões e Trabalhos Futuros

Este trabalho propôs um método para casamento entre modelos e fotos de edificações através de um conjunto de técnicas integradas que têm como estratégia principal a reconstrução

de câmera. Para fazer tal reconstrução e ser capaz de casar e catalogar as fotografias, foi preciso resolver o problema da correspondências entre elementos da imagem e do modelo, que constitui um dos problemas fundamentais da visão computacional. A abordagem proposta para resolvê-lo foi utilizar o próprio modelo da edificação em posicionado apropriadamente para restringir a busca por correspondências na imagem. Essa estratégia pressupõe que o modelo virtual seja manipulado pelo usuário para que suas arestas sirvam de guia para localizar correspondências na imagem, utilizando uma estratégia local de busca na vizinhança da projeção das arestas do modelo.

Pelo fato de ser semi-automático, o método parte de uma solução inicial fornecida pelo usuário que possibilita uma busca local por relações entre imagem e modelo em detrimento da exploração das informações globais do modelo. Em casos onde a imagem de entrada contém muito ruído e o modelo fotografado possui geometria complexa, torna-se naturalmente difícil resolver o problema da correspondência e o método aqui proposto se mostra mais dependente das ações do usuário e sujeito a imprecisão. Já para casos simples, o processo torna-se mais automatizado e robusto com relação ao posicionamento inicial do modelo pois torna-se mais simples calcular correlações entre imagem e o modelo.

Como resultado final foi desenvolvida uma aplicação que implementa o método proposto e oferece uma solução completa para o problema do registro de câmeras de fotografias com relação a seu modelo virtual. O sistema provê também diversos mecanismos para auxiliar o usuário a comparar as fotografias e o modelo bem como navegar espacialmente entre diversas imagens registradas.

Várias melhorias podem ser discutidas, em particular para o problema da correspondência entre pontos da imagem e do mundo. Neste contexto pode-se explorar o uso de técnicas de NPR para a detecção de arestas estruturais no modelo virtual e nas imagens.

Uma extensão natural seria a aplicação do método em vídeos. A técnica poderia ser estendida para explorar coerência espacial e temporal em seqüências de frames obtidas de vídeos reais. A facilidade da extensão do método para sequencias de vídeos se deve a própria natureza da técnica que busca localmente correspondências entre modelo e imagem. Quando se assume que a seqüência de quadros de um vídeo têm coerência espacial e temporal, pode-se aplicar diretamente o método proposto para ajustar ao modelo, partindo da posição de um frame inicial para o próximo frame, de forma automática. O processo se iniciaria com o posicionamento manual do usuário para o primeiro frame do vídeo, resultando na recuperação de um caminho completo de câmera ao longo de um intervalo de tempo.

Como trabalho futuro pretendemos utilizar um conjunto de fotos bem distribuídas no espaço como entrada para um método que as correlacione com o modelo virtual para melhorar a qualidade de um modelo virtual aproximado da edificação retratada. As arestas estruturais do modelo poderiam ser parametrizadas para que fossem ajustadas às marcações feitas nas fotos. Essa idéia é bem próxima à do trabalho desenvolvido por Paul Debevec em Debevec (1996), mas utilizaria modelos diretamente criados por modeladores comerciais o que traria a conveniência de uso de um bom modelador mas sem exigir precisão na modelagem. Em outras palavras, um modelo poderia ser rapidamente criado por um designer, sem medidas corretas, e suas proporções em seguida poderiam ser ajustadas com base nas fotos de entrada.

Outra possibilidade de ampliação deste trabalho seria o de utilizar um conjunto denso de fotos para reconstruir totalmente as proporções corretas do modelo utilizando apenas correspondências entre as imagens. Essa estratégia é inspirada no trabalho de Snavely et al. (2006), mas utiliza como base um modelo impreciso que seria ajustado e serviria como restrição para pontos reconstruídos a partir de correlações nas fotos utilizando geometria epipolar. O método proposto por Noah Snavely et al., utiliza apenas as fotos e não tem qualquer informação sobre

o modelo.

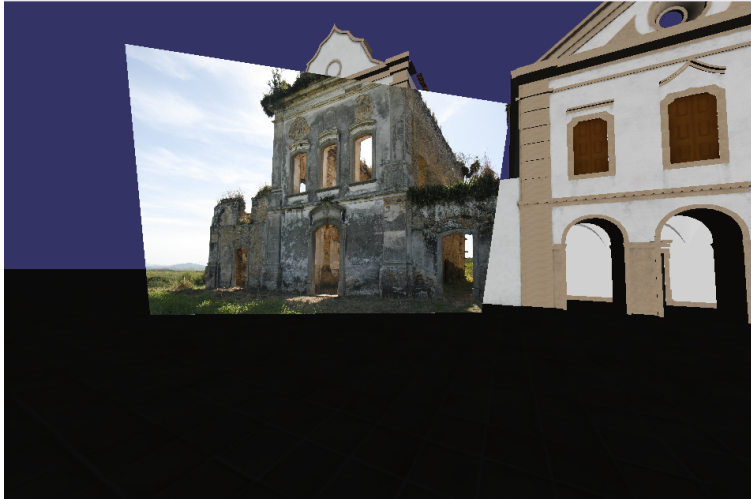


Figura 9: Casamento entre o modelo e uma fotografia do convento.

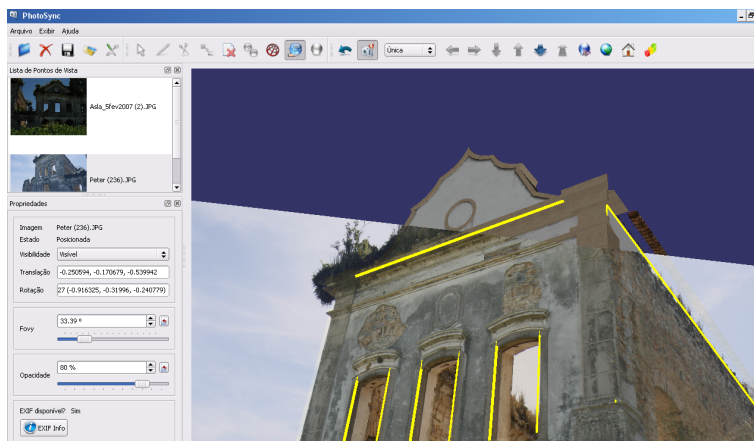


Figura 10: A mesma fotografia mostrada na Figura 9, utilizando *zoom* uma orientação diferente.

AGRADECIMENTOS

O TecGraf é um laboratório prioritariamente financiado pela Petrobras. Alberto Raposo recebe apoio do CNPq, Edital Universal 2007, Processo número 472967/2007-0.

Referências

Canny, J., 1986. A computational approach to edge detection. pp. 679–698.

Debevec, P. E., 1996. *Modeling and Rendering Architecture from Photographs*. PhD thesis, University of California at Berkeley.

Fischer, M. A. & Bolles, R. C., 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, vol. 6, pp. 381–395.



(a) O modelo virtual e suas arestas mostradas em amarelo.



(b) O casamento utilizando exatamente o mesmo ponto de vista mostrado na figura (a).

Figura 11: Casamento entre imagem e modelo do convento de Boaventura.

Gamba, P., Mecocci, A., & Salvatore, U., 1996. Vanishing point detection by a voting scheme. *IEEE International Conference on Image Processing*, pp. 301–304.

Hartley, R. & Zisserman, A., 2003. *Multiple View Geometry in Computer Vision*. Cambridge University, 2nd edition.

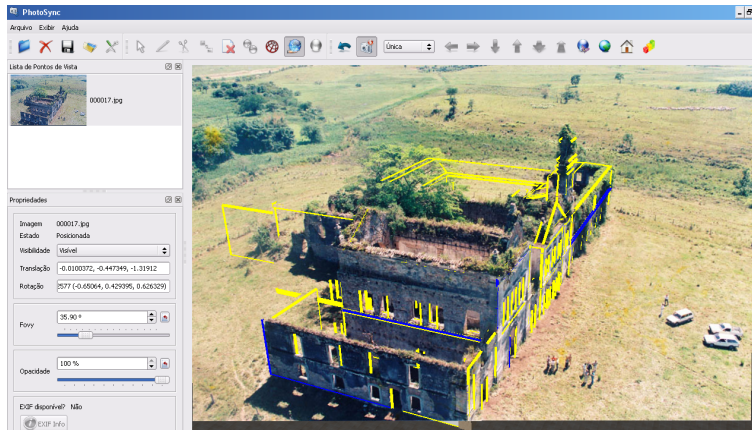
Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*.

Matas, J., Galambos, C., & Kittler, J., 2004. Progressive probabilistic hough transform. *British Machine Vision Conference*, pp. 256–265.

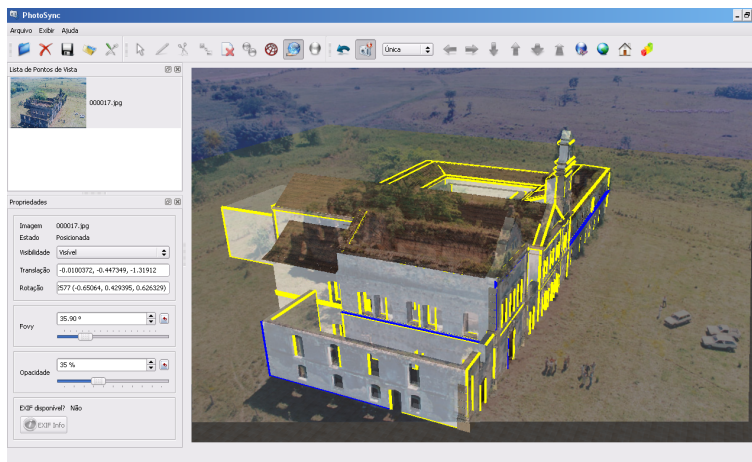
McLean, G. F. & Kotturi, D., 1995. Vanishing point detection by line clustering. *IEEE Transactions On Pattern Analysis And Machine Intelligence*, vol. 11, pp. 1090–1095.

Rother, C., 1999. A new approach for vanishing point detection in architectural environments. *Image and Vision Computing*, vol. 20(9-10), pp. 647–655.

Schaffalitzky, F. & Zisserman, A., 2000. Planar grouping for automatic detection of vanishing lines and points. Technical report, Department of Engineering Science, University of Oxford.



(a) Casamento entre o modelo e uma fotografia do convento. A figura é mostrada com 100% de opacidade. Tal figura foi obtida no começo dos anos 80 e não há informações EXIF disponíveis para ela. A calibração foi feita utilizando-se pontos de fuga.



(b) O mesmo casamento mostrado na Figura (a), utilizando-se 30% de opacidade.

Figura 12:

Snaveley, N., Seitz, S. M., & Szeliski, R., 2006. Photo tourism exploring photo collections in 3d. Technical report, Microsoft at Microsoft Research, Redmond WA 98052, USA.

Taylor, C. J. & Kriegman, D. J., 1995. Structure and motion from line segments in multiple images. *IEEE Transactions On Pattern Analysis And Machine Intelligence*, vol. 17(11), pp. 1021–1032.

Trucco, E. & Verri, A., 1998. *Introductory Techniques for 3-D Computer Vision*. Prentice Hall, Inc.