

# Reconhecimento de Fachadas de Edifícios em Imagens de Ambientes Urbanos

José F. Martins  
FEUP, ISMAI  
R. Dr. Roberto Frias  
4200-425 Porto  
jfm@fe.up.pt

Jorge A. Silva  
FEUP, INEB  
R. Dr. Roberto Frias  
4200-425 Porto  
jsilva@fe.up.pt

A. Augusto de Sousa  
FEUP, INESC Porto  
R. Dr. Roberto Frias  
4200-425 Porto  
aas@fe.up.pt

---

## Sumário

*O Reconhecimento de Fachadas de Edifícios é uma operação necessária em diversas aplicações que funcionem em Ambientes Exteriores Urbanos, das quais se salientam as de Realidade Aumentada, nomeadamente para o cálculo da pose inicial da câmara. O método apresentado neste artigo realiza o reconhecimento de uma fachada de um edifício existente numa imagem e utiliza como conhecimento prévio apenas duas imagens da mesma, adquiridas em poses distintas. Não necessita de mais informação adicional relativa à fachada, seja ela o respectivo modelo 3D, as dimensões ou a razão de forma. O método demonstrou ser capaz de reconhecer uma grande diversidade de fachadas, incluindo as que não são estritamente planas, as parcialmente visíveis e as que possuem poucos elementos visuais facilmente distinguíveis dos demais.*

## Palavras-Chave

*Realidade Aumentada em Ambientes Urbanos, Reconhecimento de Fachadas, Visão por Computador.*

---

## 1. INTRODUÇÃO

O trabalho descrito neste artigo enquadra-se num trabalho mais vasto de desenvolvimento de aplicações de Realidade Aumentada para ambientes exteriores urbanos.

O objectivo da Realidade Aumentada (RA) é adicionar, em tempo real, objectos virtuais a imagens de uma cena real, por forma a que os referidos objectos pareçam fazer parte da cena [Barfield01].

A adição de objectos virtuais a uma imagem real só é visualmente convincente se aqueles estiverem perfeitamente alinhados com a cena real. Para que tal seja possível, é importante que sejam sintetizados com uma câmara virtual cuja pose seja, idealmente, igual à pose da câmara real. A determinação da pose da câmara, em relação a um referencial escolhido, pode ser feita por diversos métodos, recorrendo a um GPS, giroscópio, acelerómetro ou a outros sensores [Rolland01], mas o método que apresenta maior potencial é o que recorre às imagens adquiridas. Uma das formas de determinar a pose consiste em estabelecer correspondências entre os pontos de um plano cuja pose é conhecida (frequentemente, uma pose frontal), usada como referência, e pontos do mesmo plano na imagem cuja pose se quer determinar. Depois de estabelecidas as correspondências é possível estimar a pose por um dos métodos descritos em [Lepetit05].

Em cenas de interiores, em que o espaço visual é de pequenas dimensões, o plano de referência é, frequentemente, um padrão especialmente concebido para ser facilmente identificado nas imagens adquiridas (padrão sintetizado). Vários *toolkits* têm sido desenvolvidos que

se baseiam nesta técnica, entre os quais se destaca o ARTToolkit [Kato99,ARTToolkit09].

Em cenas exteriores, em que o espaço visual é amplo, a utilização de padrões sintetizados não é fácil, por diversos motivos, entre os quais a grande dimensão do padrão que seria necessário e a sua colocação no ambiente envolvente. Num ambiente exterior urbano, as fachadas dos edifícios são um dos elementos mais apelativos para este fim. O facto da fachada de muitos edifícios ser, em grande parte, constituída por uma superfície plana pode facilitar o estabelecimento de correspondências e a determinação da pose.

Para a determinação da pose em que uma determinada imagem da fachada foi adquirida é necessário reconhecer a fachada nessa imagem. Para tal, seria conveniente adquirir previamente uma imagem da fachada em pose frontal. Esta imagem é, na maior parte dos casos, impossível de conseguir. Em geral, a imagem de referência tem de ser adquirida a partir do solo em pose não frontal. No entanto, visto que as fachadas não são, usualmente, estritamente planas, uma só imagem de referência poderá não ser suficiente para um reconhecimento correcto.

O estabelecimento de correspondências entre duas imagens de uma fachada adquiridas em poses diferentes não é um processo trivial, por diversos motivos: em primeiro lugar pela pose ser diferente, depois porque as fachadas raramente são totalmente planas e, finalmente, porque existem diversos factores que podem causar diferenças entre as imagens adquiridas, como variações de iluminação, da luz reflectida, oclusões ou, até mesmo, alterações

do aspecto visual da fachada (por exemplo, quando uma persiana é fechada). Os métodos de emparelhamento baseados na extracção de pontos característicos (ou pontos de interesse) das imagens [Lepetit05] têm sido muito utilizados devido à sua robustez a variações de escala, ponto de vista e iluminação, bem como a oclusões parciais.

Os pontos característicos, detectados através de operadores adequados, são caracterizados por um descritor local, sendo o emparelhamento efectuado através da comparação desses descritores. No entanto, no caso de fachadas de edifícios, devido à existência de muitos elementos que se repetem (por exemplo, as janelas), a simples comparação de descritores pode dar origem a muitos emparelhamentos errados. É, por isso, necessário executar processamentos adicionais para refinar os emparelhamentos, procurando garantir que são estabelecidos correctamente. No método proposto, este refinamento é conseguido através da aplicação de restrições topológicas.

Neste artigo, é apresentado um método para efectuar o reconhecimento de uma fachada de um edifício numa imagem adquirida numa pose qualquer. Para que o reconhecimento possa ser efectuado é necessário, em primeiro lugar, garantir que a fachada contém uma superfície plana dominante e, em segundo lugar, adquirir previamente duas imagens dessa fachada, de pontos de observação diferentes. O método proposto divide-se em duas fases: uma fase de preparação e uma fase de reconhecimento. A fase de preparação tem como objectivo adquirir e preparar toda a informação necessária para a fase seguinte, nomeadamente, os pontos característicos da fachada e a sua topologia. A fase de reconhecimento consiste em verificar, com recurso a essa informação, se a fachada em causa está presente numa imagem adquirida numa pose arbitrária, estabelecendo correspondências entre os pontos característicos da imagem da fachada nesta pose e de uma das imagens adquiridas na fase de preparação.

O artigo desenvolve-se nas seguintes secções: na secção 2 é feita uma revisão de alguns métodos que têm sido propostos para o reconhecimento de fachadas de edifícios; nas secções 3 e 4, descrevem-se as duas fases do método proposto e a sua implementação; na secção 5 faz-se a avaliação do método e, finalmente, na secção 6 são retiradas as principais conclusões sobre o trabalho realizado.

## 2. TRABALHO RELACIONADO

No contexto da Realidade Aumentada, foram desenvolvidos, ao longo da última década, diversos trabalhos que permitem reconhecer uma fachada de um edifício e utilizá-la como padrão de referência para realizar o alinhamento. Destes trabalhos, salientam-se os seguintes.

Simon *et al.* [Simon02] desenvolveram um sistema de RA que utiliza, como padrões, uma ou mais superfícies planas existentes em ambientes exteriores (entre as quais se destacam as fachadas de edifícios) ou interiores.

A fase inicial do funcionamento deste sistema consiste num processo interactivo no qual o utilizador deve deli-

mitar, na imagem adquirida, uma ou mais regiões que correspondam a superfícies planas existentes na cena real.

A pose da câmara é calculada a partir de um conjunto de correspondências estabelecidas entre os pontos característicos detectados nessas regiões em duas imagens adquiridas consecutivamente.

Os movimentos bruscos da câmara causam a perda de alinhamento que só pode ser recuperado se o utilizador colocar a câmara numa pose semelhante à última pose conhecida. Nos exemplos apresentados, as fachadas utilizadas como padrões são praticamente planas e têm poucas ou nenhuma oclusões.

O sistema de RA apresentado em [Jiang04] é híbrido visto que, para além de utilizar as imagens adquiridas da cena real, também utiliza informação proveniente de um giroscópio acoplado à câmara, para obter o alinhamento. Este sistema permite reconhecer diversas fachadas pertencentes a um ou mais edifícios cujos modelos 3D sejam conhecidos de antemão. Os autores realçam que todas as linhas visualmente significativas existentes no edifício devem existir no respectivo modelo 3D, mesmo que não correspondam a arestas salientes desse edifício. Isto é justificado pelo facto da informação visual relativa às superfícies de um objecto tridimensional ser geralmente representada no respectivo modelo 3D através de texturas, ao invés de ser representada geometricamente.

A fase inicial do funcionamento deste sistema exige a intervenção do utilizador para identificar as correspondências existentes entre as arestas detectadas na imagem adquirida e as arestas pertencentes ao modelo 3D.

A pose da câmara é calculada a partir das correspondências que se estabelecem entre as arestas detectadas na imagem adquirida e as arestas existentes numa imagem sintetizada do modelo 3D.

Os movimentos bruscos da câmara não causam a perda de alinhamento, uma vez que nesta situação as imagens adquiridas são ignoradas e a pose da câmara é obtida através do giroscópio. O alinhamento também é mantido quando ocorrem oclusões de natureza dinâmica, que podem até ocultar a maior parte da fachada, desde que sejam temporárias, visto que a pose da câmara pode ser obtida com o giroscópio. Na eventual perda de alinhamento, este pode ser recuperado se o utilizador posicionar a câmara numa pose semelhante à última pose conhecida.

Em [Reitmayr06] é apresentado um sistema de RA híbrido que utiliza imagens da cena real, assim como informação proveniente de um giroscópio e de outros sensores para possibilitar o alinhamento.

O sistema permite reconhecer fachadas de um ou mais edifícios cujos modelos 3D sejam conhecidos previamente. Os modelos 3D são constituídos pelas arestas salientes mais significativas do edifício e os restantes elementos visuais não salientes são representados no respectivo modelo através de texturas.

Na fase inicial do funcionamento não é necessária a intervenção do utilizador, mas a sua posição inicial preci-

sa de ser conhecida. Esta limitação foi ultrapassada posteriormente [Reitmayr07] com o recurso a um GPS.

O cálculo da pose inicial da câmara (tanto no início do funcionamento do sistema como na recuperação da perda do alinhamento) é realizado a partir das correspondências estabelecidas entre pontos característicos detectados na imagem adquirida e em imagens sintetizadas do modelo 3D texturado. Após o cálculo da pose inicial, a estimação da pose nas imagens seguintes é realizada a partir de correspondências estabelecidas entre arestas detectadas na imagem e as mesmas arestas do modelo 3D texturado numa imagem sintetizada.

O alinhamento é mantido, mesmo com movimentos bruscos da câmara e quando ocorrem oclusões totais, o que é possível recorrendo à informação gerada pelo giroscópio e restantes sensores. Em caso de perda de alinhamento, este pode ser recuperado se a pose da câmara for aproximada a uma de dez poses anteriores conhecidas e distintas da câmara. A execução do método de recuperação do alinhamento conduz a uma redução do desempenho do sistema mas, em contrapartida, a recuperação é mais fácil do que quando é necessária uma aproximação à última pose conhecida da câmara [Simon02, Jiang04].

O sistema de RA de Xu [Xu08] é capaz de realizar o alinhamento num ambiente interior ou exterior mesmo que não existam superfícies planas dominantes.

Antes da fase activa da RA e no caso de ambientes interiores, é necessário adquirir duas ou mais imagens, a partir de posições distintas, da cena real com a qual se pretende alinhar os objectos virtuais. A pose da câmara é calculada através da detecção de um padrão sintetizado que deve ser colocado na cena e totalmente visível em, pelo menos, duas das imagens.

Nessas imagens são detectados pontos característicos. Visto que as poses da câmara são conhecidas em ambas as imagens e que se podem estabelecer correspondências entre os pontos nelas detectados é possível reconstruir os pontos 3D cujas projecções nas imagens consistem nos pontos característicos detectados inicialmente. Esta informação é crucial para a determinação da pose da câmara durante a fase activa da RA.

Num ambiente exterior, basta obter uma imagem da cena real desde que a maioria dos elementos nela existentes (que incluem as fachadas de edifícios) esteja distante da câmara. Para simplificação do problema, os pontos detectados são considerados como bidimensionais (existentes num plano imaginário distante da câmara).

Na fase activa da RA, a pose inicial da câmara deve ser próxima de uma das poses conhecidas da câmara (uma única pose, no caso de ambientes exteriores).

A pose da câmara é calculada a partir das correspondências entre os pontos característicos detectados na imagem adquirida da cena real e os mesmos pontos detectados na imagem inicial. Os autores afirmam que mesmo com a referida simplificação, é possível obter um bom alinhamento,

desde que os movimentos da câmara sejam suaves.

O método proposto no presente artigo, embora tendo sofrido alguma influência do trabalho de [Xu08], apresenta diferenças importantes em relação aos métodos analisados. Ao contrário de [Jiang04] e de [Reitmayr06], não exige o conhecimento prévio do modelo 3D da fachada. Esta é uma importante vantagem, uma vez que o modelo 3D poderá não existir ou não estar disponível e, nessa situação, a sua construção pode ser complexa e demorada. Além disso não são utilizados outros sensores, para além da câmara, como acontece nestes dois trabalhos. Nos exemplos apresentados em [Simon02] e [Xu08], as câmaras encontravam-se distantes da fachada. Aliás, em [Xu08] essa limitação é expressamente indicada. No método proposto, existe uma maior flexibilidade no posicionamento da câmara relativamente à fachada.

Está prevista a integração dos resultados do trabalho aqui apresentado num sistema de RA que permitirá que a posição inicial do utilizador seja arbitrária, ao contrário do que sucede em [Reitmayr06] e [Xu08] e que, para recuperação do alinhamento, não exigirá que a câmara seja colocada numa pose próxima da última pose conhecida, como sucede com [Simon02] e [Jiang04].

### 3. MÉTODO PROPOSTO

O método aqui apresentado encontra-se dividido em duas fases, a de preparação e a de reconhecimento. A primeira fase tem como objectivo adquirir e preparar toda a informação relativa à fachada que será utilizada na segunda fase para fazer o reconhecimento da mesma.

A fase de preparação consiste nas seguintes etapas: (a) aquisição de duas imagens da fachada, em poses distintas, tomadas de cada um dos lados da fachada; (b) identificação da região de interesse da fachada; (c) detecção de pontos característicos; (d) identificação de pontos característicos coplanares com o plano da fachada e (e) identificação de pontos característicos colineares.

A fase de reconhecimento é composta por três etapas: (a) aquisição de uma imagem, em pose desconhecida, que, em geral, deverá conter a fachada; (b) detecção de pontos característicos; (c) emparelhamento entre os pontos detectados na imagem adquirida e os pontos detectados nas duas imagens adquiridas na fase anterior. Este emparelhamento é feito em duas fases: numa primeira, apenas com base num descritor local, associado a cada ponto característico; numa segunda, são aplicadas restrições topológicas aos emparelhamentos resultantes da primeira fase que permitem melhorar significativamente a qualidade dos emparelhamentos finais. Se o número de emparelhamentos for inferior a um determinado limiar, considera-se que a fachada não se encontra presente na imagem adquirida ou, caso esteja, possa estar demasiado oculta ou distorcida pela perspectiva para ser reconhecida.

Nas secções seguintes descreve-se, com algum pormenor, as diferentes fases do algoritmo proposto.

### 3.1 Fase de Preparação

A informação principal resultante desta fase é a seguinte: a delimitação da região de interesse da fachada; a extração de pontos característicos dessa região que sejam coplanares com o plano dominante da fachada e a caracterização desses pontos. Esta caracterização é feita de duas formas: através de um vector individual de características que serão usadas para fazer o emparelhamento com a imagem adquirida na fase seguinte; através de características topológicas que indicam, relativamente a cada ponto, se ele pertence a alguma linha de pontos característicos e qual a sua posição relativa em relação a todas as outras linhas de pontos característicos existentes.

#### 3.1.1 Aquisição das Imagens

É necessária a aquisição de duas imagens da fachada, doravante designadas por  $I_L$  e  $I_R$ . Estas imagens devem ser obtidas a partir de posições diferentes, preferencialmente de ambos os lados da fachada (Figura 1), pois isso facilitará a identificação dos pontos que são coplanares com o plano dominante da fachada (3.1.4).

#### 3.1.2 Delimitação da Região de Interesse

A região de interesse da fachada tem de ser delimitada em  $I_L$  e  $I_R$ . Esta etapa é necessária para a identificação do plano dominante da mesma (3.1.4) permitindo também restringir as regiões das imagens adquiridas nesta fase onde serão detectados os pontos característicos (3.1.3).



Figura 1: Delimitação manual da região de interesse.

A delimitação é feita manualmente indicando os cantos de dois quadriláteros, em  $I_L$  e  $I_R$ , que correspondam a um mesmo rectângulo em pose frontal (Figura 1). Este rectângulo deve ser coplanar com o plano dominante, pertencer à fachada e englobar a maior parte da mesma (idealmente, toda a fachada). No entanto, o rectângulo não necessita de existir visualmente na fachada, podendo somente existir algumas das suas arestas ou vértices.

#### 3.1.3 Detecção dos Pontos Característicos

Para realizar o reconhecimento da fachada numa qualquer imagem adquirida, que designaremos por  $I_U$ , é necessário estabelecer correspondências entre uma imagem de referência ( $I_L$  ou  $I_R$ ), na qual a fachada está identificada e a imagem  $I_U$ . Estas correspondências são estabelecidas entre pontos característicos das duas imagens. Nas imagens de referência, basta extrair os pontos característicos da região de interesse.

A detecção é realizada através do algoritmo apresentado em [Bay06]. Este algoritmo, para além de identificar as coordenadas dos pontos, disponibiliza um descritor da sua vizinhança, na imagem, constituído por um vector de

características relativas à distribuição espacial de intensidades. Este descritor tem a particularidade de ser invariante a diversos factores, como, por exemplo, a rotação, o escalamento e a iluminação. Este facto é muito importante, uma vez que contribui para um número mais elevado de emparelhamentos correctos entre imagens adquiridas em poses distintas.

#### 3.1.4 Identificação dos Pontos Coplanares com o Plano Dominante

Alguns dos pontos característicos detectados não são coplanares com o plano dominante. Isto sucede porque esses pontos podem pertencer a elementos da fachada que se encontram recuados/avançados relativamente a esse plano e, em certos casos, podem até nem sequer pertencer à própria fachada, mas a objectos que a ocluem como árvores e postes.

Uma vez que o método de reconhecimento da fachada se baseia na utilização de dois conjuntos de pontos coplanares, é fundamental garantir que se utilizam unicamente pontos pertencentes ao plano dominante. É necessário, por isso, identificar esses pontos e eliminar os restantes.

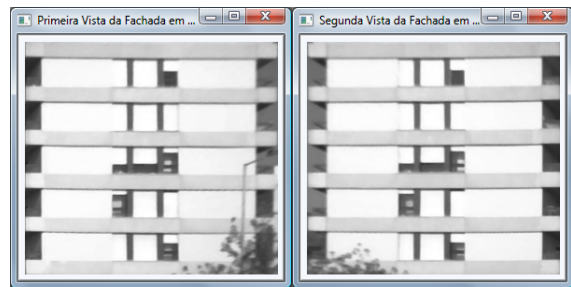


Figura 2: Regiões de interesse em pose frontal.

O processo de identificação dos pontos coplanares com o plano dominante faz uso de  $I_L$  e  $I_R$  e é realizado em duas etapas. Na primeira etapa, as duas regiões de interesse da fachada são transformadas para pose frontal (Figura 2), sendo designadas por  $F_L$  e  $F_R$ , respectivamente. Para realizar esta transformação, seria necessário conhecer as dimensões reais dessas regiões ou, alternativamente, a razão de forma. Em geral, as dimensões são desconhecidas e, devido à distorção de perspectiva, não é possível determinar directamente a razão de forma a partir das dimensões do quadrilátero que delimita a região de interesse. O cálculo da razão de forma de um rectângulo a partir de uma imagem do mesmo numa pose não frontal é feito recorrendo ao método descrito em [Martins07].

Na segunda etapa, os pontos característicos detectados são projectados para uma pose frontal. Após esta projecção, os pontos característicos pertencentes ao plano dominante que sejam detectados em  $I_L$  e  $I_R$  podem ser facilmente identificados, pois as suas posições devem ser coincidentes em  $F_L$  e  $F_R$ . Isto não acontece com pontos característicos que estejam afastados do plano dominante. No entanto, pode acontecer que um determinado elemento característico da fachada dê origem a um ponto característico numa imagem mas não na outra. A opção por

esta solução limitaria os pontos elegíveis aos detectados em ambas as imagens.

Para identificar o maior número possível de pontos pertencentes ao plano dominante, optou-se por uma solução que não necessita que um determinado ponto seja detectado nas duas imagens. Para um dado ponto existente em  $I_L$  ou  $I_R$  é calculada a sua projecção em  $F_L$  e  $F_R$ ; as vizinhanças de ambas as projecções são comparadas usando correlação normalizada e se forem suficientemente semelhantes então o ponto é identificado como pertencente ao plano dominante. Se o ponto pertencer a um elemento visual que não seja coplanar com o plano dominante da fachada, como, por exemplo, um poste de iluminação, então as respectivas vizinhanças em  $F_L$  e  $F_R$  serão visualmente distintas (Figura 3).

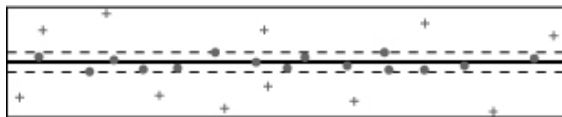


**Figura 3: Detecção de pontos característicos não-coplanares com a fachada.**

### 3.1.5 Identificação de Pontos Colineares

Uma grande abundância de elementos visuais muito semelhantes, que podem facilmente ser confundidos entre si, contribui para o estabelecimento de um elevado número de emparelhamentos errados. Uma solução possível para este problema, consistiria em só considerar os pontos singulares (pontos cujo descritor não é semelhante ao de mais nenhum ponto característico). No entanto, no caso concreto de uma fachada, esta solução poderia reduzir o número de pontos utilizáveis.

Uma alternativa consiste em associar informação de natureza topológica, a cada ponto, que o torne mais facilmente distinguível dos demais. Uma das características de uma fachada, que é útil nesta situação, reside no facto de muitos dos seus elementos característicos visuais estarem dispostos segundo linhas horizontais/verticais. Nas imagens  $F_L$  e  $F_R$ , estas linhas devem estar orientadas segundo linhas/colunas da imagem, em torno das quais deverão surgir elevadas concentrações de pontos característicos.



**Figura 4: Contabilização do número de pontos característicos na vizinhança de uma linha da imagem.**

As linhas/colunas em cuja vizinhança existir um número de pontos característicos superior a um limiar estabelecido (Figura 4) serão usadas como referência para a caracterização topológica dos pontos característicos da região de interesse.

Esta caracterização consiste em associar a cada ponto duas informações importantes: a que linha(s) horizontal

e/ou vertical pertence<sup>1</sup> e qual o lado em que se encontra relativamente a cada uma das linhas a que não pertence.

A utilização desta informação topológica permite reduzir significativamente o número de emparelhamentos errados.

Todos os pontos que não estiverem associados a nenhuma das linhas identificadas são eliminados.

## 3.2 Fase de Reconhecimento

Esta fase consiste em detectar se a fachada seleccionada na fase de preparação está presente numa imagem adquirida numa pose desconhecida ( $I_U$ ). Para isso, os pontos característicos resultantes da fase de preparação são emparelhados com os pontos característicos de  $I_U$ , tendo em conta os descritores que lhes estão associados (3.1.3) e a caracterização topológica dos pontos (3.1.5). Descrevem-se a seguir as etapas principais desta fase.

### 3.2.1 Aquisição da Imagem

A primeira etapa consiste em adquirir a imagem  $I_U$  na qual se pretende reconhecer a fachada seleccionada na fase de preparação. A pose desta imagem é desconhecida, bem como a existência da própria fachada na imagem.

### 3.2.2 Detecção de Pontos Característicos

Para que a fachada possa ser reconhecida em  $I_U$  é necessário emparelhar os pontos característicos dessa imagem com os pontos que foram detectados na fase anterior. A detecção de pontos característicos é, nesta etapa, feita sobre toda a imagem.

### 3.2.3 Emparelhamento

Para maximizar o número de emparelhamentos correctos, estabelecem-se dois conjuntos de emparelhamentos entre os pontos característicos de  $I_U$  e os de  $I_L$  e  $I_R$ .

Estes emparelhamentos são estabelecidos através do método apresentado em [Beis97]. Este método faz uso dos descritores associados a cada ponto, com o objectivo de tentar encontrar para cada ponto na primeira imagem o ponto mais semelhante na segunda imagem (4.6).

Se nenhum dos dois conjuntos de emparelhamentos possuir um número mínimo de pares assume-se que a fachada não foi detectada. Caso contrário, considera-se que a imagem ( $I_L$  e  $I_R$ ) mais próxima de  $I_U$  é aquela que deu origem ao conjunto de emparelhamentos com maior cardinal. Este conjunto é o ponto de partida para a etapa seguinte.

### 3.2.4 Eliminação de Emparelhamentos Errados

Geralmente, da etapa anterior, resultam alguns emparelhamentos errados. O seu número é variável e poderá ser afectado por uma grande diversidade de factores, como: a diferença entre as poses da câmara; a variação da iluminação; o descritor utilizado para caracterizar os pontos; o método utilizado no emparelhamento e, principalmente, o conteúdo das duas imagens cujos pontos se pretende

<sup>1</sup> Um ponto característico pode pertencer: a uma só linha horizontal, a uma só linha vertical ou, simultaneamente, a uma linha horizontal e a uma linha vertical.

emparelhar. Seria, pois, conveniente eliminar os emparelhamentos errados.

Um método muito utilizado para resolver este problema, descrito em [Hartley03], consiste em recorrer ao algoritmo RANSAC [Fischler81] para o cálculo robusto da homografia que relaciona duas imagens de um mesmo plano. Este método começa por seleccionar alguns emparelhamentos ao acaso e, com eles, calcula a respectiva homografia. Usando esta homografia, cada ponto do plano na primeira imagem é projectado na segunda imagem e é calculada a distância entre a projecção do ponto e o seu par resultante do emparelhamento. Se a distância for superior a um dado limiar, o emparelhamento é considerado errado. Este processo é repetido até se conseguir um número total de emparelhamentos correctos que possa ser considerado aceitável. Se esse número não for atingido conclui-se que não é possível calcular a homografia em questão de uma forma robusta, o que implicaria o não reconhecimento da fachada. Este método produz, geralmente, bons resultados, mesmo em situações em que a percentagem de emparelhamentos errados seja elevada. No entanto, no caso das fachadas, a sua utilização pode, em certos casos, produzir resultados insatisfatórios.

Esta situação pode ocorrer pelo facto de se partir do princípio que os emparelhamentos correctos apresentam um comportamento comum e regular que os distingue dos errados que possuem, usualmente, um comportamento aleatório. No entanto, nas fachadas, os emparelhamentos errados podem apresentar um comportamento comum e regular, como, por exemplo, quando um conjunto de pontos característicos de um andar emparelha com os seus análogos de outro andar. Se o número destes pares for suficientemente elevado, poderão ser tomados, erradamente, como correctos e assim determinar uma homografia incorrecta.

A solução aqui proposta para a eliminação de emparelhamentos errados recorre à caracterização topológica dos pontos característicos que foi feita na fase de preparação e consiste na aplicação de duas restrições, aos emparelhamentos estabelecidos, que designaremos por restrição de colinearidade e restrição de lateralidade.



**Figura 5: Linha horizontal obtida e pontos emparelhados incorrectamente (símbolo “+”).**

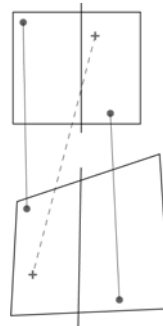
A primeira destas restrições resulta da condição de colinearidade entre os pontos coplanares ser invariante à distorção da perspectiva, isto é, os pontos que sejam colineares em  $I_L$  ou  $I_R$  também o deverão ser em  $I_U$ . Portanto, se um ponto não respeitar esta condição é porque o seu correspondente não foi determinado correctamente, devendo esse par de pontos ser eliminado do conjunto de emparelhamentos estabelecidos (Figura 5).

A verificação desta restrição é feita da forma que se descreve a seguir.

Tomam-se os pontos de  $I_U$  que foram emparelhados com um conjunto de pontos colineares de  $I_L$  ou  $I_R$  (consoante a imagem usada no emparelhamento) e calcula-se a recta que melhor aproxima esses pontos, recorrendo ao algoritmo RANSAC. A recta é considerada válida se tiver associado um número mínimo de pontos e todos os pontos que forem rejeitados como não pertencentes à recta, são considerados como pontos que violam a restrição de colinearidade.

Alguns dos emparelhamentos errados que ainda restarem podem ser eliminados por aplicação da restrição de lateralidade. A aplicação desta restrição consiste em verificar se dois pontos característicos correspondentes ocupam, nas respectivas imagens, a mesma posição relativamente às linhas de pontos característicos a que não pertencem.

Assim, para cada linha calculada através de RANSAC e para cada ponto não associado a essa linha é verificado se o lado em que o ponto se encontra relativamente à linha em  $I_U$  coincide com o lado em que ele se encontra relativamente à mesma linha em  $F_L$  ou  $F_R$  (Figura 6). Esta verificação é repetida para cada uma das linhas identificadas. Idealmente, bastaria que os lados não coincidisse uma só vez para que o ponto e o respectivo emparelhamento fossem eliminados. No entanto, poderão ocasionalmente existir linhas mal determinadas, o que pode provocar, erradamente, que os lados em que o ponto se encontra relativamente à linha não coincidam. Como forma de contornar esta situação, é contabilizado, para cada ponto, o total de vezes que os lados coincidem e o total de vezes em que tal não ocorre. No final, se o primeiro total for inferior ao dobro do segundo total, o ponto e o respectivo emparelhamento é eliminado.



**Figura 6: Verificação da restrição de lateralidade.**

### 3.2.5 Delimitação da Região de Interesse na Imagem Adquirida

Para delimitar a região de interesse da fachada em  $I_U$  projecta-se sobre esta imagem os cantos da região de interesse, identificados em  $I_L$  ou  $I_R$ . Para realizar essa operação, é necessário calcular a homografia, que relaciona ambas as imagens com base nos emparelhamentos que subsistiram da fase anterior. Se o número de emparelhamentos resultantes do processo de eliminação descrito na secção anterior não for suficiente para calcular esta homografia considera-se que a fachada não foi detectada na imagem adquirida. Conclui-se, desta forma, o processo de reconhecimento da fachada na imagem  $I_U$ , adquirida numa pose desconhecida.

## 4. IMPLEMENTAÇÃO

O método apresentado foi implementado, em grande parte, recorrendo a funcionalidades existentes na API

OpenCV [OCV09] muito utilizada no desenvolvimento de aplicações nos domínios da Visão por Computador.

A implementação divide-se em duas componentes fundamentais, a preparação e o reconhecimento, correspondentes às duas fases do método que foram descritas nas secções 3.1 e 3.2, respectivamente.

O desenvolvimento foi realizado num PC portátil com um processador *Intel Core2 T8300* a 2.4 GHz, 3 GB de RAM, e uma placa gráfica *nVidia GeForce 8600M GS* com 512 MB de memória. Na aquisição de imagens foi usada uma *webcam Logitech QuickCam Pro 9000*.

No arranque da aplicação desenvolvida é dada ao utilizador a possibilidade de calibrar a câmara, recorrendo a funcionalidades da OpenCV [Martins07]. A calibração da câmara determina os seus parâmetros intrínsecos, nomeadamente, os coeficientes de distorção das lentes. Esta informação é utilizada para a remoção da distorção das lentes em todas as imagens adquiridas e, futuramente, os restantes parâmetros intrínsecos, para o cálculo da pose da câmara.

Nas subsecções seguintes descrevem-se os aspectos mais relevantes da implementação de ambas as fases, que ainda não foram referidos anteriormente.

#### 4.1 Aquisição e Pré-processamento de Imagem

Todas as imagens foram adquiridas com uma resolução de 640×480 e pré-processadas no sentido de remover a distorção das lentes, reduzir o ruído e normalizar a intensidade da imagem, como foi descrito em [Martins07].

#### 4.2 Delimitação da Região de Interesse

A delimitação da região de interesse da fachada nas duas imagens iniciais deverá ser realizada manualmente pelo utilizador. Para aumentar a precisão deste processo, é aplicado o algoritmo de detecção de pontos característicos de Harris [Harris88] a uma região circular com um raio de 5 *pixels*, centrada em cada ponto seleccionado pelo utilizador. Se nessa região for detectado um ponto característico, é dada ao utilizador a possibilidade de optar entre o ponto que seleccionou e o ponto característico detectado.

#### 4.3 Detecção dos Pontos Característicos

Esta detecção é realizada recorrendo ao algoritmo SURF (*Speeded Up Robust Features*) [Bay06]. Este algoritmo detecta pontos característicos numa imagem, formando, para cada ponto, um vector de características (descriptor) que descreve a respectiva vizinhança. É um dos algoritmos de detecção mais rápidos [Bastos08] e possui um grande potencial para aplicações de RA.

A implementação OpenCV do SURF fornece, para cada ponto característico detectado, um conjunto de valores: as suas coordenadas na imagem, um descriptor da sua vizinhança, o sinal do Laplaciano e o tamanho, a orientação e o valor do Hessiano da vizinhança do ponto. Esta implementação também oferece a possibilidade de escolha entre um descriptor de 64 características ou um de 128 características. Com o intuito de obter a maior percentagem de emparelhamentos correctos, optou-se pelo descriptor mais longo.

Em todas as detecções realizadas foi imposto um limite mínimo de 500 no valor do Hessiano dos pontos característicos. Para valores substancialmente inferiores (por exemplo, 200), o número de pontos detectados é muito elevado, o que torna o processo de emparelhamento demorado. Para valores consideravelmente elevados (por exemplo, 1000), o número final de pontos poderá não ser suficientemente elevado para realizar um reconhecimento preciso da fachada.

Adicionalmente, para uma maior rapidez de emparelhamento e para uma maior percentagem de emparelhamentos correctos, são filtrados todos os pontos que possuam um tamanho inferior a um limite mínimo. Para garantir uma melhor distribuição espacial dos pontos pela região de interesse, é imposta uma distância mínima de 3 *pixels* entre os pontos detectados, evitando concentrações dos mesmos que tornariam o emparelhamento mais demorado. Valores superiores desta distância reduzem substancialmente o número de pontos característicos utilizáveis.

#### 4.4 Identificação dos Pontos Coplanares com o Plano Dominante

A identificação dos pontos coplanares com o plano dominante implica a comparação das vizinhanças dos pontos característicos, em  $F_L$  e  $F_R$  (3.1.4), através de uma correlação normalizada. Se o índice de semelhança for inferior a 85%, o ponto é considerado como não pertencente ao plano dominante e, por conseguinte, eliminado.

#### 4.5 Identificação dos Pontos Colineares

Para cada linha/coluna de *pixels*, de  $F_L$  e  $F_R$ , é totalizado o número de pontos característicos que pertencem a uma manga centrada nessa linha/coluna e cuja largura é 5% da largura/altura de  $F_L$  e  $F_R$ . Para todas as linhas/colunas que possuam um total de pontos característicos superior ou igual a 12, os respectivos pontos passam a ser considerados como pertencentes à vizinhança dessa linha/coluna. Os restantes pontos característicos não pertencentes à linha/coluna em questão são caracterizados quanto ao lado em que se encontram relativamente a essa linha/coluna.

O número de pontos escolhido (12) teve em conta que para um número de pontos muito elevado (por exemplo, 20) o número de linhas utilizáveis seria reduzido e para um número de pontos muito baixo (por exemplo, 4), a probabilidade da linha ser determinada incorrectamente na fase de emparelhamento (3.2.4) aumentaria.

#### 4.6 Emparelhamento

O emparelhamento é realizado entre o conjunto de pontos detectados em  $I_U$  e os conjuntos detectados em  $I_L$  e  $I_R$ .

Para tornar o processo de emparelhamento mais expedito, estes conjuntos são divididos em dois subconjuntos, consoante o sinal do Laplaciano. Visto que um mesmo elemento visual detectado em duas imagens produz, geralmente, dois pontos característicos com o mesmo sinal, é desnecessário tentar emparelhar pontos característicos com sinais diferentes.

Os emparelhamentos entre os quatro pares de subconjuntos são estabelecidos recorrendo ao método BBF (*Best-Bin-First*) [Beis97]. Este método utiliza uma árvore binária

ria do tipo k-d [Friedman77] para armazenar os descritores dos pontos de cada imagem. Os emparelhamentos são determinados através de uma pesquisa limitada a uma pequena parte da árvore, o que contribui para a rapidez deste método. O método garante uma elevada probabilidade de encontrar o emparelhamento correcto, uma vez que a pesquisa é realizada de forma a dar sempre prioridade aos ramos que conduzem a pontos cujos descritores apresentam uma distância mais reduzida ao descritor do ponto que se pretende emparelhar. Neste caso, apenas são mantidos os pares cuja distância euclidiana for inferior a 3.

O par (imagem actual, imagem inicial) que der origem a um maior número de emparelhamentos é o escolhido (3.2.3). No entanto, se esse número for inferior a 20, considera-se que a fachada não foi detectada em  $I_U$ .

Visto que o conjunto de emparelhamentos escolhido possui, geralmente, emparelhamentos errados, estes deverão ser eliminados. Esta eliminação é realizada recorrendo à informação topológica associada a cada ponto (3.1.5).

Para cada uma das linhas horizontais/verticais identificadas na fase anterior é totalizado o número de pontos pertencentes a essa linha que foram detectados e emparelhados em  $I_U$ . Se esse número for superior ou igual a 8 pontos, a equação da linha será determinada. Se os pontos dessa linha estiverem correctamente emparelhados com os seus pares em  $I_R$  e  $I_L$  então estes pontos também devem ser quase colineares. Para identificar quais desses pontos respeitam essa condição, é necessário calcular uma recta que passe na proximidade de, pelo menos, 40% deles. Esta percentagem foi escolhida tendo em conta que um valor muito elevado (por exemplo, 80%) contribui para o número de rectas determinadas seja reduzido e uma percentagem muito baixa (por exemplo, 20%) aumenta a probabilidade de determinar incorrectamente a recta. Esta recta é calculada através do algoritmo RANSAC (3.2.4). Se um ponto estiver demasiado afastado da recta calculada então é porque não respeita a condição de quase colinearidade imposta e portanto é eliminado visto que resulta de um emparelhamento errado. O ponto é considerado demasiado afastado da recta se ultrapassar uma determinada distância que é função da proximidade do utilizador à fachada (quanto mais afastado, menor será esse valor). Essa distância é o produto da largura da manga centrada na linha (4.5), em pose frontal, por um factor de escala que relaciona o tamanho da fachada em  $F_L$  ou  $F_R$  e o tamanho da mesma em  $I_U$ . Esse factor de escala é calculado como a razão entre a distância média entre duas rectas paralelas, em  $I_U$ , que possuam o maior número de pontos emparelhados e a distância dessas mesmas rectas em  $F_L$  ou  $F_R$ .

A determinação do lado em que um ponto (x,y) se encontra em relação a qualquer uma das rectas calculadas, descrita por uma equação na forma  $Ax+By+C=0$ , é feita substituindo os valores de x e y no primeiro membro da equação e analisando o sinal do resultado.

#### 4.7 Delimitação da Região de Interesse na Imagem Adquirida

A homografia entre os pontos detectados numa das imagens iniciais ( $I_L$  ou  $I_R$ ) e os seus pares em  $I_U$  é calculada através do algoritmo descrito em [Hartley03], utilizando somente os emparelhamentos que não foram eliminados na fase anterior. O cálculo da homografia é realizado recorrendo ao algoritmo RANSAC. Na implementação OpenCV, o número de emparelhamentos utilizados, com o intuito de encontrar a homografia que minimize o erro de reprojeção, é igual a 5. Com o objectivo de reduzir ainda mais o erro de reprojeção, a homografia é posteriormente refinada através do método de Levenberg-Marquardt [Hartley03].

Usando a homografia calculada, os cantos da região de interesse delimitada em  $I_L$  e  $I_R$  são projectados em  $I_U$  e, assim, é possível delimitar a região de interesse nesta última imagem. Visto que a região delimitada em  $I_L$  e  $I_R$  é sempre um quadrilátero convexo, a região análoga existente em  $I_U$  também o deverá ser. Portanto, se essa região não for convexa em  $I_U$ , conclui-se que a fachada não foi reconhecida correctamente.

#### 5. AVALIAÇÃO

A avaliação do método proposto foi realizada sobre três sequências de imagens relativas a três edifícios distintos.

O edifício da primeira sequência (Figura 8) possui uma fachada cujo plano dominante é constituído por uma grande superfície central. Esta fachada também possui regiões não coplanares (as varandas) e muitos elementos repetidos. As duas imagens iniciais usadas na fase de preparação são a 6 e 10. Na segunda sequência (Figura 9), o edifício possui uma fachada cujo plano dominante corresponde a metade da superfície total da fachada, o que produz um número elevado de elementos não coplanares com aquele plano. As duas imagens iniciais usadas na preparação são a 2 e 7. Nestas duas sequências, surgem oclusões de pequenas regiões da fachada. Na terceira sequência (Figura 10), a fachada do edifício é rica em detalhes visuais distintos, alguns dos quais não coplanares. As imagens desta sequência usadas na fase de preparação são as duas primeiras.

Na primeira sequência, a fachada foi detectada em todas as imagens, excepto nos casos extremos correspondentes às imagens 1 e 14. Estas são as imagens da sequência em que a fachada se encontra mais distante da câmara e o seu eixo óptico forma um ângulo mais reduzido com o plano dominante da fachada. Estes dois factores contribuíram decisivamente para o não reconhecimento da fachada. No entanto, a fachada foi reconhecida mesmo quando visualizada parcialmente como sucede na imagem 8, algo que também ocorre na segunda sequência, nomeadamente na imagem 1. Na terceira sequência, a câmara afasta-se progressivamente da fachada e o reconhecimento só falhou quando a distância tornou-se demasiado grande (imagem 9).

Uma vez que o método necessita que o utilizador selecione manualmente os cantos das regiões de interesse em



$I_L$  e  $I_R$ , durante esta avaliação concluiu-se que a precisão desta selecção é muito importante para a correcta determinação da região em  $I_U$ . Esta importância resulta do facto dos cantos do quadrilátero que delimita a região interesse da fachada, na fase de reconhecimento, serem essencialmente uma projecção dos pontos identificados pelo utilizador, na fase de preparação. Portanto, se o utilizador seleccionar erroneamente um dos cantos da fachada, então esse canto surgirá erroneamente reprojectado em  $I_U$  relativamente ao canto real usado como referência. No entanto, a aplicação é geralmente capaz de corrigir pequenos erros de selecção por parte do utilizador (4.2) e somente os erros mais grosseiros poderão causar um impacto negativo mais notório no resultado final.

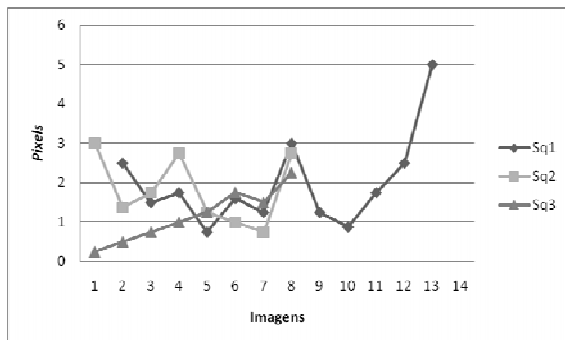


Figura 7: Gráfico do erro de reprojeção da região de interesse da fachada nas três sequências.

Em todas as sequências, a média do erro de reprojeção da fachada é inferior a 2 *pixels*, respectivamente: 1.98 *pixels*, 1.83 *pixels* e 1.16 *pixels* (Figura 7). O erro de reprojeção da fachada, em cada imagem, consiste na média das distâncias, em *pixels*, entre os cantos projectados em  $I_U$  e os cantos reais análogos existentes na mesma imagem.

O tempo de cálculo necessário, em média, para a execução do algoritmo, na fase do reconhecimento, é inferior a 350 milissegundos. Será, no entanto, possível reduzir este tempo, através da optimização de algumas das etapas.

## 6. CONSIDERAÇÕES FINAIS

Neste artigo apresentou-se um método de reconhecimento de fachadas de edifícios com o objectivo de resolver o problema do cálculo da pose inicial da câmara em aplicações de RA.

O método proposto realiza o reconhecimento de fachadas em imagens adquiridas em poses variadas, como se demonstrou nos exemplos apresentados. O erro de reprojeção da fachada é pequeno, não só dentro do espaço angular delimitado pelas duas posições em que foram adquiridas as imagens na fase de preparação, como também em posições afastadas desse espaço.

O método consegue reconhecer fachadas que sejam constituídas por muitos elementos repetidos, sofram pequenas oclusões e possuam diversos elementos não coplanares com o respectivo plano dominante. Uma vez que o méto-

do se baseia exclusivamente em imagens adquiridas, verificou-se que não funciona quando há grandes oclusões.

Apesar de, nesta fase, o método não ter sido implementado com o intuito de funcionar em tempo real possui, ainda assim, o potencial para ser integrado num sistema de RA para ambientes exteriores urbanos. Neste sistema, o método proposto poderá ser utilizado para calcular a pose inicial da câmara, no início da fase activa da RA, e para a recuperação de uma eventual perda de alinhamento.

## 7. REFERÊNCIAS

- [ARToolkit09] ARToolkit Website, Setembro de 2009. <<http://www.hitl.washington.edu/artoolkit>>
- [Barfield01] W. Barfield e T. Caudell. Basic Concepts in Wearable Computers and Augmented Reality. Fundamentals of Wearable Computers and Augmented Reality, LEA Publishers, pp 3-26, 2001.
- [Bastos08] R. Bastos e J. Dias. FIRST, Invariant Image Features for Augmented Reality and Computer Vision, ISCTE, pp 20-38, 2008.
- [Bay06] H. Bay, T. Tuytelaars e L. V.Gool. SURF: Speeded Up Robust Features. 9th European Conference on Computer Vision, pp 404-417, 2006.
- [Beis97] J.S. Beis e D.G. Lowe. Shape Indexing Using Approximate Nearest-Neighbor Search in High-Dimensional Spaces. IEEE Conf. CVPR, 1997.
- [Fischler81] M. A. Fischler e R. C. Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Comm. ACM 24, pp 381-395, 1981.
- [Friedman77] J. Friedman, J. Bentley e R. Finkel. An Algorithm for Finding Best Matches in Logarithmic Expected Time. ACM Trans. Math. Software, 3:209-226, 1977.
- [Harris88] C. Harris e M. Stephens. A Combined Corner and Edge Detector. Proceedings of the 4th Alvey Vision Conference, pp 147-151, 1988.
- [Hartley03] R. Hartley e A. Zisserman. Multiple View Geometry in Computer Vision (Second Edition). Cambridge University Press, 2003.
- [Jiang04] B. Jiang, U. Neuman e S. You. A Robust Tracking System for Outdoor Augmented Reality. IEEE VR, pp 27-31, 2004.
- [Lepetit05] V. Lepetit e P. Fua. Monocular Model-based 3D Tracking of Rigid Objects: A Survey. Foundations and Trends® in Computer Graphics and Vision. Vol. 1, No.1, pp 1-89, 2005.
- [Kato99] H. Kato e M. Billinghurst. Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System. IWAR, pp 85-94, 1999.
- [Martins07] J. Martins, J. Silva e A. Sousa. Sistema de Realidade Aumentada Baseado em Padrões Rectangulares de Dimensões Desconhecidas. 15º EPCG, pp 59-67, 2007.
- [OCV09] Open CV Website, Setembro de 2009.

<<http://www.opencv.org>>

[Reitmayr06] G. Reitmayr e T.W. Drummond. Going out: Robust Model-based Tracking for Outdoor Augmented Reality. ISMAR' 06, pp 109-118, 2006.

[Reitmayr07] G. Reitmayr e T.W. Drummond. Initialisation for Visual Tracking in Urban Environments. ISMAR'07, pp 161-172, 2007.

[Rolland01] J. P. Rolland, L. D. Davis e Y. Baillet. A Survey of Tracking Technology for Virtual Environments. Fundamentals of Wearable Computers and

Augmented Reality, LEA Publishers, pp 67-112, 2001.

[Simon02] G. Simon e M-O. Berger. Reconstructing While Registering: a Novel Approach for Markerless Augmented Reality. ISMAR' 02, pp 285-293, 2002.

[Xu08] K. Xu, K. W. Chia e A. D. Cheok. Real-time Camera Tracking for Marker-less and Unprepared Augmented Reality Environments. Image and Vision Computing, pp 673-689, 2008.

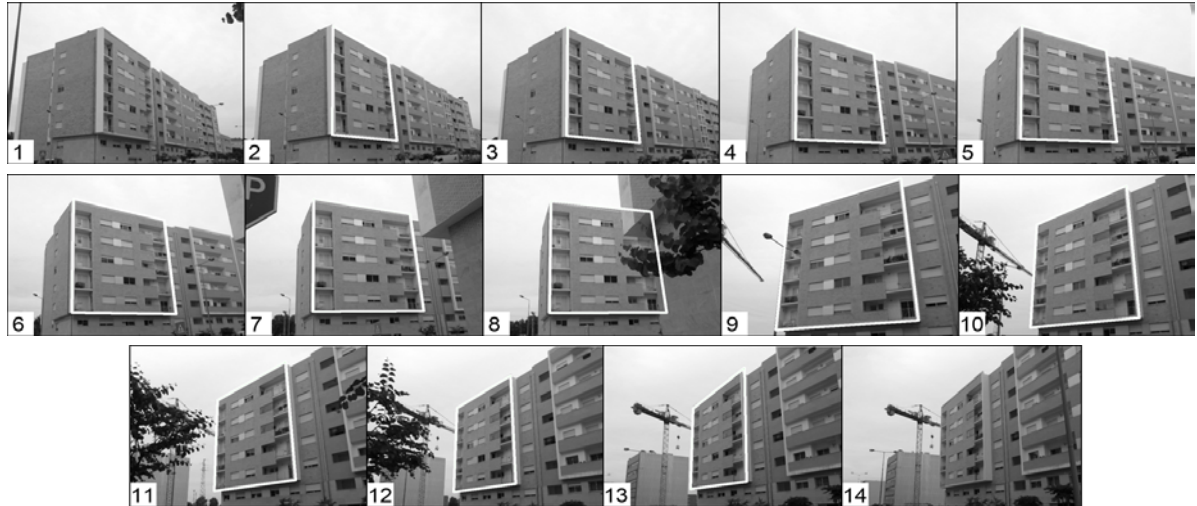


Figura 8: Primeira Sequência da Avaliação de Resultados.



Figura 9: Segunda Sequência da Avaliação de Resultados.



Figura 10: Terceira Sequência da Avaliação de Resultados.