

Adriano Barreto Huguet

Orientador: Arnaldo de Albuquerque Araújo

Coorientador: Rodrigo Lima Carceroni

# **RECONSTRUÇÃO DE CENAS URBANAS BASEADA EM ESTEREOSCOPIA E SEGMENTAÇÃO POR WATERSHED**

Dissertação apresentada ao Curso de Pós-Graduação em Ciência da Computação da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de mestre em Ciência da Computação.

UNIVERSIDADE FEDERAL DE MINAS GERAIS  
Instituto de Ciências Exatas  
Departamento de Ciência da Computação

Belo Horizonte  
2003

2003  
Adriano B. Huguet  
Todos os direitos Reservados

Dedico este trabalho aos meus pais, Rosa Maria Barreto Huguet e Luiz Augusto Simões Huguet, os quais sempre serão exemplo de sabedoria em minha vida.

# Agradecimentos

À minha esposa, Thaís Murta de Oliveira Huguet, pelo carinho e amor tão indispensáveis em todos os momentos de minha vida.

Ao Prof. Dr. Arnaldo de Albuquerque Araújo por todo o auxílio e orientação dedicados não somente no curso de Pós-Graduação como no curso de Graduação nesta mesma entidade. Ao Arnaldo um eterno abraço.

Ao Prof. Dr. Rodrigo Lima Carceroni, por toda a dedicação ao desenvolvimento de meu mestrado, prestando auxílio indispensável nas pesquisas e publicações.

Ao colega Bernardo Faria por ter aberto a minha primeira porta.

Aos companheiros do Núcleo de Processamento Digital de Imagens por toda a amizade e solidariedade recebidos nestes últimos 3 anos de convivência.

À turma do Xunil, Leonardo Gadelha, Alberto Rubens Beckler, Júlio César Soares Nunes e Fabrício Canez, por terem feito minha vida mais alegre e espirituosa.

À turma do GVX, os quais têm sido meus amigos por toda a minha vida e nunca me abandonaram nem negaram qualquer acolhimento. Me sinto honrado de tê-los ao meu lado.

A todos os colegas, funcionários e professores da UFMG que direta ou indiretamente contribuíram para a minha dissertação.

## RESUMO

---

Nesta dissertação trata-se do problema de estereoscopia sobre fotografias aéreas de baixa altitude e conseqüente alta resolução. O problema relacionado a este tipo de imagens é a existência de grandes áreas ocultas, devido ao fato do deslocamento do avião, entre imagens consecutivas, ser grande em relação às dimensões da cena coberta pelo campo visual da câmera. Em outras palavras, como as fotografias são batidas de pontos de vista bastante diferentes, existem diversas informações, tais como as laterais das edificações, que são vistas apenas em uma das imagens do par estéreo. O foco do projeto é justamente propor uma metodologia de estereoscopia para ser aplicada sobre áreas urbanas; os métodos propostos visam uma localização mais precisa das estruturas arquiteturais existentes nas imagens, preservando suas informações de bordas.

Para alcançar este objetivo, a idéia é utilizar informações específicas de fotos aéreas e unir métodos de processamento digital de imagens e visão computacional para atingir resultados mais adequados à reconstrução de modelos urbanos. Propõe-se então a utilização de resultados de segmentação da imagem como forma de direcionar os métodos de estereoscopia. Toda a pesquisa pode ser dividida em três etapas: o processo de segmentação, que foi feito através de uma variação do método de *watershed*, uma metodologia de estereoscopia para imagens com grandes áreas ocultas, e a união dos dois métodos, resultando em um modelo de estereoscopia direcionada pela segmentação por *watershed*.

Para se obter resultados mais precisos, são sugeridas alterações nos métodos genéricos de segmentação e estereoscopia existentes na literatura, que levam em consideração informações pertinentes aos objetos das cenas urbanas. Para o processo de segmentação, propõe-se uma metodologia baseada em informações geométricas e de textura dos objetos; na estereoscopia, utiliza-se restrições da base de imagens para simplificar etapas do processo, como o passo de retificação, e alteram-se os métodos de casamento denso para tratar as áreas de oclusões e alterações de luminosidade decorrentes da mudança no tamanho e quantidade de sombras existentes.

Compara-se o método de segmentação proposto com diversos modelos, existentes na literatura de processamento digital de imagens, através de análises métricas e visuais, reforçando a importância da utilização de informações de cor como limiar no processo. Os resultados de estereoscopia foram comparados com métodos propostos na literatura de visão computacional, demonstrando sua eficácia no tratamento de áreas de oclusões e bordas dos objetos.

## Sumário

---

<b>CAPÍTULO 1 – INTRODUÇÃO.....</b>	<b>10</b>
1.1 MOTIVAÇÃO.....	10
1.2 METODOLOGIA.....	10
1.3 CONTRIBUIÇÕES.....	11
1.4 ORGANIZAÇÃO DO TRABALHO.....	11
<b>CAPÍTULO 2 – MODELOS URBANOS 3D.....</b>	<b>12</b>
2.1 GEOPROCESSAMENTO.....	13
2.1.1 Aquisição de fotografias aéreas.....	14
2.1.2 Modelos altimétricos.....	16
2.2 SEGMENTAÇÃO.....	18
2.3 ESTEREOSCOPIA.....	19
2.3.1 Calibração.....	20
2.3.2 Retificação.....	21
2.3.3 Casamento Denso.....	23
2.3.4 Análise Interativa.....	24
<b>CAPÍTULO 3 – SEGMENTAÇÃO.....</b>	<b>25</b>
3.1 SEGMENTAÇÃO POR WATERSHED.....	25
3.2 WATERSHED COLORIDO.....	27
<b>CAPÍTULO 4 – ESTEREOSCOPIA.....</b>	<b>33</b>
4.1 METODOLOGIA.....	33
4.2 RETIFICAÇÃO.....	35
4.2.1 <i>Algoritmo de 8 Pontos</i> .....	36
4.2.2 <i>Singular Value Decomposition</i> .....	37
4.2.3 Matriz Essencial.....	38
4.2.4 Técnica Utilizada.....	39
4.3 CASAMENTO DENSO.....	42
4.3.1 Programação Dinâmica.....	43
4.3.2 Filtros Utilizados.....	44
4.4 UTILIZAÇÃO DA SEGMENTAÇÃO.....	47
<b>CAPÍTULO 5 – ANÁLISE E TESTES.....</b>	<b>51</b>
5.1 ANÁLISE DOS RESULTADOS.....	51
5.2 VISUALIZAÇÃO DOS RESULTADOS.....	61
<b>CAPÍTULO 6 – CONCLUSÃO.....</b>	<b>64</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>65</b>

## LISTA DE FIGURAS

---

	Pag
Figura 1 - Ortofoto gerada a partir de fotografias aéreas e modelo de terreno.....	12
Figura 2 - True-ortofoto gerada pela empresa Istar.....	13
Figura 3 – Modelagem manual das edificações.....	13
Figura 4 – Duas faixas de vôo sobre a região de Niteroi, Rio de Janeiro.....	14
Figura 5 – Áreas sobrepostas em um par de quadros.....	14
Figura 6 – Fotografias digitais com áreas de oclusão no detalhe.....	15
Figura 7 – Sensores aéreos da Leica Geosystems.....	16
Figura 8 – DTM original obtido por estereoscopia.....	17
Figura 9 – Par estereo retificado e a disparidade gerada.....	18
Figura 10 – Fotografia aérea e seu respectivo mapa de intensidade de bordas.....	19
Figura 11 – Esboço de uma utilização de fotografias métricas geo-referenciadas...	20
Figura 12 – Três diferentes visões do grid de calibração.....	21
Figura 13 – Par estereo e seus respectivos epipólos.....	22
Figura 14 – Linhas epipolares.....	22
Figura 15 – Par estereo retificado.....	22
Figura 16 – Par estereo e respectivo mapa de disparidade.....	23
Figura 17 – Analisador de Estereo manual.....	24
Figura 18 – Segmentação por watershed.....	26
Figura 19 – Segmentação por watershed utilizando CBMA.....	26
Figura 20 – Segmentação por watershed de áreas de suburbio.....	27
Figura 21 – Segmentação por watershed de centro urbano.....	29
Figura 22 – Segmentação por watershed de centro urbano.....	30
Figura 23 – Segmentação por watershed de centro urbano.....	31
Figura 24 – Segmentação por watershed de centro urbano.....	32
Figura 25 – Fotografias digitais obtidas com a DX3700.....	34
Figura 26 – Segmentação e imagem de disparidade.....	35
Figura 27 – Geometria Epipolar.....	38
Figura 28 – Retificação de par estereo.....	39
Figura 29 – Retificação de par estereo.....	40
Figura 30 – Varredura de linhas epipolares.....	40
Figura 31 – Sobreposição de cortes em um par retificado.....	41
Figura 32 – Caminho Mínimo sobre o mapa de disparidade.....	42
Figura 33 – Programação dinâmica sobre o mapa de disparidade.....	43
Figura 34 – Filtros e imagem de disparidade.....	44
Figura 35 – Filtros e imagem de disparidade.....	45
Figura 36 – Disparidades e mapas referentes a linha central das imagens.....	46
Figura 37 – Calibração da disparidade.....	48
Figura 38 – Disparidades com diferentes saltos no caminho mínimo.....	49
Figura 39 – Programação dinâmica sobre o mapa de disparidade.....	50
Figura 40 – Bordas segmentadas.....	50
Figura 41 – Intersecção de objetos.....	51
Figura 42 – Comparação entre métodos de segmentação por <i>watershed</i> .....	52

Figura 43 – Comparação entre métodos de segmentação por <i>watershed</i> .....	53
Figura 44 – Comparação entre métodos de segmentação por <i>watershed</i> .....	54
Figura 45 – Comparação entre métodos de segmentação por <i>watershed</i> .....	55
Figura 46 – Comparação entre métodos de segmentação por <i>watershed</i> .....	56
Figura 47 – Comparação entre métodos de segmentação por <i>watershed</i> .....	57
Figura 48 – Resultado métrico dos algoritmos de <i>watershed</i> .....	58
Figura 49 – Resultados de estereoscopia.....	59
Figura 50 – Resultados de estereoscopia.....	60
Figura 51 – Óculos 3D, <i>chromadepth</i> e <i>anaglyphy</i> .....	61
Figura 52 – Visualização <i>anaglyphy</i> .....	62
Figura 53 – Visualização <i>chromadepth</i> .....	62
Figura 54 – Visualização <i>anaglyphy</i> .....	63
Figura 55 – Visualização <i>chromadepth</i> .....	63

# CAPÍTULO 1

## INTRODUÇÃO

Nesta dissertação considera-se o problema de recuperar a estrutura tridimensional de uma área metropolitana a partir de um conjunto de fotografias aéreas de baixa altitude, onde os efeitos de perspectiva e oclusão são significativos. Para tanto, utiliza-se como base imagens obtidas com uso de equipamentos populares e acessíveis. Grande parte das imagens utilizadas nesta tese são fotografias de regiões centrais da cidade de Belo Horizonte, MG. Estas fotografias contêm um elevado número de edificações, que constituem um problema à parte em estereoscopia devido as áreas de oclusões geradas por elas.

### 1.1 - MOTIVAÇÃO

A larga utilização de Sistemas de Informações Geográficas (SIG) tem refletido em extensas pesquisas relacionadas com reconstrução de estruturas arquiteturais à partir de imagens. Esse assunto tem sido bastante estudado na áreas de Processamento de Imagens, Visão Computacional e Computação Gráfica. No entanto, a maior parte da literatura sobre SIG trabalha com imagens capturadas a grandes altitudes, reduzindo ou eliminando as dificuldades causadas pela perspectiva e pela oclusão, mas exigindo o uso de equipamentos caros e de alta-resolução, como câmeras métricas de médio formato.

A comunidade de Visão Computacional, por outro lado, tem dedicado seu esforço ao desenvolvimento de algoritmos que funcionam de maneira eficaz sobre qualquer tipo de base de imagens [14,15,1,37,29,20,24]. Diversas técnicas têm mostrado resultados eficazes inclusive na presença de grandes oclusões [7,26,33,25], porém o custo computacional é elevado e essas técnicas falham em não utilizar nenhum conhecimento específico sobre cenas urbanas para obter resultados mais precisos.

A utilização de conhecimentos referentes ao problema de reconstrução de modelos urbanos tem demonstrado uma enorme melhoria nos resultados apresentados [23,40]. Porém, estas técnicas, oriundas da comunidade de Computação Gráfica, apresentam um alto nível de intervenção manual, tornando o processamento sobre grandes bases de dados extremamente difícil e caro.

O objetivo deste trabalho é o de combinar os esforços das várias áreas envolvidas no problema de reconstrução para definir uma metodologia específica para a reconstrução de grandes áreas urbanas a partir de fotografias de baixa altitude obtidas com câmeras de baixa-resolução. A estratégia adotada é a de se utilizar conhecimentos válidos sobre as cenas urbanas, para guiar o processo de reconstrução.

### 1.2 - METODOLOGIA

A idéia principal da metodologia proposta consiste em utilizar técnicas de Processamento Digital de Imagens (PDI) para gerar informações que guiarão o processo de reconstrução geométrica referente à base de imagens utilizadas.

O processo pode ser desmembrado em três etapas, a extração de características da base, a reconstrução utilizando estereoscopia, e uma pós-correção do modelo através do conhecimento de suas características básicas. O resultado obtido será analisado numericamente e visualmente, através de técnicas de Visão Computacional e PDI.

### **1.3 - CONTRIBUIÇÕES**

Em diferentes etapas da pesquisa desenvolvida, podem ser localizadas contribuições para a literatura. Para o processo de extração de características foi desenvolvida uma variação do método de segmentação por *watershed* [16] baseada na utilização das cores da imagem como limiares do processo. Esse incremento resultou em uma melhoria significativa no processo de segmentação, devido à enorme variação de cores em áreas urbanas, e a frequente uniformidade dessas cores no interior de estruturas individuais, como ruas, casas, lajes, árvores, sombras e laterais de edifícios.

No processo de estereoscopia, pode ser citada como contribuição uma alternativa simples e prática de retificação, a qual pode ser aplicada especificamente em pares de fotografias obtidas em orientações quase paralelas, como é o caso de fotografias aéreas. Também, pode-se citar uma variação do método de geração de disparidades, utilizando um conjunto de filtros propostos exclusivamente para o tipo de cena a ser tratado.

Por fim, o último passo da metodologia inova ao inserir a segmentação de imagens no processo de estereoscopia, visando corrigir distorções indevidas do mesmo [12,13,15,16]. A metodologia proposta neste passo final é descrita em um artigo publicado nos anais do ICIAP 2003, sob o título de *Towards Automatic 3D Reconstruction of Urban Scenes from Low-Altitude Aerial Images* [4].

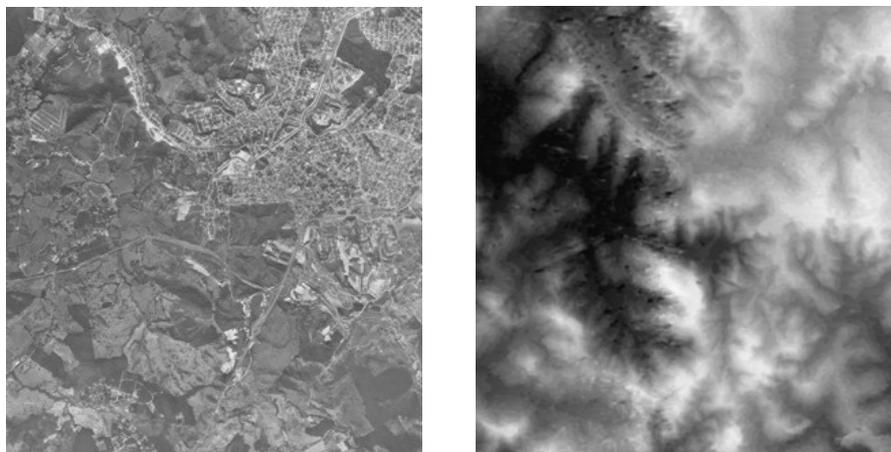
### **1.4 – ORGANIZAÇÃO DO TRABALHO**

No Capítulo 2, apresenta-se técnicas e métodos utilizados para a geração de modelos urbanos 3D, envolvendo Geo-Processamento e Visão Computacional. O Capítulo 3 descreve o método de *Watershed*, que será utilizado para calibração do processo de estereoscopia, descrito no Capítulo 4. O Capítulo 5 contém testes feitos utilizando a metodologia proposta por esta dissertação, e comparações com outros métodos existente no mercado.

## CAPÍTULO 2

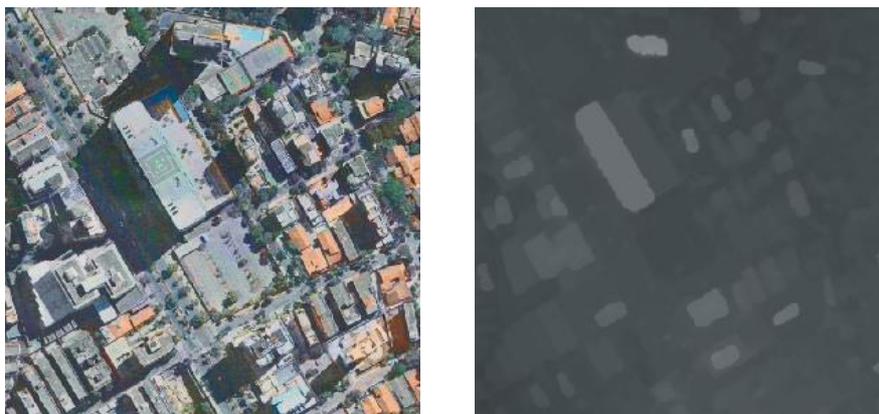
### MODELOS URBANOS 3D

A visualização tridimensional de áreas urbanas tem sido alvo de pesquisas e desenvolvimentos na área de computação gráfica, sendo largamente utilizada para os mais variados fins, como realidade virtual, jogos, simuladores, efeitos especiais, em projetos arquiteturais e para planejamento urbano. Em Sistemas de Informações Geográficas (SIGs) a visualização 3D normalmente é feita a partir de imagens georreferenciadas e suas respectivas informações de altitude, como se pode visualizar no exemplo fornecido na Figura 1.

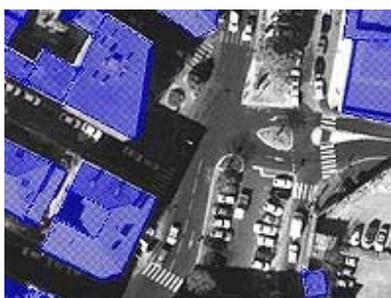


**Figura 1** - Ortofoto e seu respectivo modelo de terreno, material gerado a partir de fotografias aéreas, do acervo da CEMIG, batidas sobre a cidade de Betim, MG.

A informação de altitude da Figura 1 apresenta uma resolução de 3 m/*pixel*, sendo incapaz de possuir informações válidas para identificar construções urbanas. Um modelo de maior resolução, como o mostrado na Figura 2, contém informações suficientes para ser considerado como um modelo urbano, com sua informação de altitude tendo uma resolução de 50 cm/*pixel*. Esta forma de visualização urbana não necessita de intervenção manual, como é o caso da elaboração de um ambiente virtual através de ferramentas de análise de estéreo (Figura 3), tem seu foco de uso justamente em grandes bases de dados, nas quais se torna inviável uma montagem manual dos objetos arquiteturais das cenas.



**Figura 2** - *True-orthophoto* gerada pela empresa Istar [41], utilizando fotos aéreas em tons de cinza e colorida artificialmente pela Infostrata [42], e seu respectivo modelo de superfície.



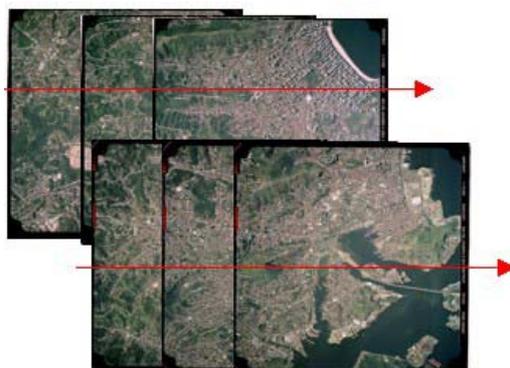
**Figura 3** – Modelagem manual das edificações executada pela Empresa Istar [41].

## 2.1 – GEOPROCESSAMENTO

A visualização tridimensional de áreas urbanas tem sido largamente utilizada para os mais variados fins, como realidade virtual, jogos, simuladores, efeitos especiais, em projetos arquiteturais e para planejamento urbano. Em Sistemas de Informações Geográficas, o levantamento e a geração dos dados necessários são feitos normalmente para áreas significativas, sendo grande parte do processamento feito de forma automática. Nesta seção, apresenta-se como as fotografias são normalmente adquiridas no processo, e duas classes de informações de altimetria.

### 2.1.1 – AQUISIÇÃO DE FOTOGRAFIAS AÉREAS

O voo sobre as áreas a serem mapeadas é feito através de passadas paralelas, ou faixas, de forma que ocorra pelo menos 60% de sobreposição entre as imagens obtidas em uma mesma faixa e 30% de sobreposição lateral, entre faixas consecutivas, como pode ser visualizado na Figura 4. A sobreposição horizontal é a utilizada no processo de marcação de altimetria, enquanto que a sobreposição lateral representa uma folga que será utilizada na composição dos resultados das diversas faixas.



**Figura 4** – Duas faixas de voo sobre a região de Niterói, Rio de Janeiro, imagens cedidas pela Infostrata [42].

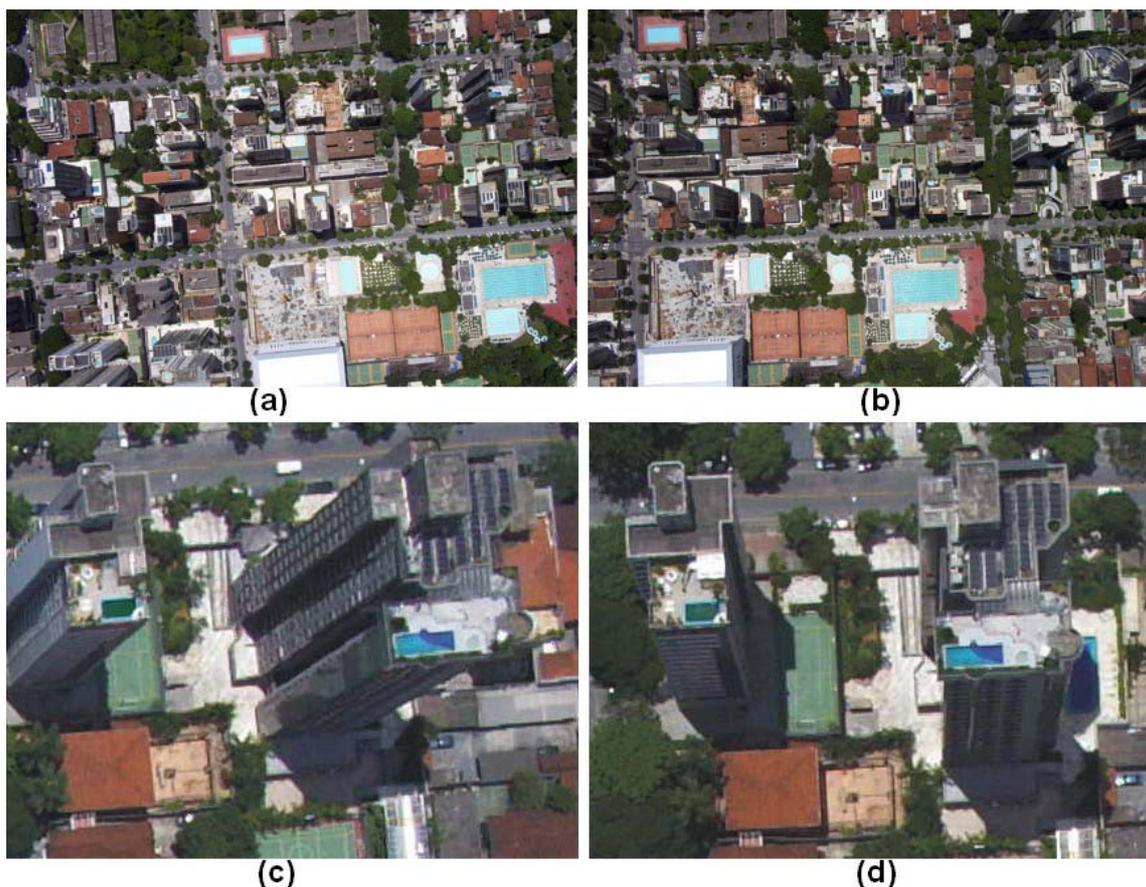
O exemplo da Figura 4 mostra seis imagens obtidas por câmeras métricas de médio formato. Essas câmeras possuem características facilitadoras para o geo-processamento, porém são extremamente caras. Cada quadro da faixa de voo é uma imagem escaneada de um diapositivo, que é a imagem revelada do negativo da fotografia. O diapositivo tem exatamente o mesmo tamanho do negativo, normalmente algo entre 23 x 23cm, e estas imagens podem ser escaneadas a até 10.000 x 10.000 *pixels*, ou seja, uma imagem sobre uma área de 4 x 4 km terá uma resolução de 0.4 m/*pixel*.



**Figura 5** – Áreas sobrepostas em um par de quadros.

Estas imagens possuem uma leve oclusão e mudança de perspectiva entre as mesmas nos centros urbanos, como pode ser percebido na Figura 5. Visando justamente utilizar exemplos complexos em termos de oclusão e perspectiva, e desenvolver um método robusto o suficiente para funcionar em imagens obtidas por qualquer câmera, foi utilizada nesta dissertação uma câmera de uso comum, a DX3700 da Kodak. Em oposição às câmeras métricas de médio formato, as câmeras comuns apresentam grandes distorções radiais, não são calibradas, e adquirem muito menos informações. Porém são muito mais acessíveis, sendo mais de 40 vezes mais baratas que as de médio formato.

A DX3700 é uma câmera digital com precisão de 3 *Megapixels*; que é quase 30 vezes menor que a fornecida pelas imagens da Figura 4. Portanto, o voo foi feito a uma altitude menor que 500 metros em relação ao solo. As imagens obtidas possuem oclusões extremas e grandes mudanças de perspectivas. Como pode ser percebido pela Figura 6, elas também são de uma resolução espacial elevada, com  $10\text{cm}/\text{pixel}$ , sendo portanto perfeitas para os testes executados neste projeto.



**Figura 6** – Fotografias digitais obtidas com a DX3700 sobre o Minas Tenis Clube I, BH, MG. (a,b) Par estéreo, (c,d) áreas de oclusão em detalhe.

## 2.1.2 – MODELOS ALTIMÉTRICOS

Ainda hoje, o processo de geração de altimetria é frequentemente feito de forma analógica, através da geração do modelo de curvas de nível. Entretanto, existem inúmeros softwares que permitem a geração automática, como as ferramentas ENRI [43], ERDAS [44] e outros componentes da LEICA Geosystems [45]. Esses softwares utilizam tratamento estereoscópico para a geração da informação altimétrica, como também é o caso dos produtos vendidos pela empresa ISTAR [41]. Atualmente a montagem dessas informações pode ser feita independentemente das fotografias, através de sensores laser, como sistemas LIDAR de scanearmento aéreo a laser fornecidos pela LEICA [45] (Figura 7). Os sistemas a laser possuem alta precisão e não necessitam de processamento intenso para a geração dos dados, porém têm como pontos fracos a necessidade de manutenção de um avião especificamente adaptado para o material e o preço, que atinge patamares superiores a algumas centenas de milhares de dólares.

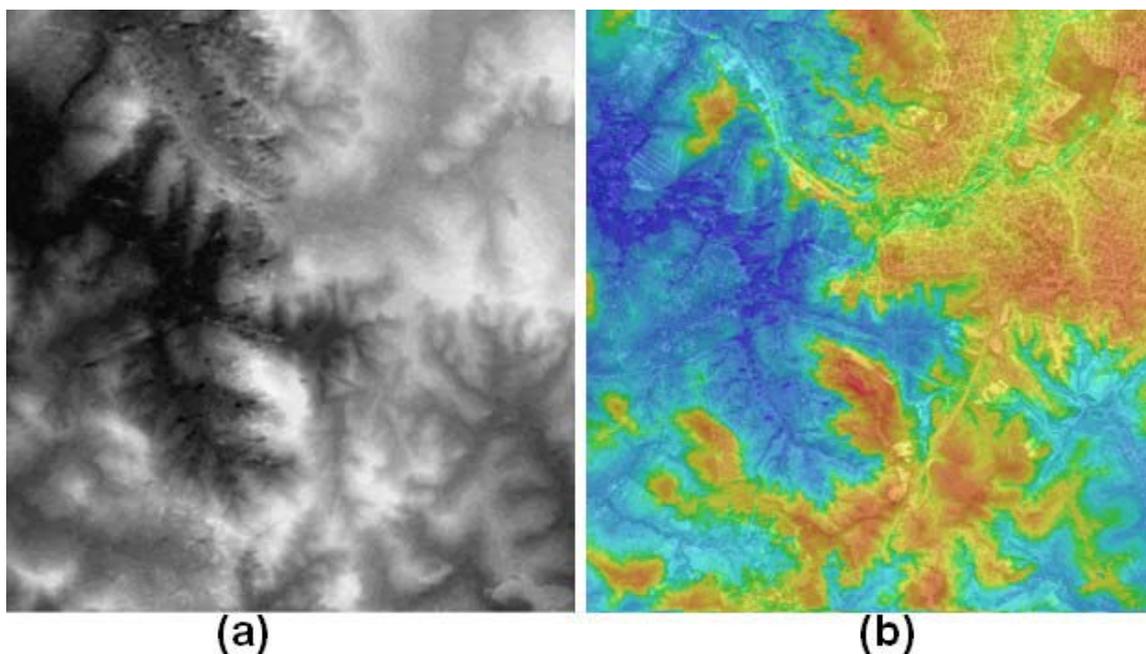


**Figura 7** – Sensores aéreos da Leica Geosystems, (a) sensor digital ADS40, (b) *scanner laser ALS50*.

De forma geral, os modelos de altitude gerados automaticamente podem ser de dois tipos distintos, modelos de terreno (*digital terrain model* DTM ou modelo digital de terreno MDT ou *digital elevation model* DEM) e modelos de superfície (DSM, *digital surface model*). A diferença básica entre os dois tipos de modelos está no fato de que o modelo de terreno representa apenas informações da geografia do local; já o modelo de superfície parte justamente para o caminho oposto, tentando fazer uma representação fiel da altitude de todos os objetos presentes na cena, como árvores e edificações. Normalmente os DSM são feitos em resolução bem maior, para permitir a visualização de informações não inerentes ao relevo. No caso da utilização de estereoscopia para a geração de DSM, é necessário o tratamento de áreas de oclusão criadas pelas estruturas arquiteturais, sendo esta a motivação desta dissertação. Já as DTM podem ser feitas inclusive sobre resoluções bem menores, pois a única informação relevante será a do relevo. Enquanto um arquivo DSM deverá apresentar uma resolução inferior a 1 *m/pixel*, um arquivo DTM com 5 *m/pixel* é considerado de excelente resolução.

## MODELOS DE TERRENO

Os modelos de terreno são os modelos de aquisição mais simples e mais utilizados em geoprocessamento. Sua obtenção pode ser feita inclusive através de modelos de curvas de nível gerados com marcação de pontos geodésicos em trabalho de campo, sem a necessidade de um voo. Em sistemas de informações geográficas, estes modelos são utilizados inclusive na ortoretificação das fotografias aéreas. Nesse processo os quadros obtidos no voo são distorcidos conforme o DTM referente aos mesmos; os resultados desse processo são chamados de ortofotos e podem ser facilmente compostas em uma única imagem. Na Figura 8, tem-se um exemplo de DTM e sua visualização junto com uma ortofoto utilizando a técnica de *chromadepth* (descrita na Seção 5.2, Visualização dos Resultados).

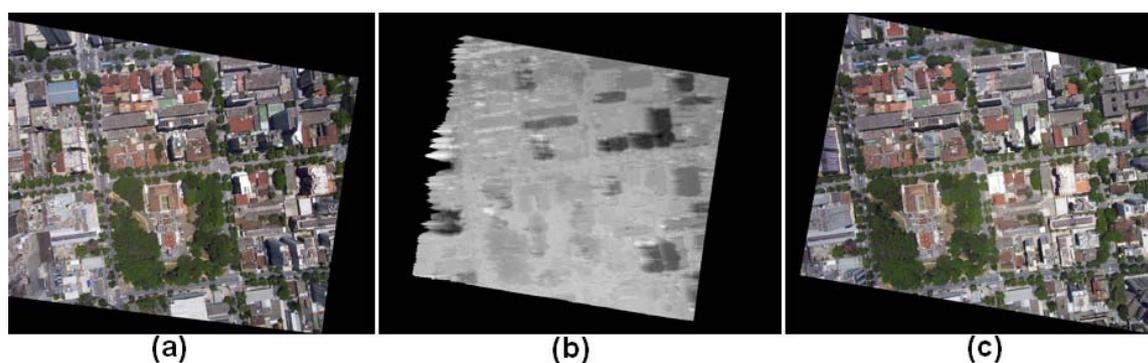


**Figura 8** – (a) DTM original obtido por estereoscopia, (b) sobreposição da DTM com cores da representação por *chromadepth* e da respectiva ortofoto.

## MODELOS DE SUPERFÍCIE

Os modelos de superfície apresentam uma resolução significativamente maior que os DTM, como pode ser observado no DSM gerado pela ISTAR [41] na Figura 2. Eles podem ser originados através de instrumentos a laser, como o empregado pela Orbisat [46]

e pela Leica [45] (Figura 7-b), e também podem ser obtidos por estereoscopia. Este último método é o objeto de estudo desta dissertação. Na Figura 9, pode-se visualizar um par estéreo retificado de imagens de alta-resolução e sua respectiva disparidade (seção 2.3.3), a qual poderá ser utilizada na montagem da DSM. Um exemplo de DSM é o resultado produzido pela Istar [41] (Figura 2).



**Figura 9** – (a) e (c) par estereo retificado, (b) disparidade gerada entre o par.

## 2.2 – SEGMENTAÇÃO

O objetivo do uso da segmentação no processo proposto é justamente o de se utilizar informações específicas de cenas urbanas para aprimorar o processo de reconstrução. Fotografias de cenas urbanas possuem uma enorme variedade de características visuais, constituindo em uma fonte de dados excelente para o processo de segmentação. Nessa etapa, lida-se com o problema específico de segmentar objetos arquiteturais e artefatos presentes no meio urbano, como ruas, árvores, sombras, laterais de edificações, lajes e telhados.

A segmentação de imagens é um problema de decomposição semântica, exigindo portanto soluções específicas para determinados problemas que requerem uma intensa intervenção humana. A escolha da técnica apropriada para determinada aplicação deve levar em conta a natureza da base de dados a ser utilizada e os resultados esperados do algoritmo, sendo os resultados analisados em função de sua forma, contornos, regiões e texturas dos objetos localizados. No caso específico deste trabalho tenta-se localizar elementos que possuem bordas com uma definição razoável é com uma certa homogeneidade de textura.

Dentre os métodos tradicionais de segmentação pode-se citar: *region growing and merging* [2,28], *deformable models* [19], *bubbles* [12] e *watershed segmentation* [13]. Os dois primeiros métodos possuem problemas de inicialização e convergência, necessitando de certa intervenção manual para um resultado mais acertado. A técnica de *bubbles* possui uma inicialização automática, porém depende de ajuste fino de um conjunto de parâmetros para atingir uma convergência completa, além de ser bastante lento quando comparado com

a técnica que iremos utilizar, que é uma variação de *watershed segmentation* [13], que será chamada de *watershed colorido*.

O método de *watershed* faz parte de uma classe de técnicas que tratam a imagem como uma superfície 3D, onde as elevações são correlacionadas com valores de intensidade do *pixel*, como pode-se visualizar na Figura 10. Como o método leva em consideração apenas informações obtidas a partir da borda do mapa de intensidade (Figura 10), inova-se ao adicionar nesta técnica uma informação referente à textura dos objetos. No método de *watershed colorido*, a informação de cor é predominante no processo, sendo a responsável por todo o funcionamento do método, como será melhor descrito na Seção 3.2 .



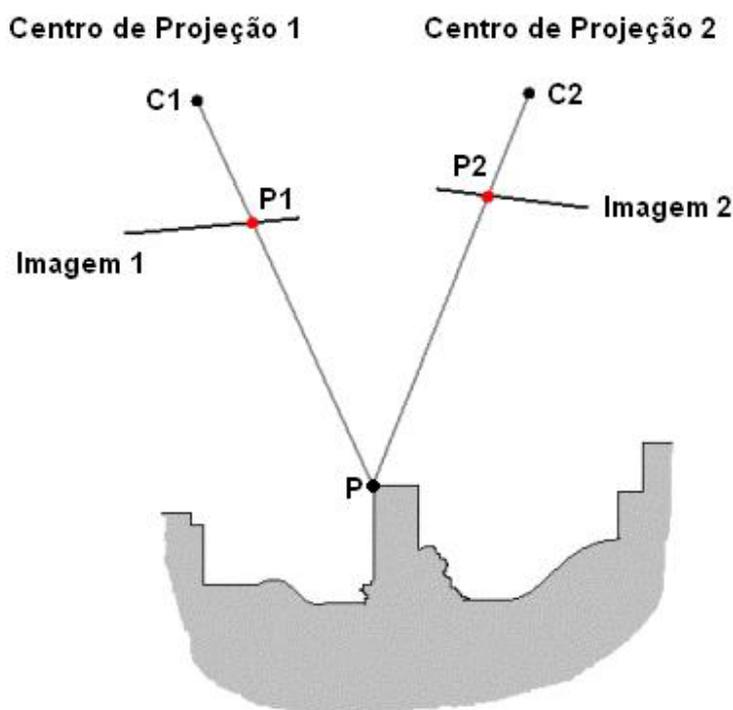
**Figura 10** – Fotografia aérea e seu respectivo mapa de intensidade de bordas.

### 2.3 – ESTEREOSCOPIA

Diversas técnicas de reconstrução 3D de uso geral elaboradas pela comunidade de Visão Computacional conseguem lidar com oclusões. Em particular, uma grande evolução na área foi o desenvolvimento de técnicas de *Space Carving* (escultura espacial) [14,15,1,37,29,20,24], as quais obtêm resultados globalmente consistentes mesmo na presença de grandes oclusões. Reconstruções através de técnicas de *dense stereo* (estereoscopia densa) também toleram a existência de descontinuidades [7,26,33,25]. O problema destas técnicas reside no fato de serem extremamente gerais, além de possuírem um alto custo computacional. Aqui o objetivo é a utilização do conhecimento específico relacionado com cenas urbanas para obter reconstruções mais precisas, optando-se por um método de estereoscopia simples e com baixo custo computacional, para poder ser empregado sobre grandes áreas, corrigido através do resultado de segmentação. Nesta seção, descrevem-se e exemplificam-se os passos referentes ao processo de estereoscopia utilizados no processo, passando pela calibração da câmera, retificação dos pares estéreo e marcação de disparidade através de correlação entre cada par de linhas das imagens retificadas. Finalmente, é demonstrado como funciona o processo de estereoscopia manual.

### 2.3.1 – CALIBRAÇÃO

As imagens utilizadas em geoprocessamento devem ser obtidas através de câmeras métricas, ou seja, que tenham seus parâmetros, como distância focal e centro do foco previamente calibradas, sendo as mesmas corrigidas com relação a sua distorção radial. A razão da calibração é tornar as operações de geoprocessamento mais simples, facilitando o posicionamento da imagem no espaço. Pode-se perceber pela Figura 11 que em um processo de geração de altimetria as imagens são tratadas de forma simplificada. Outro motivo da calibração é facilitar os demais passos da estereoscopia, que são a retificação e a determinação da disparidade. Os resultados obtidos e os métodos propostos para imagens calibradas retornam um resultado bem mais apurado do que sobre as imagens originais sem nenhum conhecimento prévio dos parâmetros da câmera.

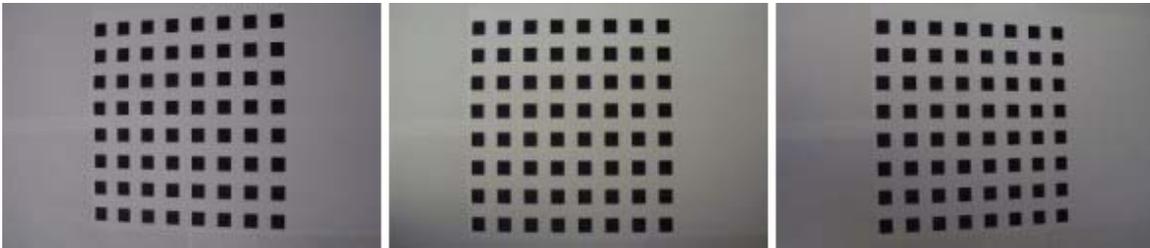


**Figura 11** – Esboço de uma utilização de fotografias métricas georreferenciadas.

A escolha do método de calibração se baseou na precisão dos resultados e na facilidade de obtenção dos mesmos. Os métodos existentes podem ser divididos em duas classes distintas, *Photogrammetric calibration* [21,30] e *Self-calibration* [36,27,35]. Os métodos baseados em auto-calibração são extremamente flexíveis, porém possuem um resultado com pouca precisão, devido ao grande número de parâmetros a serem estimados [35]. A calibração fotogramétrica pode ser executada através da observação de um objeto cuja geometria 3D é conhecida com bastante precisão [21]. Nesse caso, a calibração é extremamente eficiente. A calibração utilizando um plano precisamente transladado entre

as diferentes imagens [30] também retorna excelentes resultados, porém todos estes métodos exigem um custo elevado de tempo e equipamento para se conseguir a precisão necessária.

Zhengyou Zhang [49] propôs um método de calibração híbrido, envolvendo idéias de ambos os tipos, extremamente flexível. Seu método se baseia no conhecimento *a priori* de uma série de pontos em um plano no espaço e a localização dos mesmos nas imagens. Erros no processo de marcação automática dos pontos nas imagens são bem aceitos pelo algoritmo, sendo este o motivo de sua utilização nesta dissertação. Para calibrar a câmera, simplesmente desenha-se um *grid* em uma ferramenta de edição de imagens, e coloca-se sobre uma mesa branca; a seguir bate-se 3 fotos em ângulos diferentes (Figura 12), que é a quantidade mínima de visões necessárias para se efetuar a calibração. O algoritmo fornece os parâmetros internos da câmera e de distorção radial, os quais serão utilizados para corrigir as imagens, através do método proposto em [49].



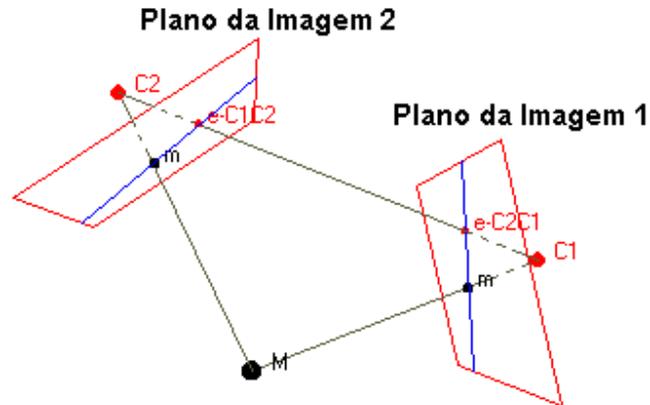
**Figura 12** – Três diferentes vistas do *grid* de calibração.

### 2.3.2 – RETIFICAÇÃO

A etapa de retificação é outro passo importante no processo de localização dos *pixels* correspondentes em um par estéreo. Algoritmos que procuram a correspondência de forma direta entre os *pixels* de um par de imagens possuem um custo computacional elevado e resultados inferiores às técnicas executadas sobre pares retificados. Diversos métodos de retificação foram propostos, dentre os quais pode-se citar Cox e Roy [34] e Liebowitz e Zisserman [8]; entretanto, a maioria destes métodos são de difícil implementação, sendo a solução que propomos simples, rápida e bastante eficaz para imagens aéreas. A idéia básica do processo de retificação consiste em diminuir o espaço de busca da correspondência da imagem inteira para apenas uma linha da mesma. Pela Figura 13, pode-se visualizar a linha localizada em azul, chamada de linha epipolar, onde  $eC1C2$ ,  $eC2C1$  são chamados de epipolos. Um epipolo é o ponto em uma imagem onde o centro de seu par seria visto. Cada linha epipolar de uma imagem tem uma linha correspondente na outra imagem. Para se calcular a correlação entre as linhas, é necessário o conhecimento da matriz essencial entre as duas imagens, a qual será melhor descrita na Seção 4.2 .

Dado que a matriz essencial é conhecida, e que pode-se calcular as linhas epipolares, a etapa de retificação pode ser finalizada, distorcendo-se a imagem de forma a que suas linhas epipolares fiquem paralela. Na Figura 14, pode-se observar uma série de linhas epipolares e na Figura 15, o par retificado com as linhas paralelas umas às outras. A técnica utilizada nesta dissertação, Seção 4.2.4, se baseou em propriedades específicas de

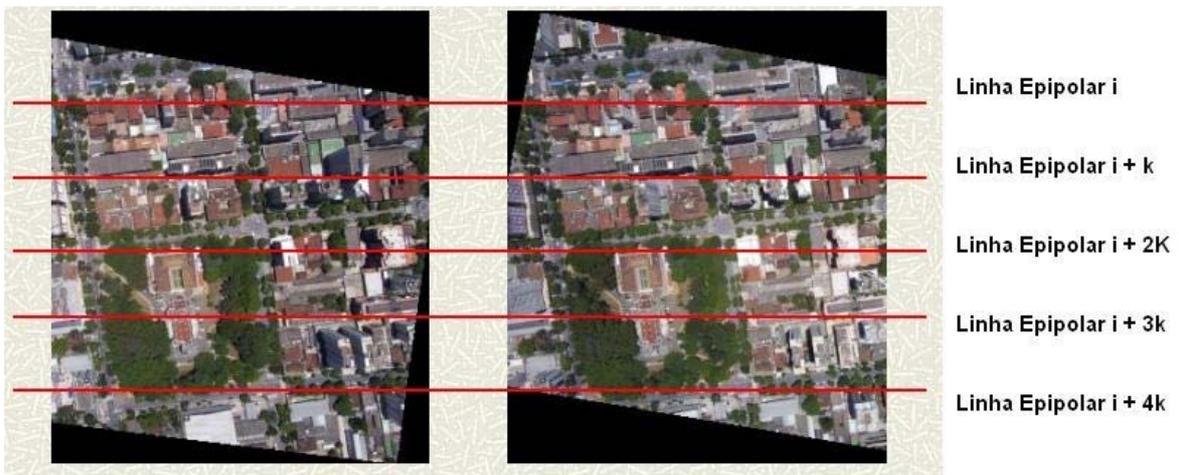
fotografias aéreas para retificar os pares, resultando em um algoritmo eficaz, rápido e simples, evitando a necessidade de cálculos complexos, como nos demais algoritmos propostos na literatura.



**Figura 13** – Par estéreo e seus respectivos epíolos.



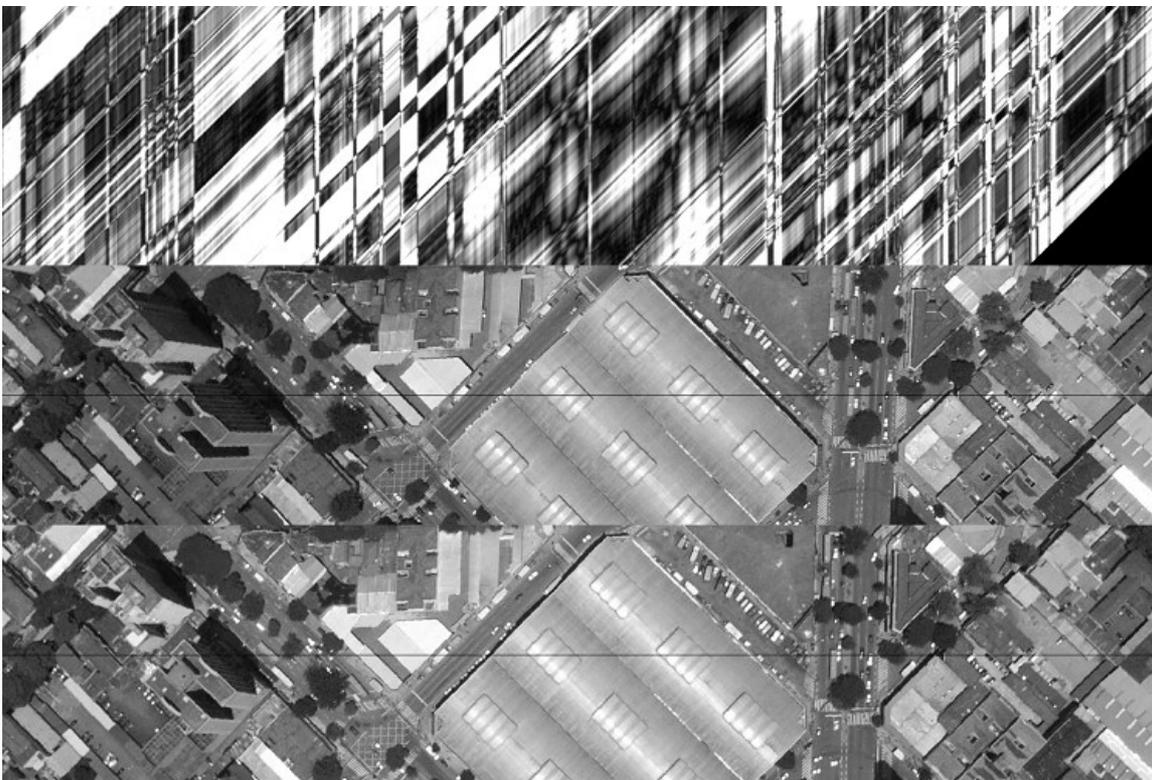
**Figura 14** – Linhas epipolares em um par estéreo.



**Figura 15** – Par estéreo retificado.

### 2.3.3 – CASAMENTO DENSO

O mapa de disparidade identifica o mapeamento espacial entre os *pixels* de um par de imagens retificadas. É importante observar na Figura 15 que esse mapeamento pode ser realizado independentemente para cada par de linhas retificadas, ou seja, a disparidade é apenas o deslocamento horizontal relativo entre cada par de *pixels*, como se pode observar na Figura 16. O problema central da estereoscopia é justamente descobrir qual o deslocamento válido para cada *pixel*, adotando-se nesta dissertação como base um Mapa de Disparidade, onde esse mapa consiste em uma série de testes com diferentes valores de deslocamento armazenados em um único *grid* (Figura 16).



**Figura 16** – Par estéreo comparado sob o ponto de vista de um par de linhas retificadas, com o mapa de disparidade referente ao par de linhas epipolares demarcada em preto sobre as imagens.

Cada linha vertical do mapa de correlação é referente ao elemento de mesma posição horizontal na imagem retificada. Cada elemento dessa linha vertical representa um resultado para um determinado valor de deslocamento, onde esse resultado é apenas um filtro de média aplicado entre os valores do *pixel* de uma imagem e o *pixel* referente ao deslocamento proposto na outra imagem. O segredo do processo de disparidade está justamente nos filtros a serem utilizados; o filtro que será proposto (Seção 4.3.2) também se

baseou em informações específicas referentes a fotografias aéreas. O resultado esperado em uma análise sobre o mapa é o caminho mínimo que o percorre horizontalmente, esse caminho é facilmente identificado, através de uma análise visual da Figura 16, como sendo a faixa escura horizontal presente no centro do *grid*; a forma como este caminho será localizado está descrita na Seção 4.3.1.

### 2.3.4 – ANÁLISE INTERATIVA

Em oposição ao que é feito nesta dissertação, o processo de marcação de disparidade pode ser feito de forma manual. Na Figura 17, pode-se observar um exemplo de um Analisador de Estéreo. Este processo permite uma marcação precisa e uma definição perfeita das estruturas presentes na cena; porém o custo da intervenção humana é enorme, tornando difícil e penoso o tratamento de grandes regiões.



**Figura 17** – Analisador de Estéreo manual desenvolvido pela HGT [47].

O processo de análise manual pode também ocorrer como pós-processamento a um reconhecimento automático de altimetria, visando uma melhoria em áreas de interesse específico, como prédios, viadutos, entre outros artefatos.

# CAPÍTULO 3

## SEGMENTAÇÃO

O processo de segmentação consiste em localizar diferentes objetos na imagem, normalmente delimitados por bordas, mudança de intensidade ou coloração. A técnica de *watershed* que estaremos utilizando baseia-se nas bordas existentes na imagem. Ainda neste capítulo propomos a utilização da variação de cor para melhorar ainda mais o resultado da segmentação.

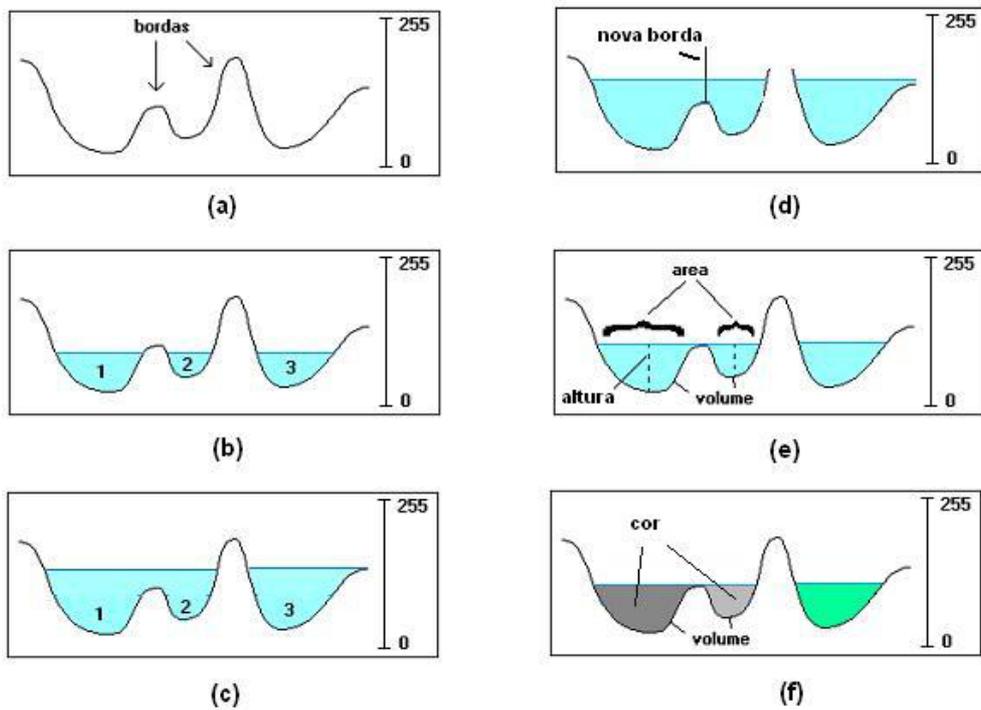
### 3.1 – SEGMENTAÇÃO POR WATERSHED

A técnica de segmentação de imagens através do método *watershed* [9,10,16] explora analogias que podem ser estabelecidas entre a imersão de um relevo topográfico a partir de seus mínimos e a simulação de imersão sobre uma imagem digital, vista como um relêvo artificial, onde a cada *pixel* da imagem se atribui uma altitude proporcional à sua intensidade. No método *watershed*, a segmentação da imagem resulta de um processo de simulação de imersão, por etapas, realizado a partir de pontos preferenciais do relêvo (mínimos locais, marcadores impostos através de sequências de filtragens ou marcadores colocados manualmente). Assim, o processo de crescimento é interrompido e os *pixels* correspondentes são marcados como pertencentes à fronteira.

Na maioria dos casos, a simulação de imersão é feita sobre a imagem gradiente obtida a partir da imagem original. A imagem gradiente é empregada com a finalidade de se ressaltar os contornos das regiões de interesse (Figuras 10 e 18-a). Entretanto, por causa do ruído, estas imagens apresentam um elevado número de mínimos locais que produzem, após a imersão total, um número igualmente excessivo de regiões. Este problema é conhecido como *super-segmentação*.

Para reduzir o problema da *super-segmentação* há duas classes de soluções: na primeira, modifica-se a imagem gradiente de forma que suas bacias correspondam às regiões de interesse, enquanto na segunda, procura-se remover bordas irrelevantes. A imposição de marcadores, que é uma exigência da primeira classe, força uma modificação da homotopia (alteração na “topologia do relêvo” e conseqüentemente dos pontos onde a imersão é iniciada). Por sua vez, a utilização de mecanismos de agregação das regiões não significativas, segundo um critério de similaridade, que exemplifica a segunda classe, procura remover as bordas irrelevantes. Dependendo da implementação, o mecanismo de agregação pode estar embutido no processo de crescimento de regiões ou pode ser realizado *a posteriori*. Diversas variantes da Segunda classe de soluções foram propostas na literatura. Entre elas está o método interativo de segmentação ABA [17] que utiliza o algoritmo de agregação de bacias CBMA – *Catchment Basis Merging Algorithm* [18]. O CBMA, na sua versão original, procura suprimir as regiões não significativas da imagem (Figura 19), no decorrer do processo de crescimento e emprega, para isso, um ou mais parâmetros *geométricos* (área, profundidade, volume da bacia de capacitação), (Figura 18e) para decidir em agregar ou não bacias vizinhas. A recente inclusão [3] no CBMA de um parâmetro com base na similaridade entre as cores das regiões adjacentes, permite que a segmentação também possa ser realizada combinando-se independentemente parâmetros geométricos e de cor. Neste último caso, no entanto, a agregação é realizada como pós-

processamento, ou seja, *a posteriori*. Nesta dissertação, é apresentada uma solução alternativa para o problema da super-segmentação, *Colored Watershed*, descrita na Seção 3.2, que pertence à Segunda classe e realiza a agregação das bacias de captação, durante o processo de imersão com base em um critério de similaridade entre as médias das cores de duas regiões adjacentes.



**Figura 18** – Segmentação por watershed. (a) Imagem de bordas, (b) início do preenchimento dos lagos, (c) estágio avançado de preenchimento com encontro de lagos, (d) nova borda formada para efetuar a separação dos objetos, (e) atributos geométricos dos objetos, (f) critério de similaridade por cor ou *Watershed colorido*.

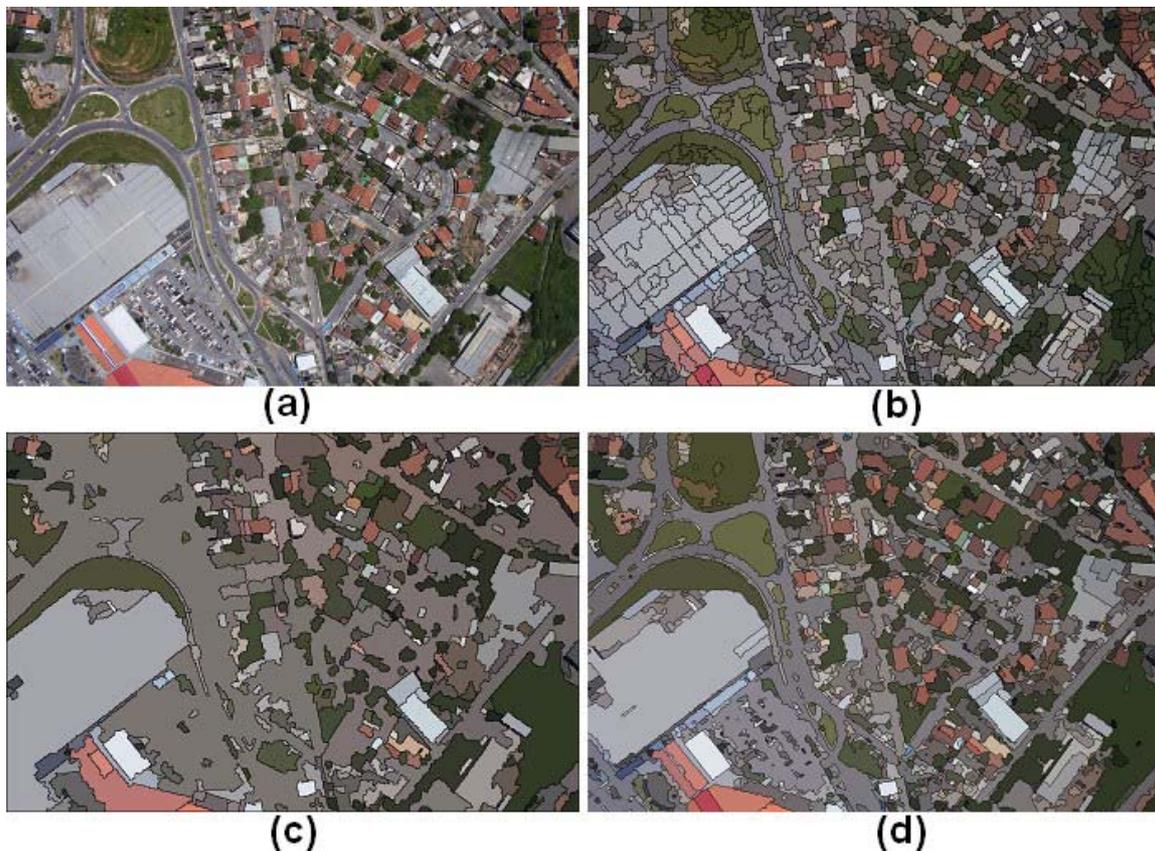


**Figura 19** – Segmentação por *watershed* utilizando CBMA com atributos geométricos. (a) Imagem original, (b) imagem segmentada.

### 3.2 – WATERSHED COLORIDO

Dentre as variações de *watershed*, a utilização de atributos em uma implementação (CBMA) apresentou um resultado mais adequado para nossa aplicação, porém os objetos reconhecidos ainda não são similares aos obtidos por segmentações manuais baseadas em uma análise semântica. Isso ocorre devido à não uniformidade geométrica das estruturas presentes em uma área urbana, que podem ser pequenas, como árvores isoladas e telhados de casas, ou grandes, como ruas, agrupamentos de árvores e grandes construções. Na Figura 19-b, pode-se perceber que o método de CBMA (atributos geométricos) super-segmentou as pavimentações e o agrupamento de árvores presente em torno da igreja.

Para resolver esta dificuldade, foram implementados dois métodos de segmentação baseados em *watershed* que utilizam informação de cor no processo de segmentação. Os resultados produzidos por estes métodos podem ser observados nas Figuras 20-c e 20-d e serão definidos como *cor a posteriori* e *watershed colorido* respectivamente.



**Figura 20** – Segmentação por watershed de áreas de subúrbio. (a) Imagem original, (b) segmentação com CBMA geométrico, (c) segmentação com *cor a posteriori*, (d) segmentação com *watershed colorido*.

O método de *cor a posteriori* é a maneira mais simples de corrigir as imperfeições apresentadas pelo CBMA com limiares geométricos; o resultado obtido pela segmentação

(Figura 19-b) pode ser colorido artificialmente através da média de cor dos *pixels* de cada objeto na imagem original. O resultado podendo ser observado na Figura 21-b. Após o processo de segmentação e colorização artificial dos objetos, o *cor a posteriori* agrupa os objetos vizinhos que possuem uma diferença de cor menor que um determinado limiar, eliminando a linha que existe entre eles. Na Figura 20-c, pode-se observar o resultado da aplicação de *cor a posteriori* sobre o resultado obtido por CBMA geométrico (Figura 20-b). O resultado da aplicação do filtro de cor reduz dramaticamente a super-segmentação apresentada pelo método original.

Neste projeto, propõe-se a utilização de uma maneira mais elaborada de se introduzir o limiar de cores em um CBMA, através da modificação do NHWS, de modo a que no lugar de parâmetros geométricos seja utilizada, apenas a informação de cor, como pode ser visto na Figura 18-f. Durante o processo de alagamento dos vales da imagem de bordas (Figura 10) o critério a ser utilizada para a união ou não de diferentes objetos será o parâmetro de cor, onde essa informação é o resultado da média dos *pixels* do objeto em cada etapa do processo de alagamento, sendo necessária portanto a utilização da imagem original durante o processo. Este método será chamado de *Watershed colorido* e seu resultado pode ser observado na Figura 20-d.

```

Watershed Colorido:

Entradas: S(k), grupo de regiões segmentadas na altura k;
              C(k+1), a seleção baseada na altura k;
              Tc, o limiar de cor;
              Tv, o limiar de volume;

Output:   S(k+1), regiões segmentadas na altura k+1;

S(k+1) <- 0;
For each área selecionada b ∈ C(k+1):
    r <- 0;
    If b intersepta mais do que uma região em S(k) and ao menos um
        par destas regiões tem uma média de cor que difere mais que Tc:

        For each área intersectada r':

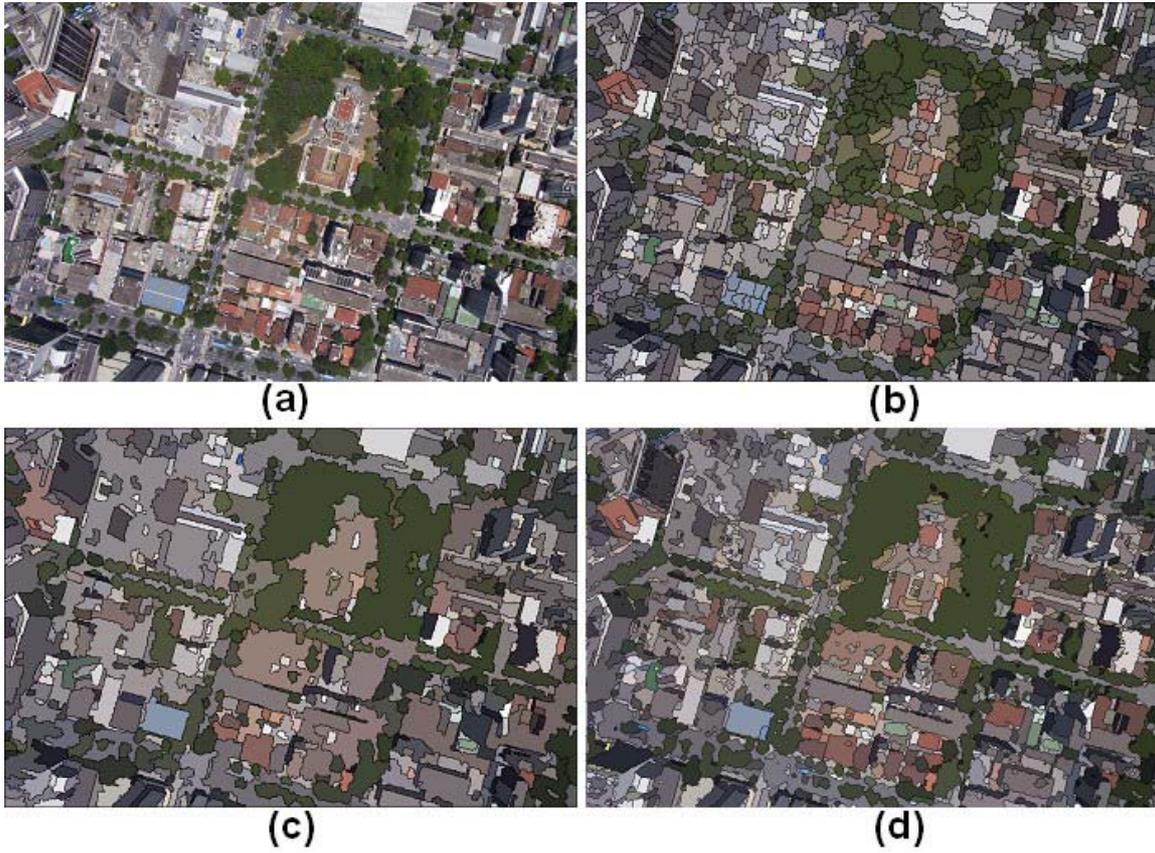
            If volume(r') > Tv:
                S(k+1) <- S(k+1) U {r'};
            Else:
                r <- r U r'; (Região de interseptação)
            End for each:

    Else:
        R <- b:

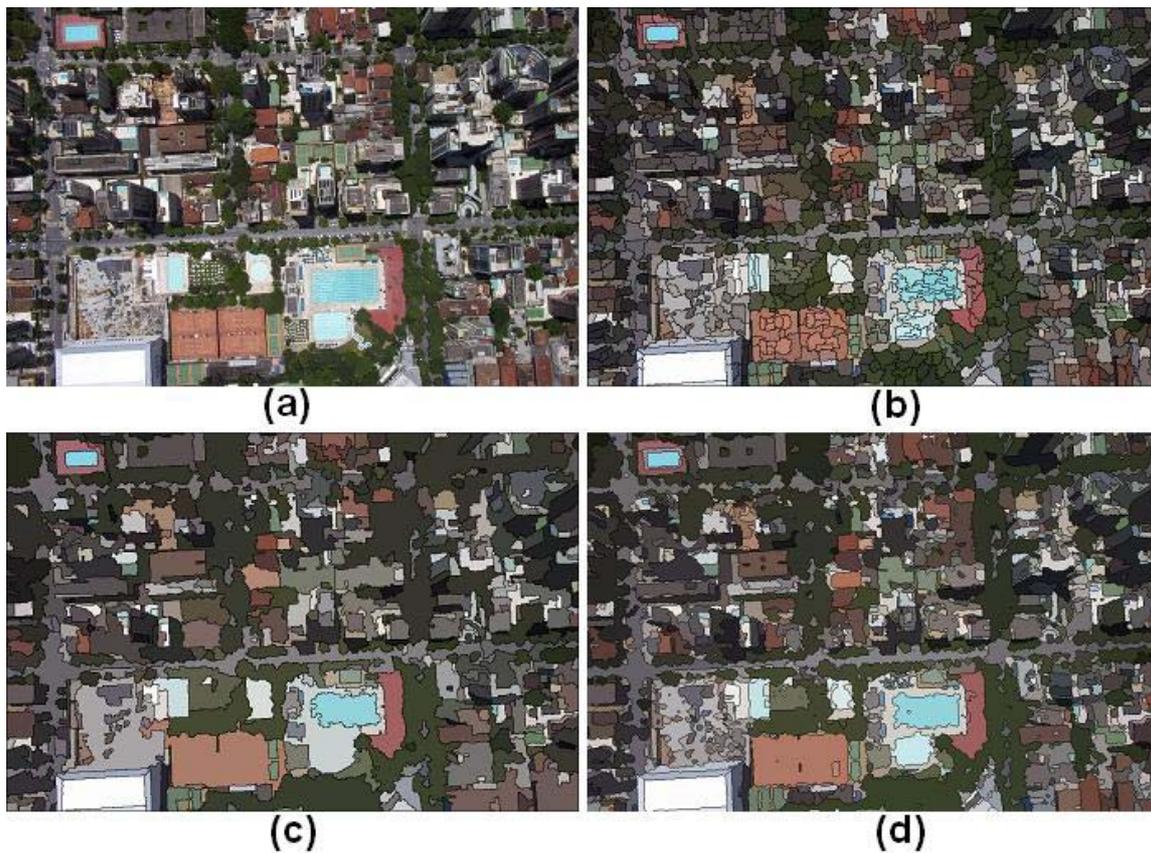
        S(k+1) <- S(k+1) U {r'};
End for each.

```

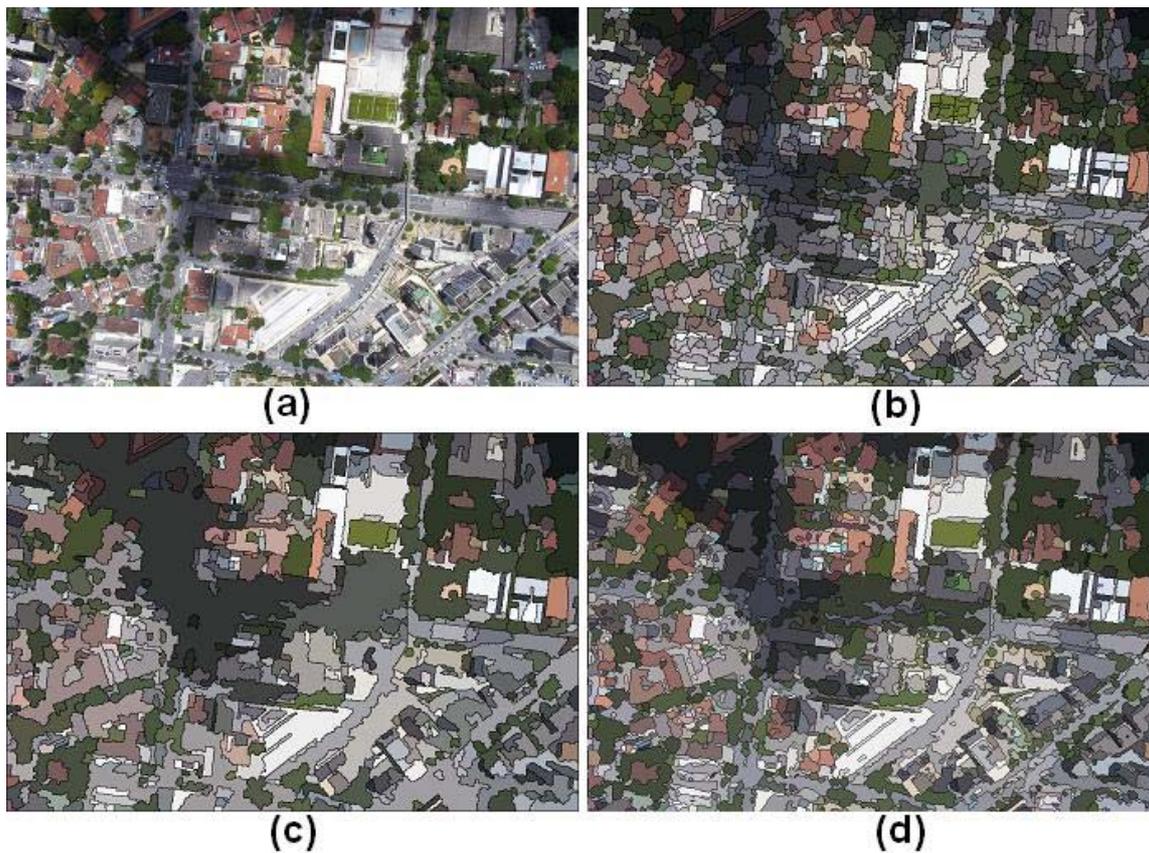
**Código 1:** Algoritmo de *watershed* baseado no uso de comparações de cor *on-the-fly*.



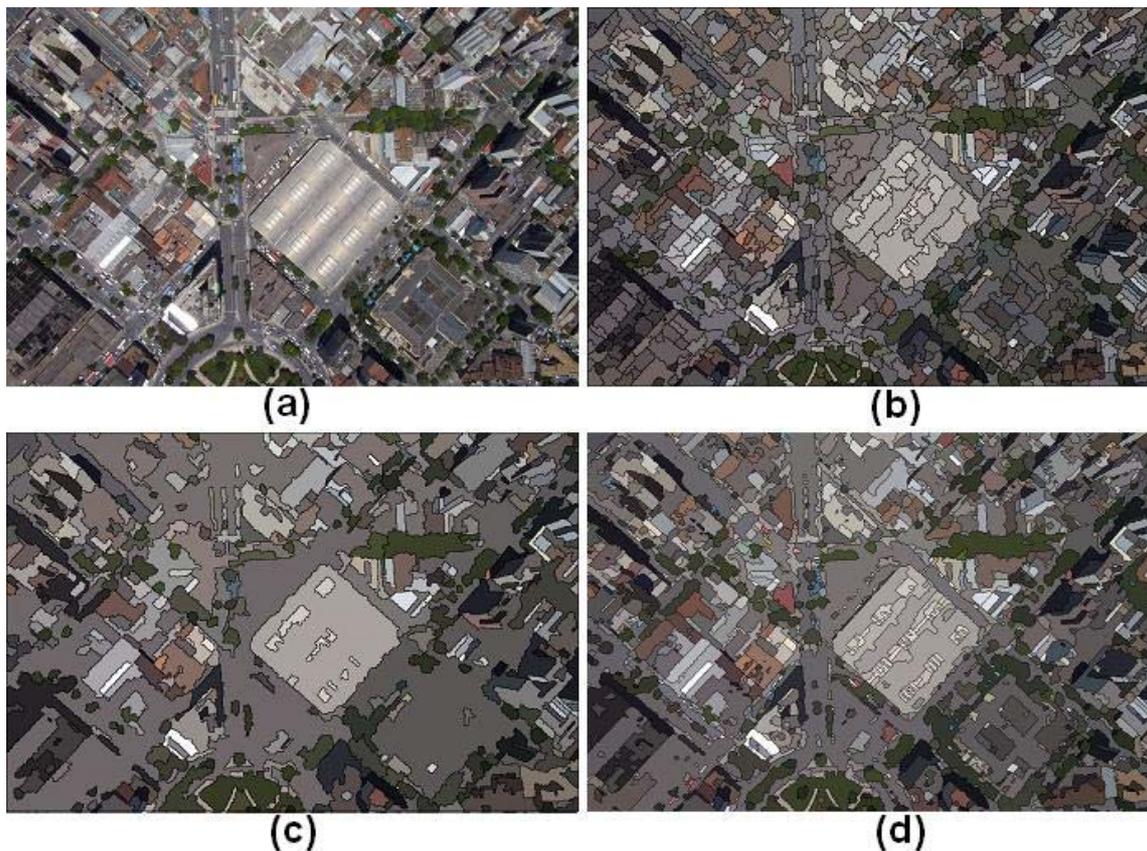
**Figura 21** – Segmentação por *watershed* em áreas de centro urbano. (a) Imagem original, (b) segmentação com CBMA geométrico, (c) segmentação com *cor a posteriori*, (d) segmentação com *watershed colorido*.



**Figura 22** – Segmentação por *watershed* em áreas de centro urbano. (a) Imagem original, (b) segmentação com CBMA geométrico, (c) segmentação com *cor a posteriori*, (d) segmentação com *watershed colorido*.



**Figura 23** – Segmentação por *watershed* em áreas de centro urbano. (a) Imagem original, (b) segmentação com CBMA geométrico, (c) segmentação com *cor a posteriori*, (d) segmentação com *watershed colorido*.



**Figura 24** – Segmentação por *watershed* em áreas de centro urbano. (a) Imagem original, (b) segmentação com CBMA geométrico, (c) segmentação com *cor a posteriori*, (d) segmentação com *watershed colorido*.

Pode-se perceber nitidamente pelos resultados apresentados nas Figuras 20 e 21 que a aplicação de limiares de cor, pelos dois métodos propostos, apresentou um resultado muito mais semelhante a uma análise semântica da imagem do que o algoritmo de CBMA geométrico. Comparando os resultados de *Watershed colorido* (Figura 20-d e Figura 21-d) com os resultados de *cor a posteriori* (Figura 20-c e Figura 21-c), pode-se perceber que o primeiro método retornou um resultado mais significativo que o *a posteriori*; uma análise numérica mais detalhada da comparação dos métodos será feita na Seção 5.1.

A explicação para o resultado ter sido diferente entre as duas propostas da utilização de cor se baseia no seguinte argumento, a aplicação posterior de cor apresenta reflexos, muitas vezes errônea, do resultado de CBMA geométrico, pois o utiliza como base de seu processo; já o *watershed* com cor é mais independente de parâmetros geométricos. Essa utilização de cor internamente ao *watershed*, que será denominada de *on the fly*, é uma implementação muito mais robusta do que o primeiro método. O resultado foi que o reconhecimento de elementos por cor *on the fly* reconheceu melhor diversos tipos de objetos, independentemente de seu tamanho e forma.

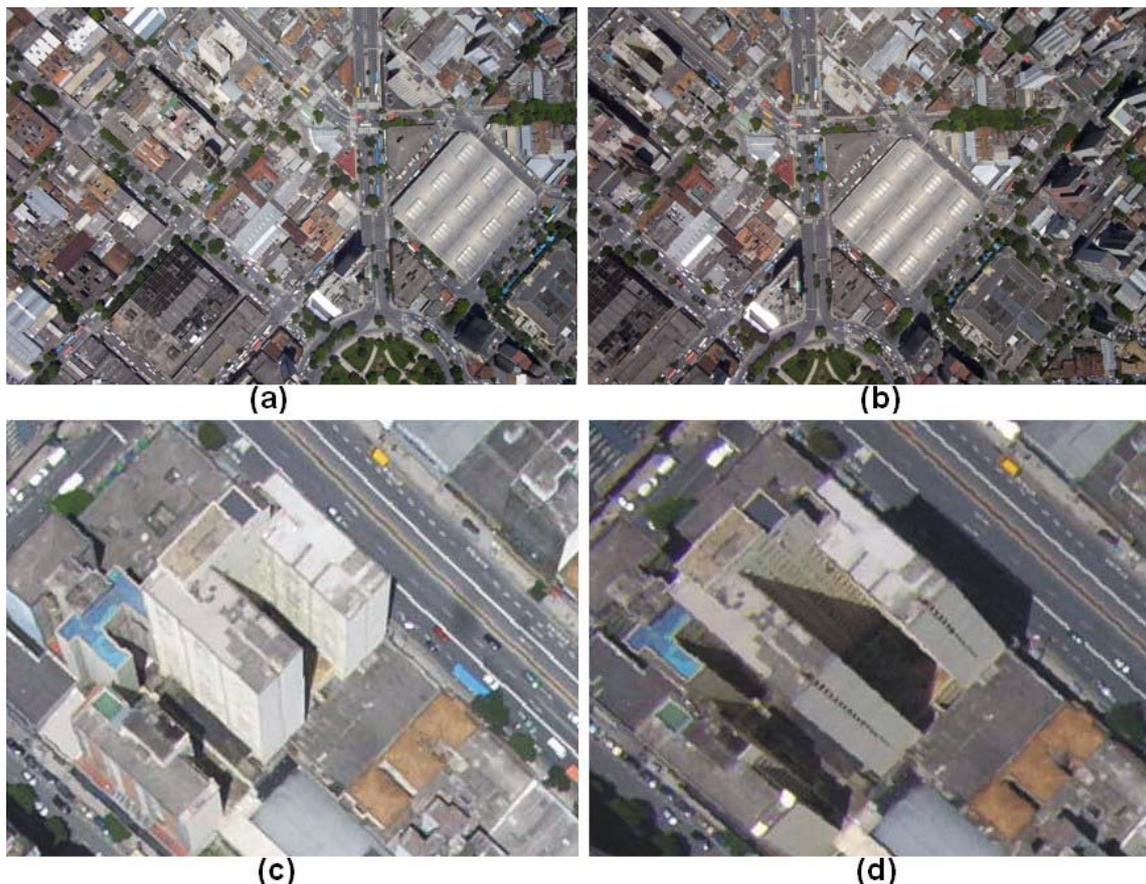
## CAPÍTULO 4 – ESTEREOSCOPIA

---

### 4.1 – METODOLOGIA

A estereoscopia consiste em fazer uma correlação entre *pixels* de um par de imagens, chamado de par estéreo, sendo que este mapeamento deverá ser otimizado de forma a minimizar erros apresentados no resultado final, que é chamado de imagem de disparidade (Figura 9 e 17). Estes métodos são chamados de estereo densos, ou *Dense Stereo*, e possuem um grande problema de otimização. Pesquisas nesta área têm focado em localizar restrições no conjunto de possíveis mapeamentos de forma a diminuir o custo computacional do processo.

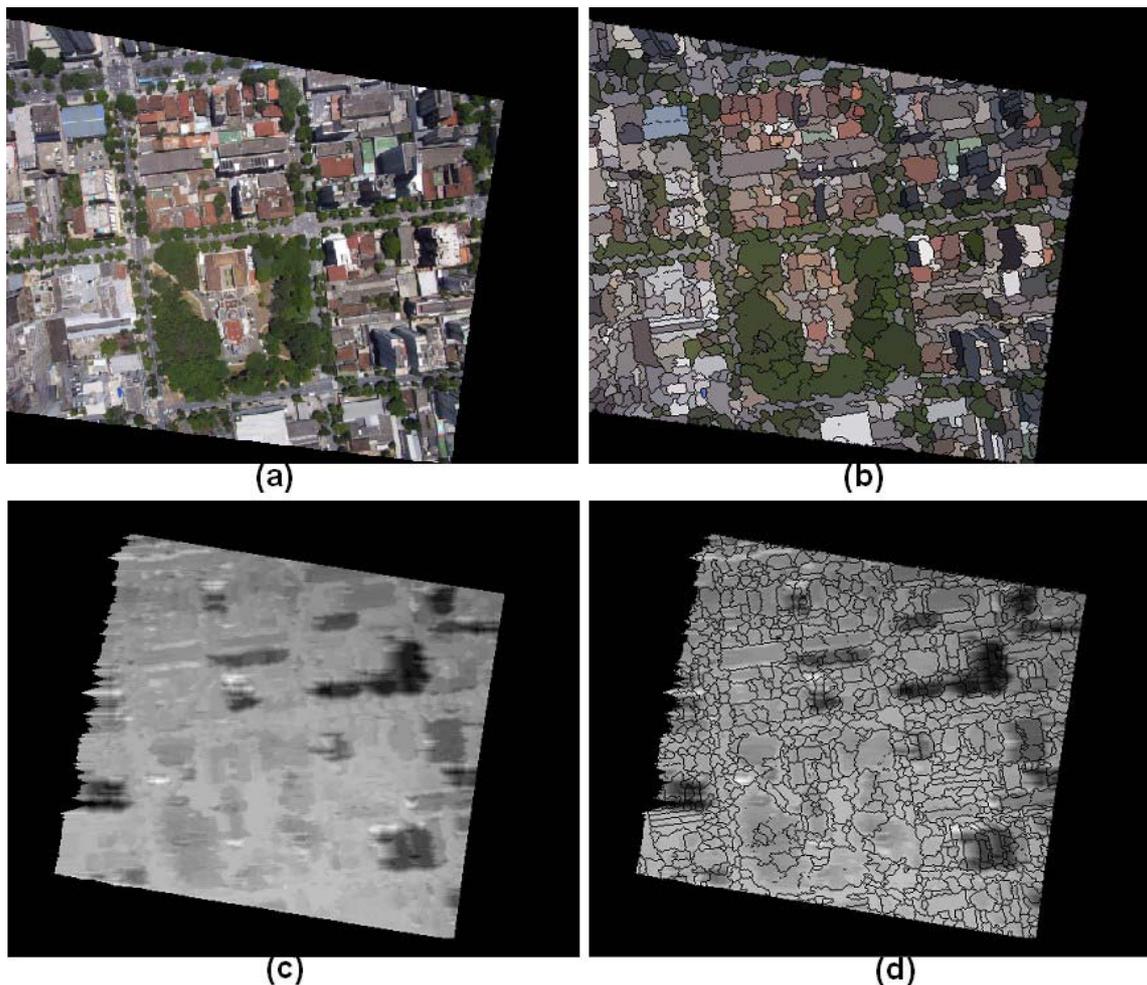
Nesta dissertação trata-se especificamente de estereoscopia em pares de imagens com grandes áreas ocultas (Figura 25). Métodos desenvolvidos especificamente para permitir alto nível de descontinuidade possuem um peso computacional extremamente elevado, devido ao fato de estarem mapeando a disparidade considerando um espaço de possibilidades muito maior do que métodos que não tratam do problema de grandes descontinuidades. Nosso método é rápido o suficiente para permitir o tratamento de grandes áreas, permitindo descontinuidades e com baixo custo computacional.



**Figura 25** – Fotografias digitais obtidas com a DX3700 sobre o Mercado Central, BH, MG. (a), (b) Par estéreo, (c) e (d) áreas de oclusão em detalhe.

Para resolver o problema, que é o de desenvolver um algoritmo com um baixo custo computacional e com suporte a regiões descontínuas, opta-se por restringir a existência de descontinuidades apenas às bordas localizadas pelo algoritmo de segmentação utilizado, *Watershed colorido*, descrito na Seção 3.2. Na figura 26 pode-se visualizar uma imagem de disparidade sobreposta pelas linhas segmentadas da imagem original.

O método escolhido utiliza como entrada um par de imagens retificadas (Figura 26-a) e executa uma otimização independente para cada par de linhas epipolares (Figura 15). O algoritmo reduz o problema de localizar o mapeamento para cada linha epipolar em um problema de menor caminho (Figura 16 e Seção 2.3.3). Este problema pode ser resolvido eficientemente via programação dinâmica (Seção 4.3.1). Tradicionalmente, essa solução permite apenas pequenas descontinuidades nas áreas de oclusão, borrando o resultado de disparidade nos locais de grandes oclusões (Figura 26-c). A novidade na implementação proposta neste projeto está justamente em permitir grandes descontinuidades justamente nas bordas localizadas pela segmentação da imagem original (Figura 26-d), reduzindo dramaticamente o borrimento do resultado final, como será mostrado nas Seções 4.4 e 5.1.



**Figura 26** – (a) Imagem retificada, (b) segmentação por *watershed colorido*, (c) disparidade localizada, (d) sobreposição das linhas segmentadas na imagem de disparidade.

## 4.2 – RETIFICAÇÃO

A etapa de retificação consiste em preparar o par de imagens para o processo de localização da disparidade dos *pixels*, o qual será explicado na Seção 4.3. Basicamente, esta etapa localiza os pares de linhas epipolares nas duas imagens (Figura 14) e gera duas imagens retificadas, nas quais as linhas epipolares ficam paralelas umas às outras (Figura 15). Para se localizar as linhas epipolares é necessário conhecer a matriz essencial entre o par de imagens. Pela Figura 13, pode-se perceber que para cada ponto em uma imagem existe uma linha correspondente a este ponto na outra imagem que cruza o epipolo. A correspondência entre essas linhas (Figura 13) é dada justamente pela matriz essencial, através da Equação 1, onde  $u'$  e  $u$  são quaisquer *pixels* de imagens distintas do par estéreo pertencentes às linhas epipolares correspondentes, e  $E$  é a matriz essencial.

$$\mathbf{u}'^T \mathbf{E} \mathbf{u} = 0$$

**Equação 1** – Solução linear para a matriz essencial.

O processo utilizado para a localização da matriz essencial foi o algoritmo de 8 pontos, proposto por Longuet-Higgins [11] e defendido por Hartley [31]; o motivo da utilização deste método é o fato dele ser rápido e de fácil implementação. Após a localização da matriz essencial, o próximo passo é gerar as imagens retificadas. No caso específico das imagens tratadas, utilizaram-se as restrições da base a ser tratada para implementar um método rápido e eficiente para a retificação de imagens aéreas. Este método de retificação será descrito na Seção 4.2.4.

#### 4.2.1 – ALGORITMO DE 8 PONTOS

O algoritmo de 8 pontos baseia-se no conhecimento de 8 pares de *pixels* correspondentes entre as imagens, chamados em SIG de *tie points* e em Visão Computacional de pontos fiduciais. A partir deste conhecimento, a matriz essencial pode ser calculada através de um conjunto de equações lineares. Mais pontos podem ser utilizados no processo, resultando em maior qualidade do resultado apurado. A propriedade primordial da matriz essencial é o fato de possuir informações relativas à geometria epipolar da imagem (Figura 27).

Na Equação 1, tem-se a solução linear para a matriz essencial, que é a base do algoritmo de 8 pontos; os pontos  $\mathbf{u}$  e  $\mathbf{u}'$  serão escritos como  $(u,v,1)$  e  $(u',v',1)$  respectivamente, o valor 1 corresponde a distância focal e será utilizado para facilitar as contas. Resolvendo a Equação 1 com os valores de  $\mathbf{u}$  e  $\mathbf{u}'$ , tem-se o resultado mostrado na Equação 2, que é uma equação linear do problema.

$$\mathbf{u}\mathbf{u}'^T\mathbf{E}\mathbf{1}\mathbf{1} + \mathbf{u}\mathbf{v}'^T\mathbf{E}\mathbf{2}\mathbf{1} + \mathbf{u}\mathbf{E}\mathbf{3}\mathbf{1} + \mathbf{v}\mathbf{u}'^T\mathbf{E}\mathbf{1}\mathbf{2} + \mathbf{v}\mathbf{v}'^T\mathbf{E}\mathbf{2}\mathbf{2} + \mathbf{v}\mathbf{E}\mathbf{3}\mathbf{2} + \mathbf{u}'\mathbf{E}\mathbf{1}\mathbf{3} + \mathbf{v}'\mathbf{E}\mathbf{2}\mathbf{3} + \mathbf{E}\mathbf{3}\mathbf{3} = 0$$

**Equação 2** – Equação linear.

De posse de 8 pontos ou mais, tem-se 8 ou mais equações iguais à Equação 2, o que será nosso conjunto de equações lineares. Esse conjunto pode ser representado como a Equação 3, onde  $\mathbf{E}$  é um vetor de 9 posições contendo os parâmetros da matriz essencial ( $E_{11}, E_{21}, E_{31}, E_{12}, E_{22}, E_{32}, E_{13}, E_{23}, E_{33}$ ) e  $\mathbf{A}$  é a matriz de coeficientes lineares cujas linhas são da forma  $(uu', uv', u, vu', vv', v, u', v', 1)$ .

$$\mathbf{AE} = \mathbf{0}$$

**Equação 3** – Equação linear.

O cálculo de E pode ser feito portanto, através da matriz A, conforme proposto por Hartley [31]. Utiliza-se uma decomposição desta matriz, chamada de *Singular Value Decomposition* [48] (SVD) para obter os valores de E.

#### 4.2.2 – SINGULAR VALUE DECOMPOSITION

Escolhe-se o método de SVD para resolver o problema de equações lineares da Equação 3; os métodos SVD se baseiam no seguinte teorema de álgebra linear: qualquer matriz A (M x N), onde o número de linhas M é maior ou igual ao número de colunas N, pode ser escrita como o produto de uma matriz U (M x N), uma matriz diagonal W (N x N) e uma matriz V (N x N) transposta (Equação 4).

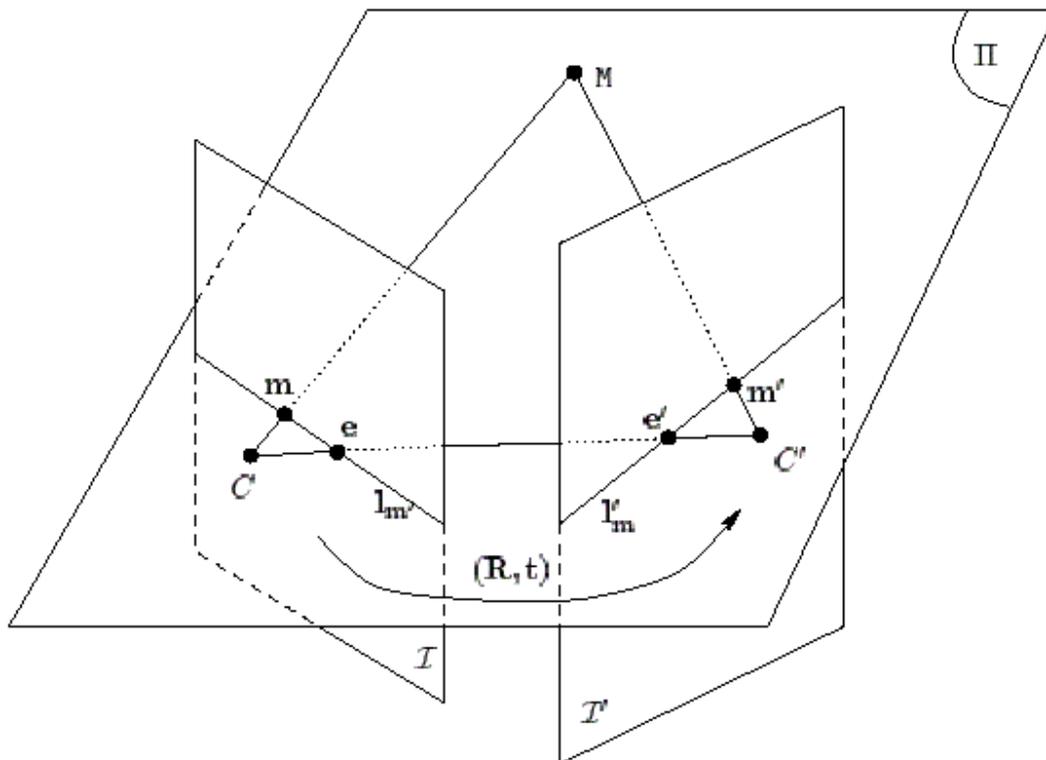
$$\begin{pmatrix} \mathbf{A} \end{pmatrix} = \begin{pmatrix} \mathbf{U} \end{pmatrix} \cdot \begin{pmatrix} w_1 & & & \\ & w_2 & & \\ & & \dots & \\ & & & w_N \end{pmatrix} \cdot \begin{pmatrix} \mathbf{V}^T \end{pmatrix}$$

**Equação 4** – *Singular Value Decomposition* de uma matriz A.

A solução da decomposição de A para o problema será referente ao menor valor w da matriz W (  $w_1 \dots w_n$  ); normalmente este menor valor no problema de estereoscopia é o último valor,  $w_n$ . Descoberto o menor valor da diagonal de W, a solução será a linha de  $\mathbf{V}^T$  referente a este valor. Caso o elemento de W escolhido seja o  $w_n$ , então a linha N de  $\mathbf{V}^T$  será a matriz essencial, onde a linha N de  $\mathbf{V}^T$  é a coluna N de V.

### 4.2.3 – MATRIZ ESSENCIAL

No problema específico deste trabalho, a utilização da matriz essencial será para localizar os epíolos e os pares de linhas epipolares; o cálculo de rotação e translação a partir da matriz essencial [5] não será utilizado nesta dissertação. Os dois epíolos do par estéreo e a matriz essencial serão utilizados no processo de retificação descrito na Seção 4.2.4.

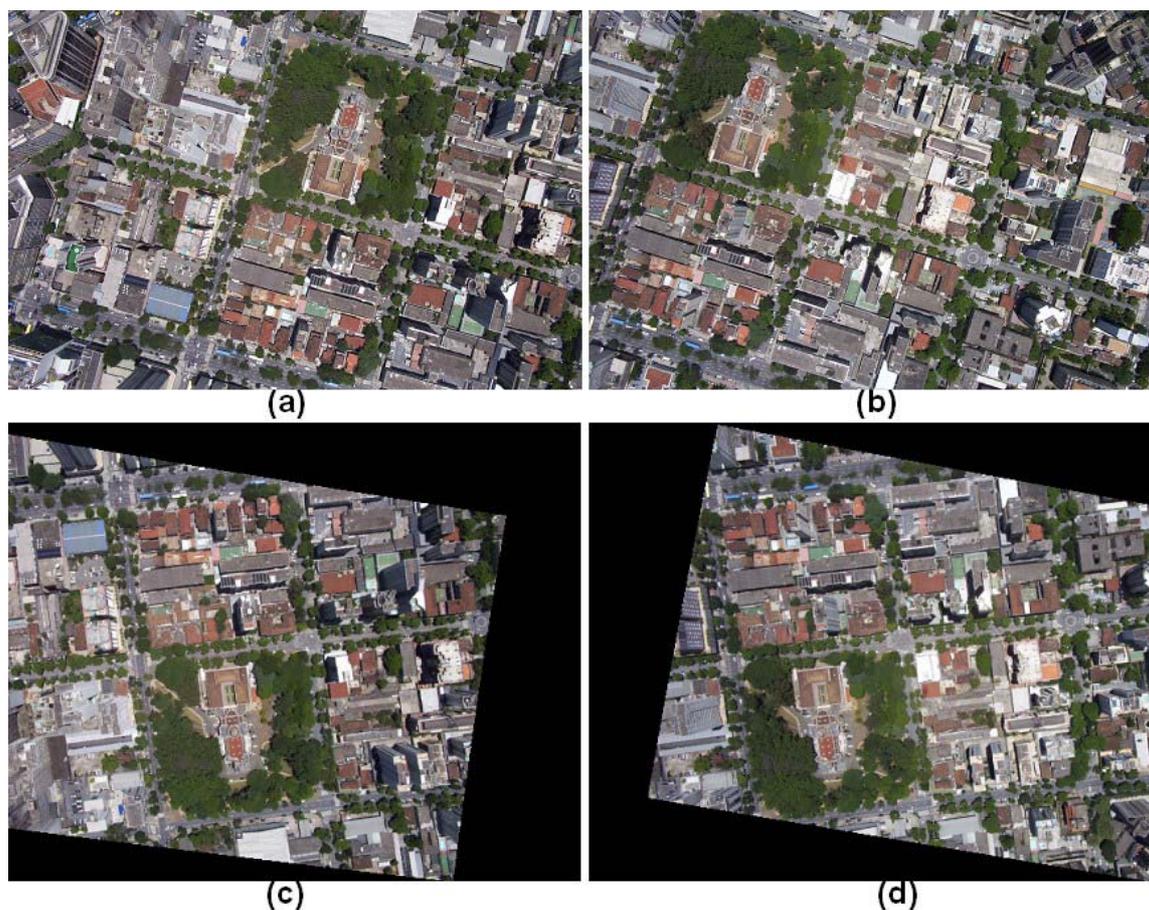


**Figura 27** – Geometria Epipolar.

A técnica de 8 pontos foi utilizada para localizar a matriz essencial entre o par, e através desta localizar as linhas epipolares. Na Figura 27, pode-se visualizar o par estéreo  $(I, I')$ , os epíolos  $(e, e')$  e o par de linhas epipolares  $(l_m, l'_m)$ . A matriz essencial acumula as informações de rotação e translação  $(R, T)$  entre as imagens.

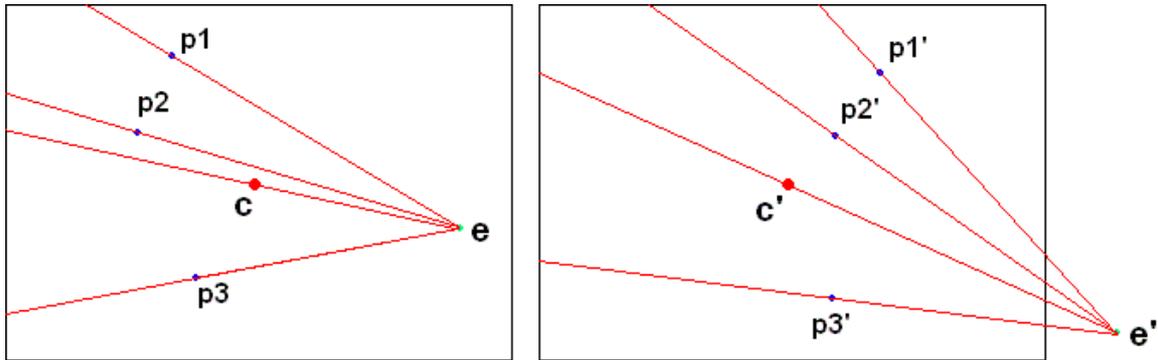
#### 4.2.4 – TÉCNICA UTILIZADA

A técnica proposta para a retificação das imagens apresentou um resultado excelente, apesar de ser extremamente simples, como se pode observar na Figura 30. Uma das peculiaridades da técnica é que o par retificado resultante fica previamente corrigido com relação ao deslocamento de posição, caracterizado pelo vôo do avião, como pode-se observar na Figura 28, facilitando em muito o processo de localização de disparidade (Seção 4.3).



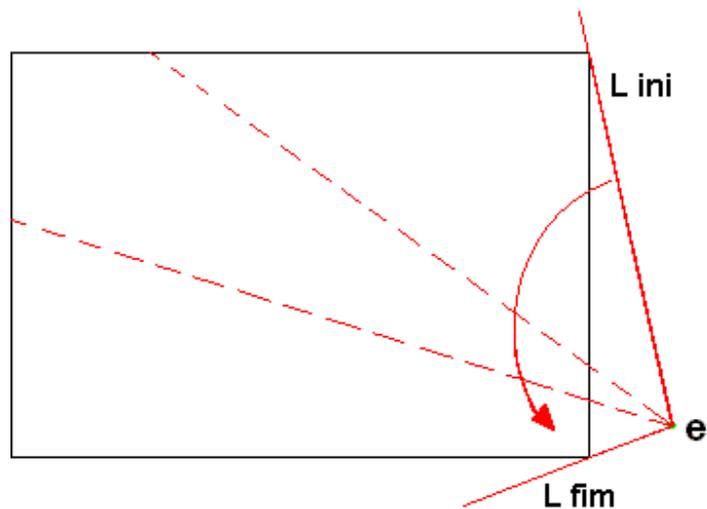
**Figura 28** – Retificação de par estereo. (a) e (b) Imagens originais, (c) e (d) imagens retificadas.

A técnica se baseia na restrição de que as fotografias aéreas são batidas mais ou menos com a mesma orientação com relação ao solo, perpendiculares ao mesmo, ou como é dito em fotogrametria, as fotos são batidas em *NADIR*. Portanto o problema de retificação pode ser transformado em um problema de duas dimensões, se tornando muito mais simples.



**Figura 29** – Retificação de par estéreo.  $(e, e')$  Epípólos,  $(p1, p1')$   $(p2, p2')$   $(p3, p3')$  3 pares de pontos utilizados no *8-point algorithm*,  $(c, c')$  centros das imagens.

A Figura 29 mostra os pontos centrais, epípólos e pontos comuns, ou *tie-points*, entre as imagens do par estéreo; o processo de retificação adotado pode ser facilmente exemplificado por através de uma análise desta figura. Pode-se dividir o algoritmo em duas etapas. Na primeira é localizado o par de linhas epipolares e na segunda etapa este par de linhas é adicionado às imagens retificadas.



**Figura 30** – Varredura de linhas epipolares.

Pela Figura 30, pode-se perceber o funcionamento da primeira etapa, que consiste em um *loop* no qual são localizadas todas as linhas epipolares entre *Lini* e *Lfim* existentes na imagem. Para cada linha localizada, ela é multiplicada pela matriz essencial para retornar seu par, e ambas as linhas são salvas nas imagens retificadas. A segunda etapa consiste em colocar as linhas na melhor posição possível nas imagens retificadas; isso é feito através de uma simples comparação entre os centros das imagens, os epípólos e os *tie-points* localizados.

Na Figura 29, os pontos (p1,p2,p3) e seus correspondentes (p1',p2',p3') estão a uma distância média de seus epíolos chamadas de  $dm$  e  $dm'$  respectivamente. Os centros das duas imagens (c,c'), por sua vez, estão a uma distância  $dc$  e  $dc'$  de seus epíolos. De posse dos parâmetros  $dm$ ,  $dm'$ ,  $dc$ ,  $dc'$ , cada uma das linhas epipolares de I e I' será desenhada na imagem retificada na posição  $x$  dada pelas Equações 5-a e 5-b, respectivamente.

$$\begin{aligned} 5a & - \quad \text{PosX} = dm - \{ [(dm - dc) - (dm' - dc')] / 2 \} \\ 5b & - \quad \text{PosX}' = dm' - \{ [(dm' - dc') - (dm - dc)] / 2 \} \end{aligned}$$

**Equação 5** – Correção na posição das linhas epipolares durante o processo de montagem do par retificado.

Apesar de extremamente simples, este processo garante que as imagens retificadas estejam corretamente posicionadas com relação à média de disparidade dada pelos seus *tie-points*, o que auxilia o processo de localização de disparidade a ser descrito na Seção 4.3, pois reduz drasticamente o espaço de busca da disparidade. Na Figura 31 visualiza-se dois cortes de um par retificado sobrepostos; a figura 32 também exemplifica este resultado.

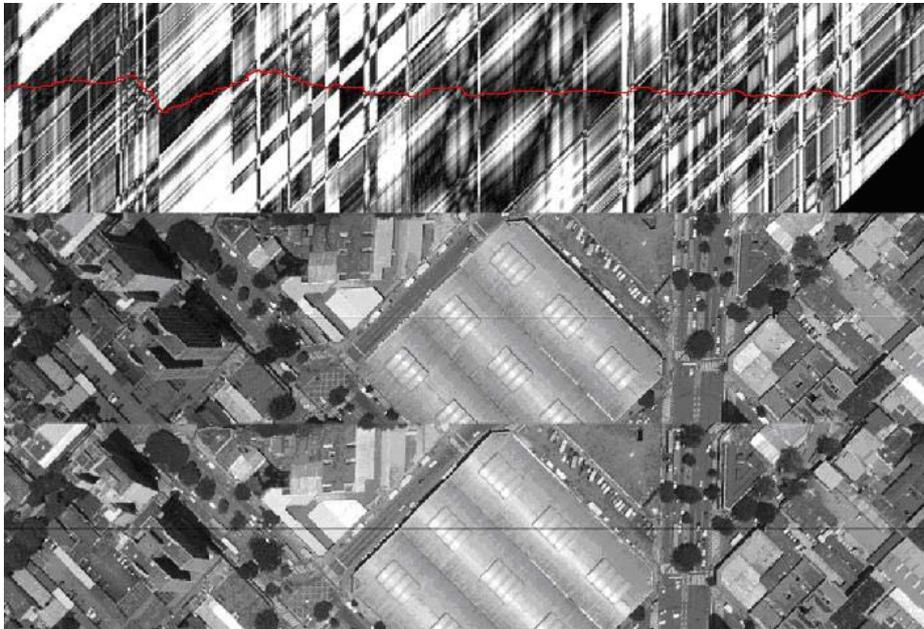


**Figura 31** – Sobreposição de cortes em um par retificado.

### 4.3 – CASAMENTO DENSO

O resultado obtido pelas seções anteriores é uma preparação para o passo que será executado nesta seção; as imagens retificadas são uma simplificação para o processo de obtenção da disparidade. Este processo será executado independentemente sobre cada par de linhas epipolares, que agora estão paralelas umas às outras, o problema se reduzindo a localizar a disparidade sobre cada par de linhas das imagens retificadas.

O processo que iremos descrever será executado sobre cada uma das  $n$  linhas das imagens retificadas. Basicamente o algoritmo pode ser dividido em duas partes distintas, a montagem do mapa de disparidade e a posterior localização do caminho mínimo no mesmo (linha vermelha na Figura 32).



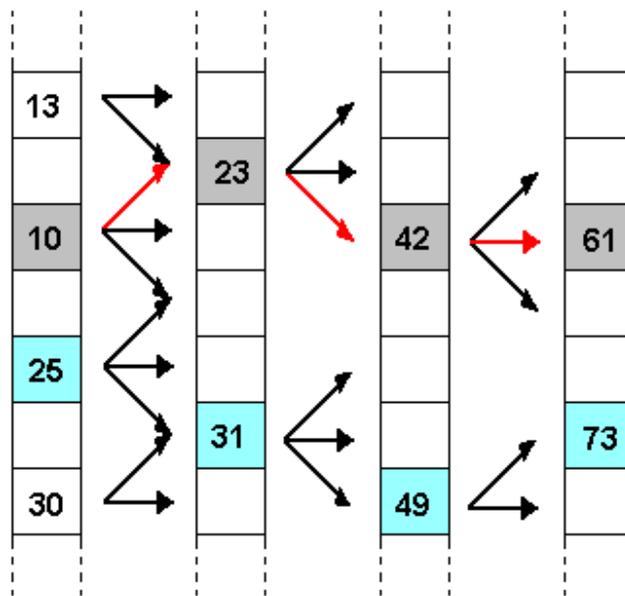
**Figura 32** – Caminho Mínimo sobre o mapa de disparidade.

Conforme já foi dito na Seção 2.3.3, o mapa de disparidade é feito da seguinte forma: para cada elemento  $I$  da linha epipolar de uma imagem do par retificado, foram testados  $N$  tamanhos diferentes de deslocamento, e os resultados são colocados na coluna  $I$  do mapa de disparidade, onde esse resultado é apenas um filtro aplicado entre os valores do *pixel* de uma imagem e o *pixel* referente ao deslocamento proposto no seu par. O segredo do bom resultado do mapa está justamente no filtro a ser utilizado, o qual será descrito na Seção 4.3.2. Após a montagem do mapa de disparidade, o caminho mínimo (Figura 32) do mesmo será localizado através de programação dinâmica (Seção 4.3.1). Este caminho é o resultado final de disparidade sobre a respectiva linha epipolar.

### 4.3.1 – PROGRAMAÇÃO DINÂMICA

A localização do resultado correto de disparidade será feita através do caminho mínimo no mapa de disparidade (Figura 32) referente a cada par de linhas epipolares. O processo de obtenção desse caminho é feito através de programação dinâmica.

Cada elemento do mapa possui um valor que varia de 0 a 255, de preto ao branco, onde quanto menor o valor, mais chances ele terá de ser um resultado válido de disparidade. Facilmente, pode-se perceber que existem inúmeros caminhos possíveis no mapa; comparar todos os caminhos é extremamente oneroso computacionalmente, portanto o objetivo é reduzir o processo a um problema de soluções locais.



**Figura 33** – Programação dinâmica sobre o mapa de disparidade.

O método é bastante simples; inicia-se com a primeira coluna do mapa de disparidade, percorre-se de 1 a N os elementos desta coluna, testando para cada um deles se é o menor elemento vinculado a I elementos da próxima coluna (no caso da figura 33 foram testados 3 caminhos locais, ou seja,  $I = 3$ ); caso ele seja o menor elemento, então a escolha de caminhos locais para aquela coluna foi feita. A próxima coluna a ser tratada já tem o valor de seus elementos somados com os elementos escolhidos da primeira coluna, o processo continuando até que a última coluna seja tratada. O caminho mínimo do mapa será aquele que terminar com o menor valor na última coluna. No caso da Figura 33, observa-se que o menor caminho foi o com valor 61, realçado em vermelho.

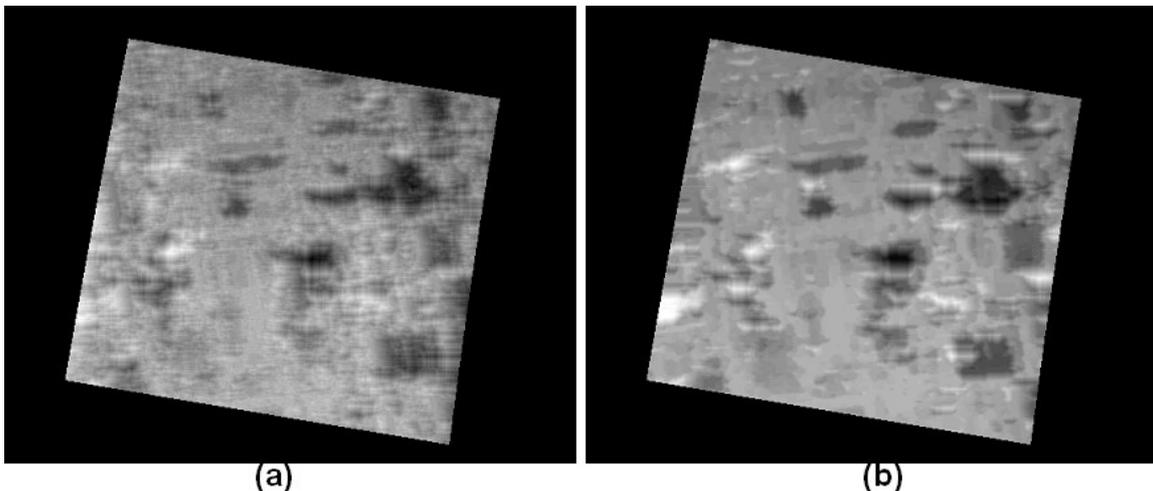
### 4.3.2 – FILTROS UTILIZADOS

O problema do filtro está relacionado diretamente com a qualidade do mapa de disparidade a ser gerado. Basicamente ele consiste na operação matemática que é feita entre pontos de uma imagem e pontos de seu par, para gerar um valor de semelhança entre estes pontos. O filtro mais simples consiste em um valor absoluto da diferença entre dois *pixels* comparados (Equação 6); por exemplo, este filtro aplicado sobre os *pixels* de valor 11 e 15 retornaria 4 como valor de diferença.

$$\text{Diferença} = \text{ValorAbsoluto}(\text{elementoI} - \text{elementoI}') )$$

**Equação 6** – Filtro de diferença simples.

O resultado da aplicação deste filtro apresenta bastante ruído (Figura 34-a), resultante do fato dele ser vulnerável a pequenos ruídos na imagem; a solução para este problema foi utilizar um filtro que abranja mais do que apenas um *pixel*. Neste caso, a comparação entre elementos do par de imagens levaria em consideração as redondezas dos elementos, sendo menos sujeito aos pequenos ruídos. O filtro utilizado nesta dissertação está descrito na Equação 7 e é um filtro de tamanho 9x9, ou seja, ele atinge 81 vezes mais elementos do que o filtro unitário. Como se pode observar na Figura 34-b, seu resultado possui bem menos ruídos que o resultado na Figura 34-a (filtro unitário).



**Figura 34** – Disparidade obtida por um filtro unitário (a) e por um filtro com tamanho 9x9 (b).

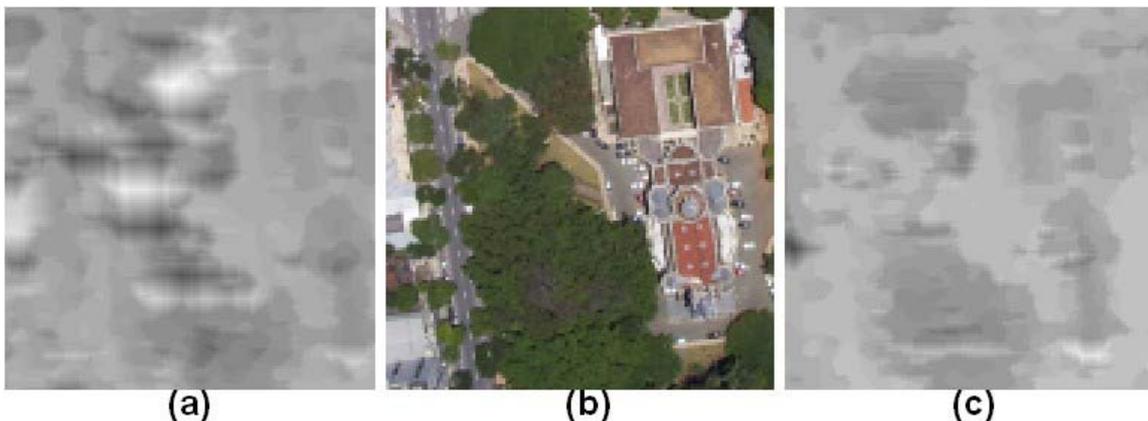
```

0, 0, 0, 0, 1, 0, 0, 0, 0
0, 0, 0, 1, 1, 1, 0, 0, 0
0, 0, 1, 1, 2, 1, 1, 0, 0
0, 1, 1, 2, 2, 2, 1, 1, 0
1, 1, 2, 2, 2, 2, 2, 1, 1
0, 1, 1, 2, 2, 2, 1, 1, 0
0, 0, 1, 1, 2, 1, 1, 0, 0
0, 0, 0, 1, 1, 1, 0, 0, 0
0, 0, 0, 0, 1, 0, 0, 0, 0

```

**Equação 7** – Filtro de diferença com tamanho 9x9 e distribuição progressiva de pesos.

O filtro com janela superior a 1x1 (unitário) utiliza a média das diferenças obtidas por cada posição do filtro. Neste sistema, utiliza-se pesos para cada uma das posições do filtro, valorizando as posições mais próximas do centro (Equação 7). Apesar deste filtro (9x9) ter melhorado muito o resultado, pode-se perceber claramente que ele apresenta diversos erros, caracterizados pelas manchas brancas na imagem (Figura 35-a).



**Figura 35** – (a) aplicação de filtro 9x9 com média simples, (b) imagem original, (c) filtro 9x9 com controle de intensidade.

Estes erros aparecem justamente em lugares onde existe uma maior homogeneidade de informações, como no conjunto de árvores presentes na Figura 35-b. A explicação para este fato é bastante simples. Cada uma das imagens do par estereo, em fotografias aéreas, é batida em posição bastante diferente, sendo os objetos vistos de ângulos distintos, mudando muito as informações referentes a sombras e luminosidades dos objetos. Em uma análise da fotografia inteira (Figura 6-a e 6-b, Figura 25-a e 25-b), percebem-se diferenças apenas nas áreas de oclusão; entretanto, existem diferenças locais nas imagens referentes à mudança do reflexo da luz solar e das sombras.

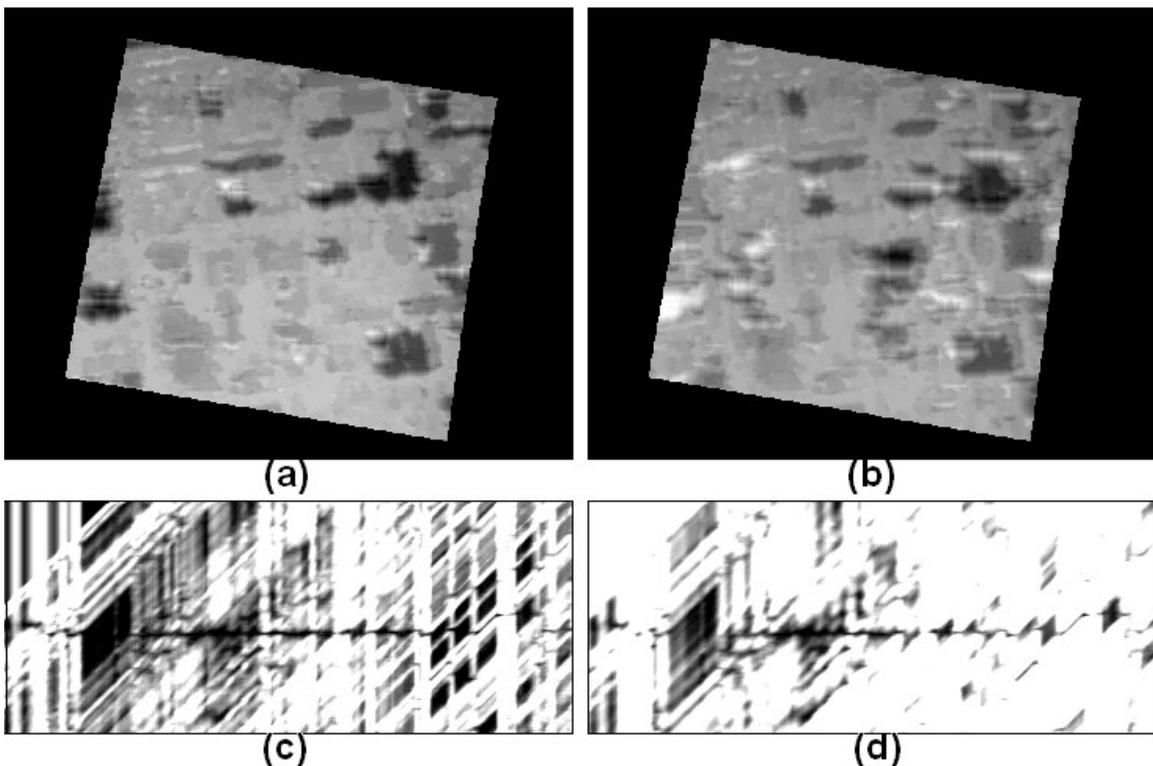
Visando corrigir este defeito, utiliza-se um filtro semelhante ao proposto na Equação 7, porém com um pré-processamento antes de sua aplicação. Como o problema a ser resolvido está relacionado com a luminosidade, a idéia foi simplesmente remover

qualquer influência da mesma no resultado do filtro. Para fazer este efeito, equaliza-se a intensidade dos *pixels* a serem tratados pelo filtro e posteriormente aplica-se sobre as informações equalizadas. A Equação 8 demonstra a equalização sobre uma janela 3x3.

Pode-se visualizar pelo resultado na Figura 35-c que o resultado da aplicação prévia da equalização removeu os erros presentes na utilização simples do filtro (Figura 35-a). O processo de equalização consiste apenas em aumentar ou diminuir os elementos utilizados de uma das imagens ( $I'$ ) em um mesmo valor, de forma que a média dos novos elementos desta imagem seja igual a média dos elementos da outra imagem ( $I$ ).

$$\begin{array}{r}
 I - \begin{array}{c} 3, 4, 5 \\ 2, 3, 4 \\ 2, 1, 4 \end{array}, \quad I' - \begin{array}{c} 5, 5, 6 \\ 4, 5, 5 \\ 4, 3, 5 \end{array} \\
 \\
 I' \text{ equalizado} - \begin{array}{c} 3, 3, 4 \\ 2, 3, 4 \\ 2, 1, 3 \end{array}
 \end{array}$$

**Equação 8** – Aplicação da equalização previamente à aplicação do filtro 9x9.



**Figura 36** – Disparidades e mapas referentes à linha central das imagens. (a) Disparidade utilizando filtro com suavização de intensidade, (b) disparidade sem suavização de intensidade, (c) mapa de disparidade da linha central da imagem utilizando filtro com suavização de intensidade, (d) mapa de disparidade sem suavização de intensidade.

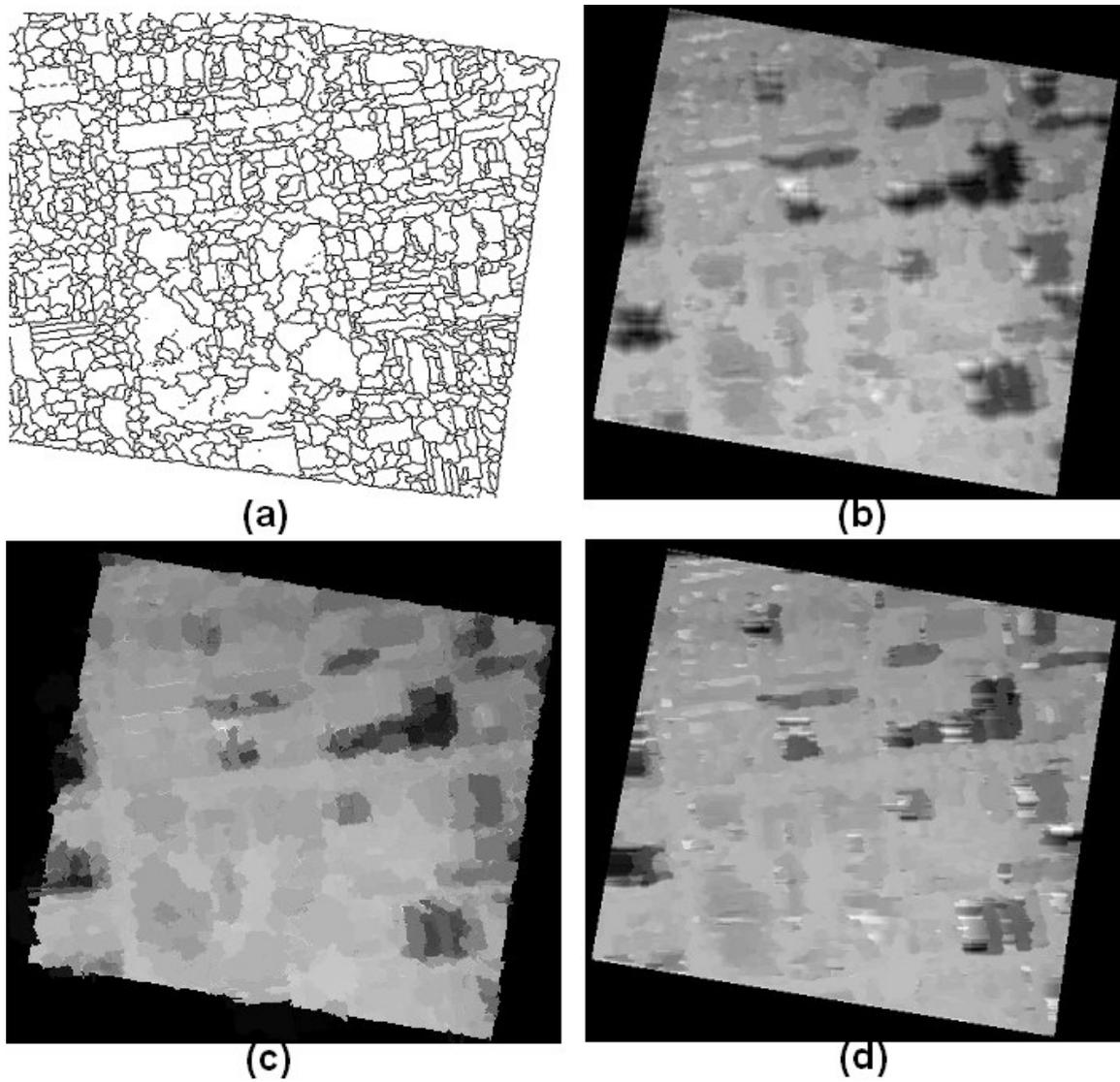
A comparação entre os resultados do processo com e sem a equalização fica mais interessante quando se analisa o mapa de disparidade gerado (Figuras 36-c e 36-d). Com uma análise visual, pode-se afirmar que é muito mais fácil descobrir o caminho mínimo no mapa sem a equalização (Figura 36-d) do que no mapa com a equalização (Figura 36-c), pois a Figura 36-d apresenta muito menos ruído no mapa de disparidade. O motivo pelo qual o efeito foi o inverso do esperado está no fato de que a linha que visualmente enxerga-se na Figura 36-d (sem equalização) está incompleta nos lugares onde houve maior mudança na luminosidade, enquanto o caminho mínimo da Figura 36-c apresenta-se completo, independente da existência de ruídos no mapa de disparidade.

#### 4.4 – UTILIZAÇÃO DA SEGMENTAÇÃO

Esta seção representa o fechamento das idéias propostas nesta dissertação; o objetivo é utilizar o resultado obtido pelo *Watershed colorido* (Seção 3.2) para aprimorar o resultado do processo de estereoscopia proposto nas Seções 4.2 e 4.3. Da mesma forma como foi proposto o algoritmo de *watershed* (Seção 3.2), sendo dividido em duas abordagens, *a posteriori* (Figura 20-c e 21-c) e *on-the-fly* (Figura 20-d e 21-d), a calibração pela segmentação também será executada sobre estes dois pontos de vista. Neste caso ambos os resultados serão válidos, podendo ser escolhidos um ou outro dependendo da aplicação.

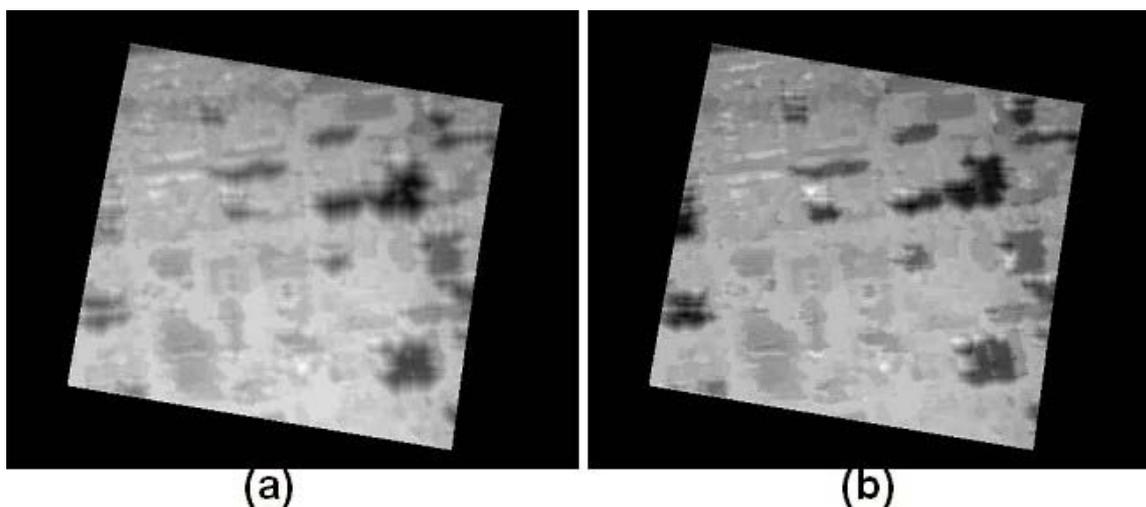
O método de calibração *a posteriori* (Figura 37-c) é apenas um filtro de média baseado nos objetos localizados pelo *watershed* (Figura 37-a) e a disparidade (Figura 37-b) localizada pelo processo proposto na Seção 4.3. Este método funciona da seguinte forma, para cada objeto segmentado é calculada a média de disparidade dos *pixels* deste objeto na imagem de disparidade (Figura 37-b). Todos os *pixels* do objeto na imagem de disparidade são substituídos pela média calculada (Figura 37-c). A disparidade obtida por este método de calibração é uma imagem mais sólida dos objetos, sendo mais fácil de perceber, tratar e visualizar as edificações presentes nas áreas urbanas. Entretanto, pequenas características da superfície são eliminadas, prevalecendo apenas informações vinculadas aos objetos da segmentação.

Apesar do resultado obtido *a posteriori* ter sido bastante eficaz em resolver o problema proposto neste projeto, que é o reconhecimento dos objetos urbanos na estereoscopia, propõe-se outro método de calibração pela segmentação, o qual é mais fiel e elimina menos informações de disparidade. Este método será executado durante o processo de localização do caminho mínimo no mapa de disparidade (Seção 4.3.1) e será chamado de *on-the-fly*. O resultado da aplicação do método *on-the-fly* (Figura 37-d) reconhece os limites dos objetos da imagem, em oposição ao borrão localizado no resultado original (Figura 37-b), e mantém as características da superfície. Alguns erros podem ser visualizados no método *on-the-fly*, porém a maior parte deles é referente a áreas ocultas e sua remoção será objeto de estudos futuros.



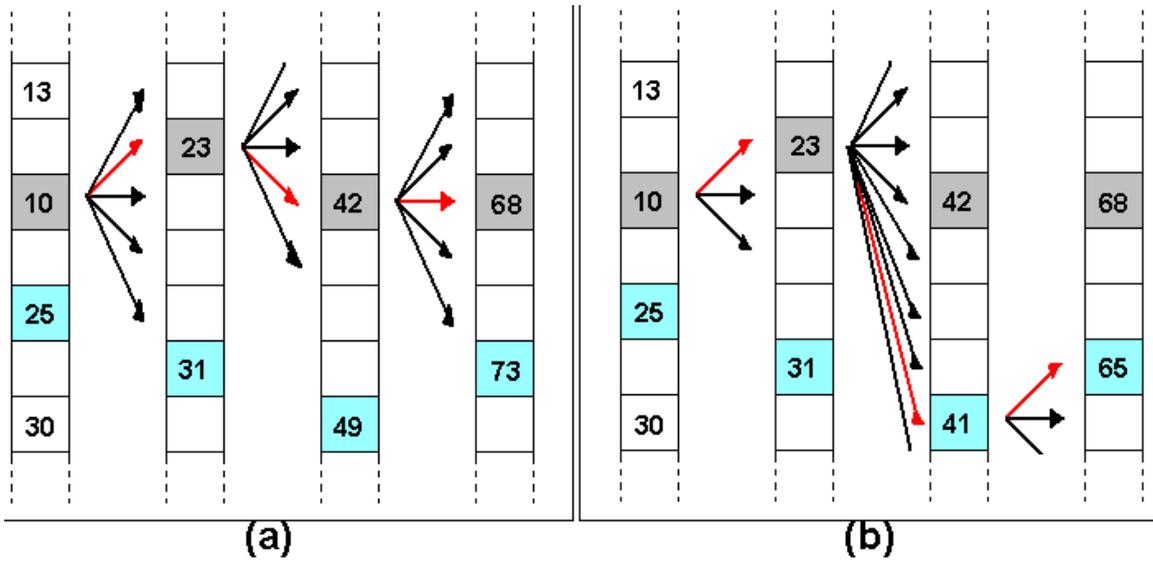
**Figura 37** – Calibração da disparidade, (a) segmentação obtida por *Watershed colorido*, (b) disparidade com suavização de intensidade, (c) disparidade calibrada *a posteriori* pela segmentação, (d) disparidade calibrada *on-the-fly* pela segmentação.

O algoritmo de calibração *on-the-fly* se baseia na variação da liberdade de movimento do algoritmo de caminho mínimo no mapa de disparidade (Seção 4.3.1). Na Figura 33 (Seção 4.3.1), pode-se perceber que existe uma liberdade de movimento do caminho no mapa de disparidade de apenas um degrau para cada coluna do mapa. O resultado obtido especificamente com essa liberdade está demonstrado na Figura 38-a. Pode-se perceber que as edificações se tornam um borrão na imagem de disparidade, o que se deve ao fato de haver pouca liberdade, não permitindo ao algoritmo o reconhecimento de saltos na disparidade, caracterizados pelas bordas das edificações.

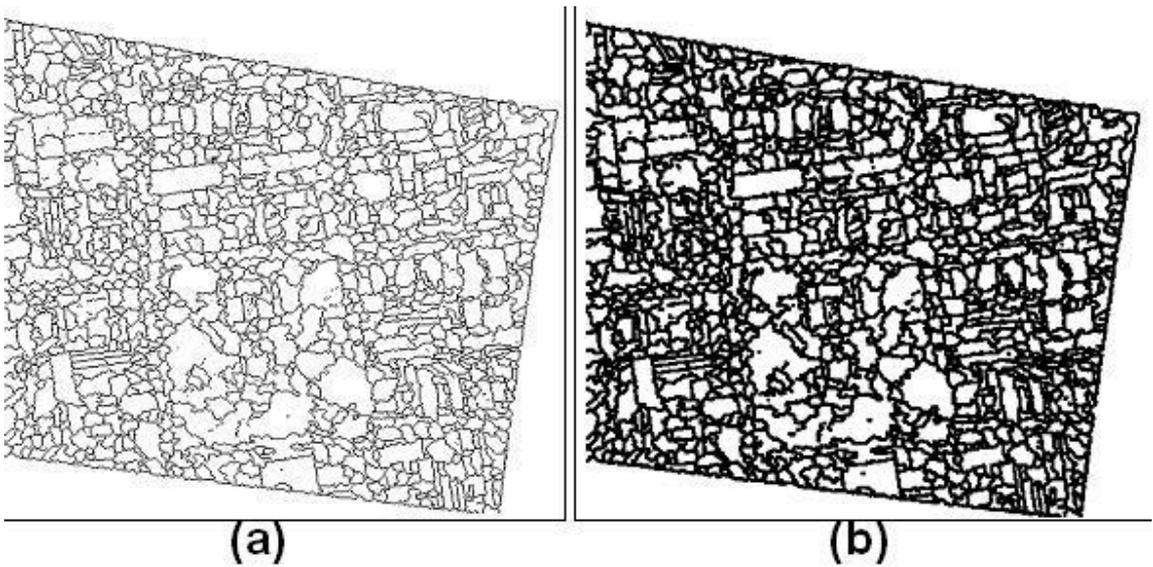


**Figura 38** – Disparidades com diferentes saltos no caminho mínimo, (a) salto simples, (b) salto duplo.

Para resolver este problema, foi utilizado uma liberdade de salto dupla (dois degraus para cima ou para baixo), como pode ser observado na Figura 39-a; o resultado dessa liberdade maior é a Figura 38-b. Pode-se perceber que houve uma melhora significativa no reconhecimento das bordas das edificações. A utilização da calibração *on-the-fly* visa justamente realçar as bordas através de um grande aumento na liberdade do caminho mínimo (Figura 39-b) justamente em áreas localizadas como bordas pelo algoritmo de segmentação (Figura 40). Pode-se perceber pela Figura 37-d que as bordas das edificações são preservadas neste método.



**Figura 39** – Programação dinâmica sobre o mapa de disparidade, (a) salto duplo, (b) salto seletivo pela calibração *on-the-fly*.



**Figura 40** – Bordas segmentadas, (a) borda simples, (b) borda ampliada.

## CAPÍTULO 5 – ANÁLISE E TESTES

### 5.1 – ANÁLISE DOS RESULTADOS

O projeto pode ser dividido em três fases, uma nova implementação do método de *watershed*, chamada de *watershed colorido*, uma metodologia de estereoscopia para fotografias aéreas, incluindo uma forma simples e eficaz de retificação, filtros com equalização de intensidade e programação dinâmica com deslocamento variável e, finalmente, a utilização do resultado da segmentação para calibrar o processo de estereoscopia. Todas estas fases apresentaram resultados interessantes, os quais serão expostos nesta seção através de análises métricas e visuais.

Na Figura 42 simula-se uma comparação entre 3 métodos de segmentação por *watershed*, baseado em informações geométricas [17] (CBMA), como área e volume, baseado em uma posterior união por cores (*cor a posteriori*) e com a utilização de segmentação por cor *on-the-fly* (*watershed colorido*). A análise comparativa destes três métodos, descritos no Capítulo 3, foi feita utilizando-se como referência uma segmentação manual (Figura 42-b). Basicamente, segmenta-se uma série de mais ou menos 40 objetos em 5 imagens diferentes (Figura 47-b); essa segmentação manual foi utilizada para avaliar metricamente os resultados dos métodos automáticos. A avaliação é feita comparando-se os objetos segmentados manualmente (objetos reais) com o resultado dos demais métodos, segundo a equação indicada na Figura-41, ou seja, a qualidade do método de segmentação é igual à percentagem de área do objeto real R mapeada pelo objeto segmentado S vezes a percentagem de área do objeto segmentado S mapeada pelo objeto real R:  $Q = (|R \cap S| / |R|) \times (|S \cap R| / |S|) = (|R \cap S|^2 / (|R| \times |S|))$ .

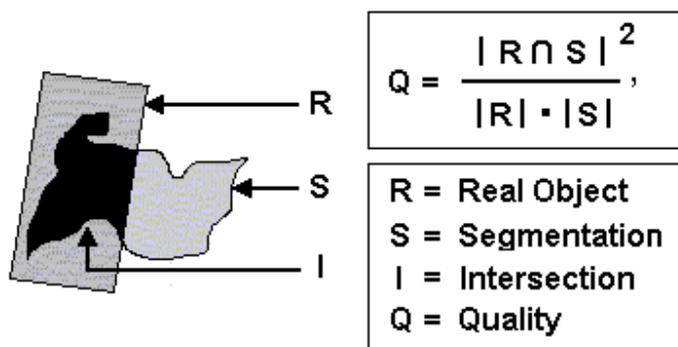
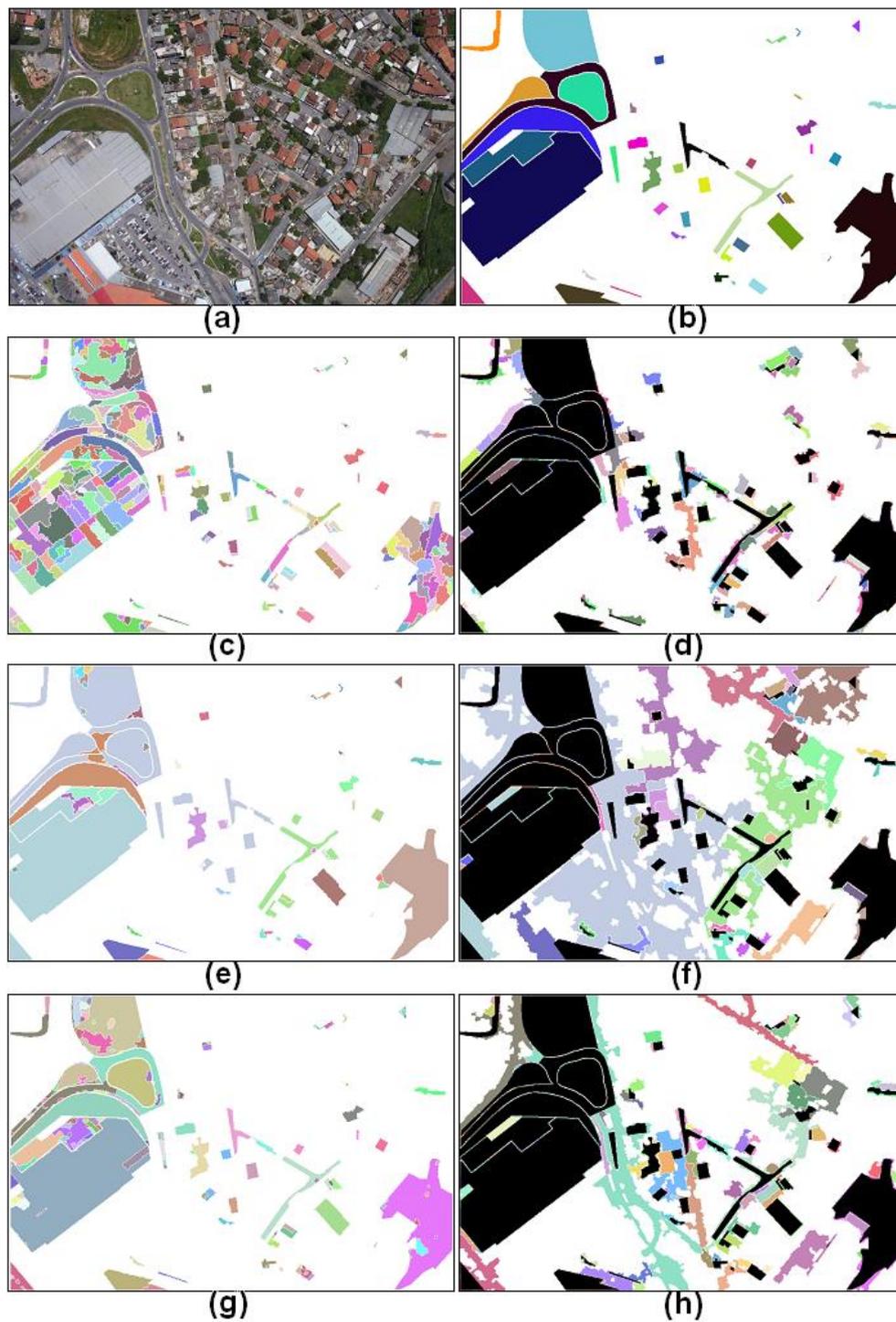
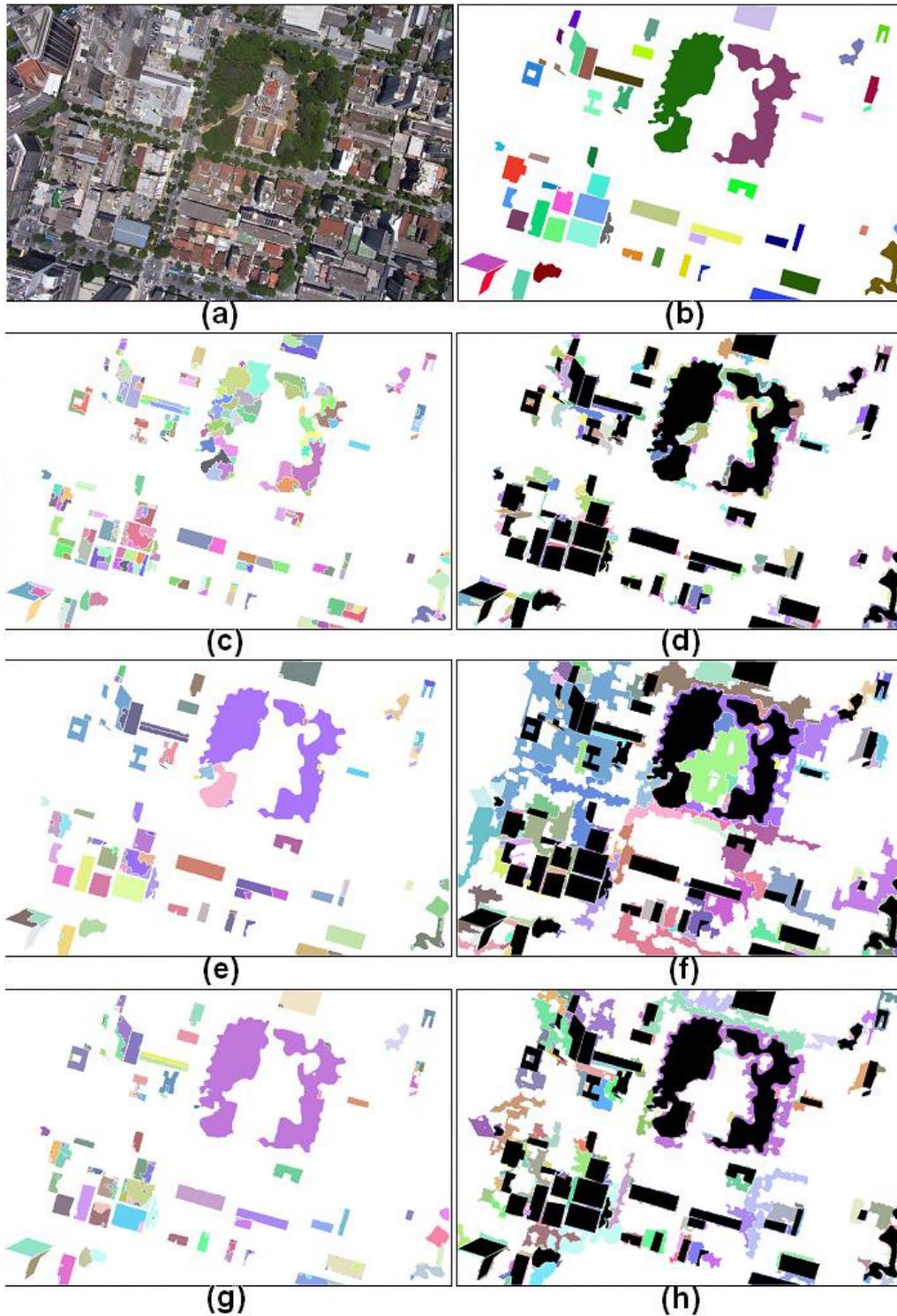


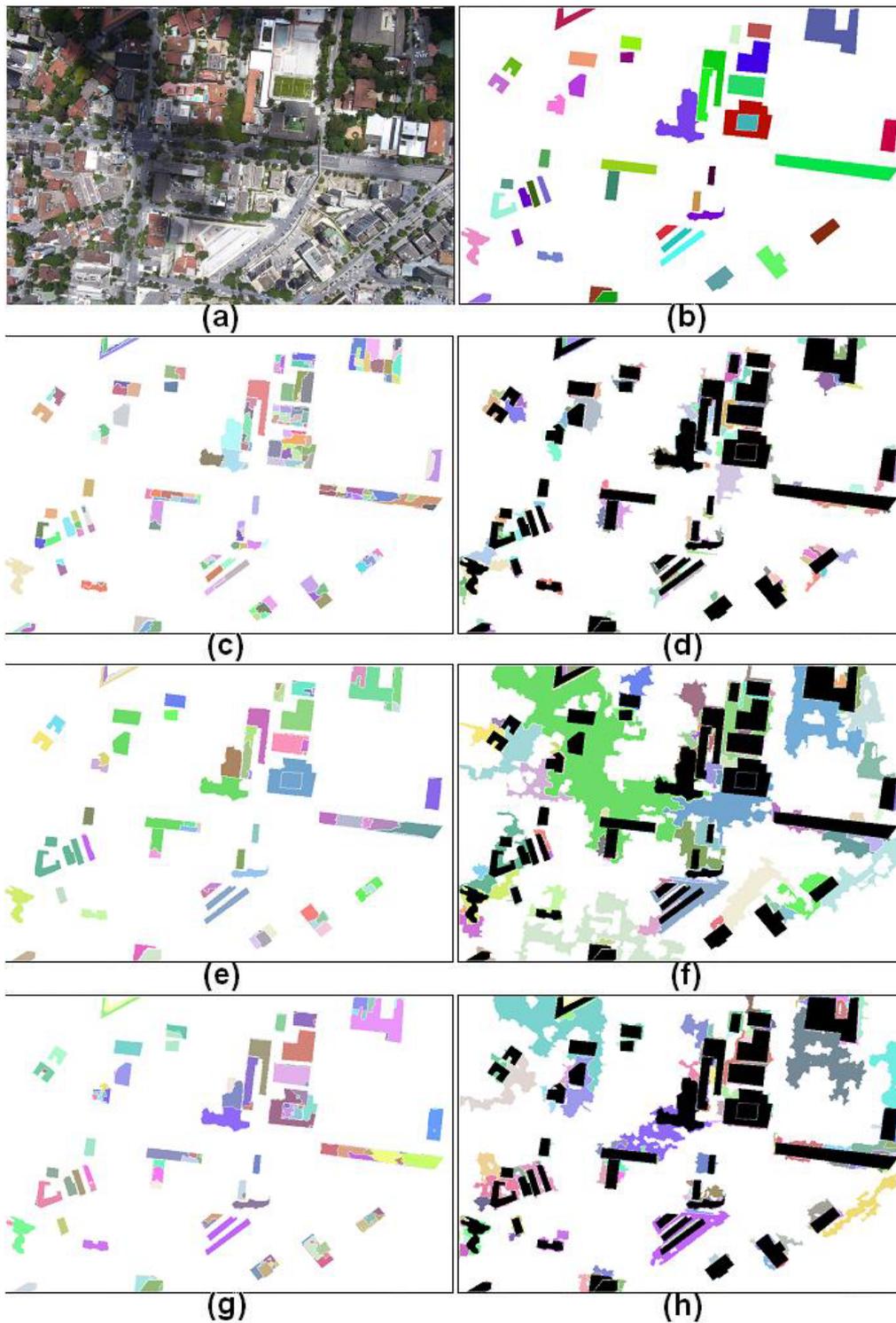
Figura 41 – Interseção de objetos.



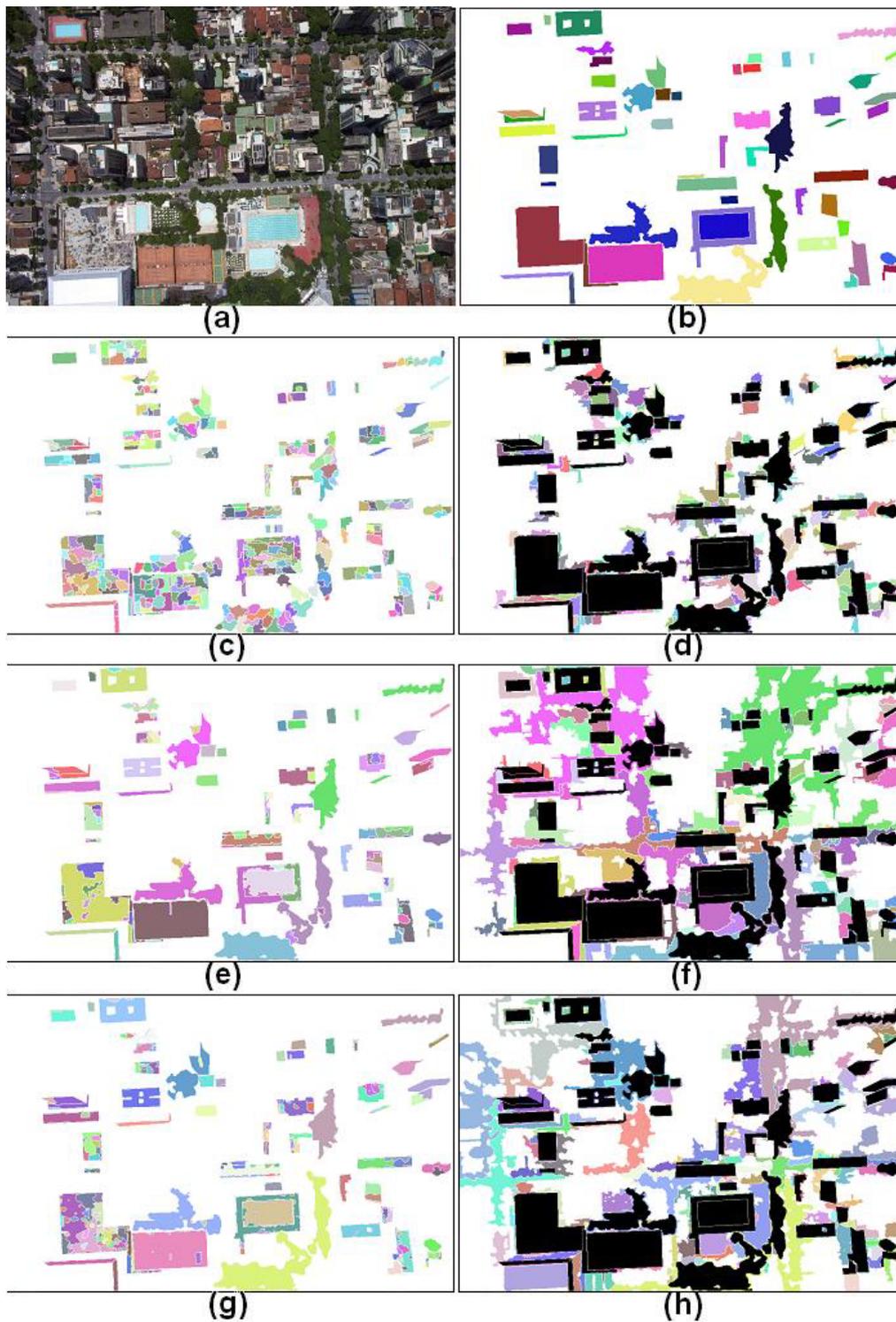
**Figura 42** – Comparação entre métodos de *watershed*, (a) imagem original, (b) objetos selecionados, (c) segmentação por CBMA, (d) resíduos da segmentação por CBMA, (e) utilização de *cor a posteriori*, (f) resíduos da aplicação de *cor a posteriori*, (g) segmentação com *watershed colorido*, (d) resíduos de *watershed colorido*.



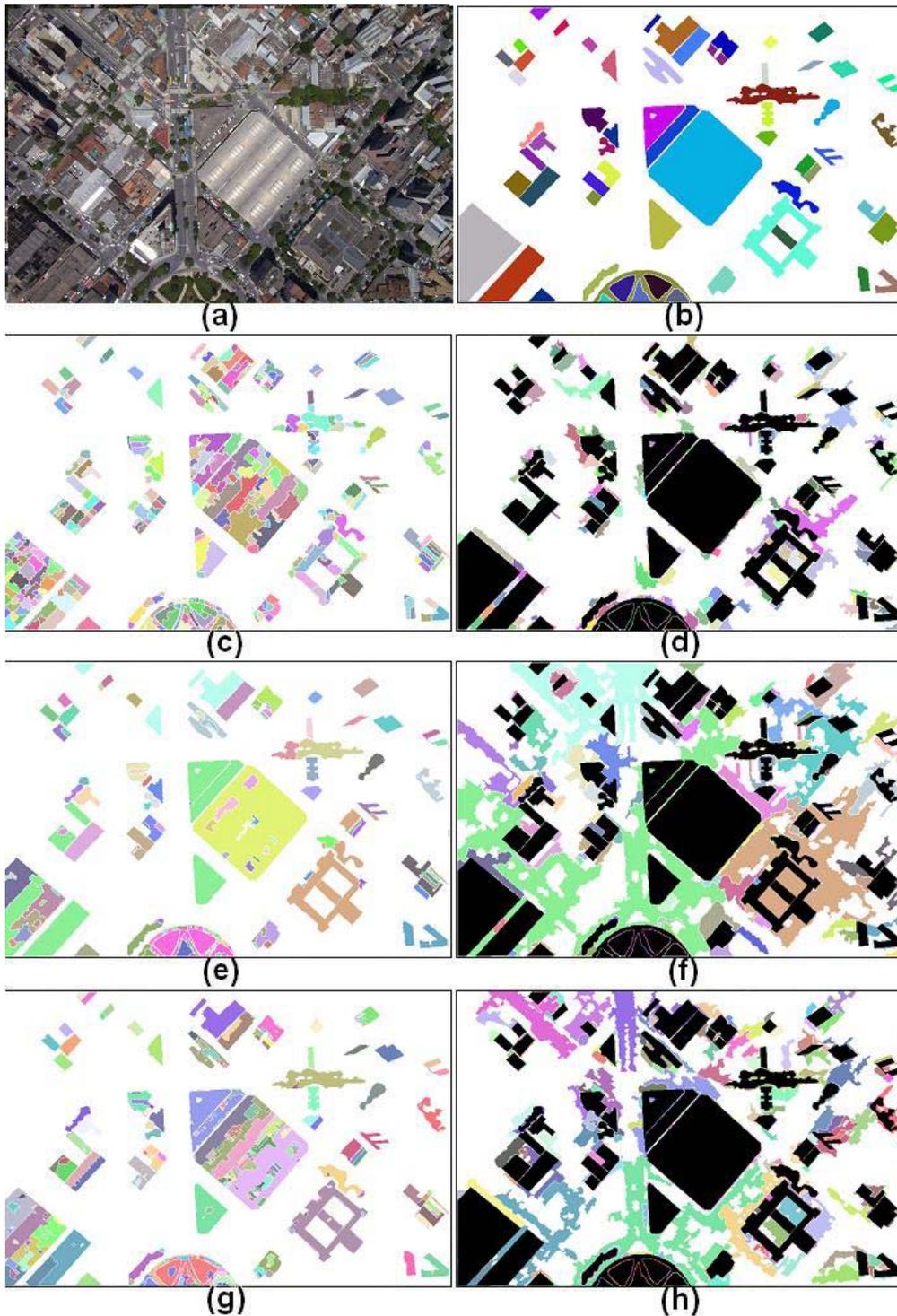
**Figura 43** – Comparação entre métodos de *watershed*, (a) imagem original, (b) objetos selecionados, (c) segmentação por CBMA, (d) resíduos da segmentação por CBMA, (e) utilização de *cor a posteriori*, (f) resíduos da aplicação de *cor a posteriori*, (g) segmentação com *watershed colorido*, (d) resíduos de *watershed colorido*.



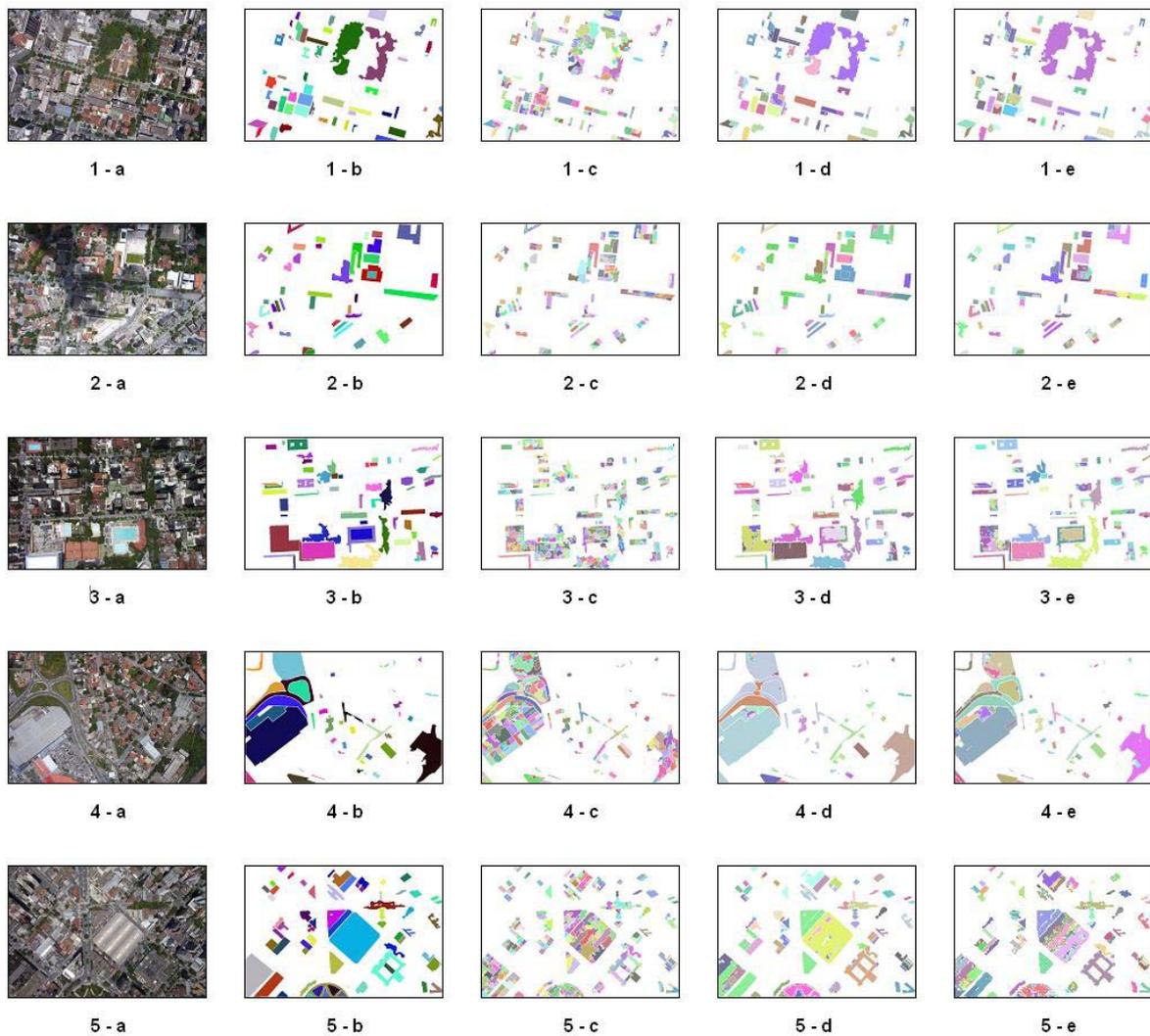
**Figura 44** – Comparação entre métodos de *watershed*, (a) imagem original, (b) objetos selecionados, (c) segmentação por CBMA, (d) resíduos da segmentação por CBMA, (e) utilização de *cor a posteriori*, (f) resíduos da aplicação de *cor a posteriori*, (g) segmentação com *watershed colorido*, (d) resíduos de *watershed colorido*.



**Figura 45** – Comparação entre métodos de *watershed*, (a) imagem original, (b) objetos selecionados, (c) segmentação por CBMA, (d) resíduos da segmentação por CBMA, (e) utilização de *cor a posteriori*, (f) resíduos da aplicação de *cor a posteriori*, (g) segmentação com *watershed colorido*, (d) resíduos de *watershed colorido*.



**Figura 46** – Comparação entre métodos de *watershed*, (a) imagem original, (b) objetos selecionados, (c) segmentação por CBMA, (d) resíduos da segmentação por CBMA, (e) utilização de *cor a posteriori*, (f) resíduos da aplicação de *cor a posteriori*, (g) segmentação com *watershed colorido*, (h)resíduos de *watershed colorido*.

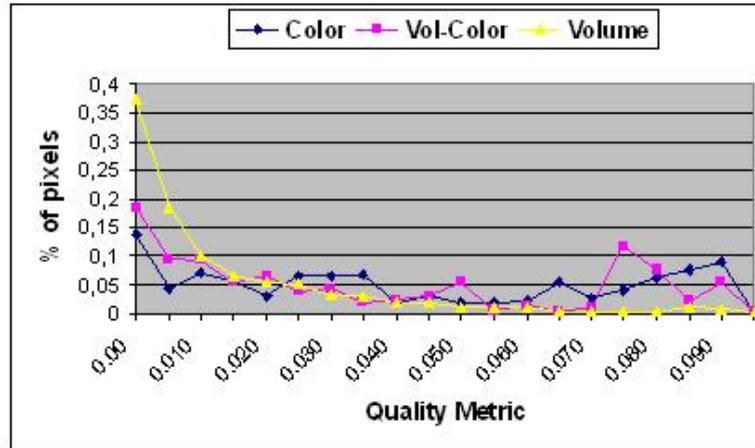


**Figura 47** – Comparação entre métodos de segmentação por *watershed*, (a) imagem original, (b) segmentação manual, (c) CBMA geométrico, (d) *cor a posteriori*, (e) *watershed colorido*.

Para melhor visualizar os resultados obtidos, eles foram divididos em duas categorias para cada método testado, segmentação interna dos objetos reais (Figura 42-c,e,g) e resíduos obtidos pela segmentação (Figura 42-d,f,h). Na Figura 42, pode-se perceber que o método de CBMA apresentou menos resíduos (Figura 42-d), porém segmentou demais os objetos reais (Figura 42-c). Já os métodos baseados em cor mantiveram uma coerência nos objetos reconhecidos (Figura 42-e,g), sendo que o método de *cor a posteriori* apresentou muito mais resíduos que o *watershed colorido*.

Foi realizada uma análise numérica dos resultados de testes feitos nas imagens da Figura 47 segundo a equação da Figura 41; pelo gráfico da Figura 48, pode-se perceber que o método de *watershed colorido* (Color) obteve os melhores resultados, seguido do método *cor a posteriori* (Volume) e do CBMA (Volume). As médias de acerto para os três métodos foram de 43%, 37% e 15% respectivamente. Como se pode ver pelo gráfico da Figura 48, o método baseado em cor apresenta uma distribuição de pontos concentrados nos valores mais elevados de acerto.

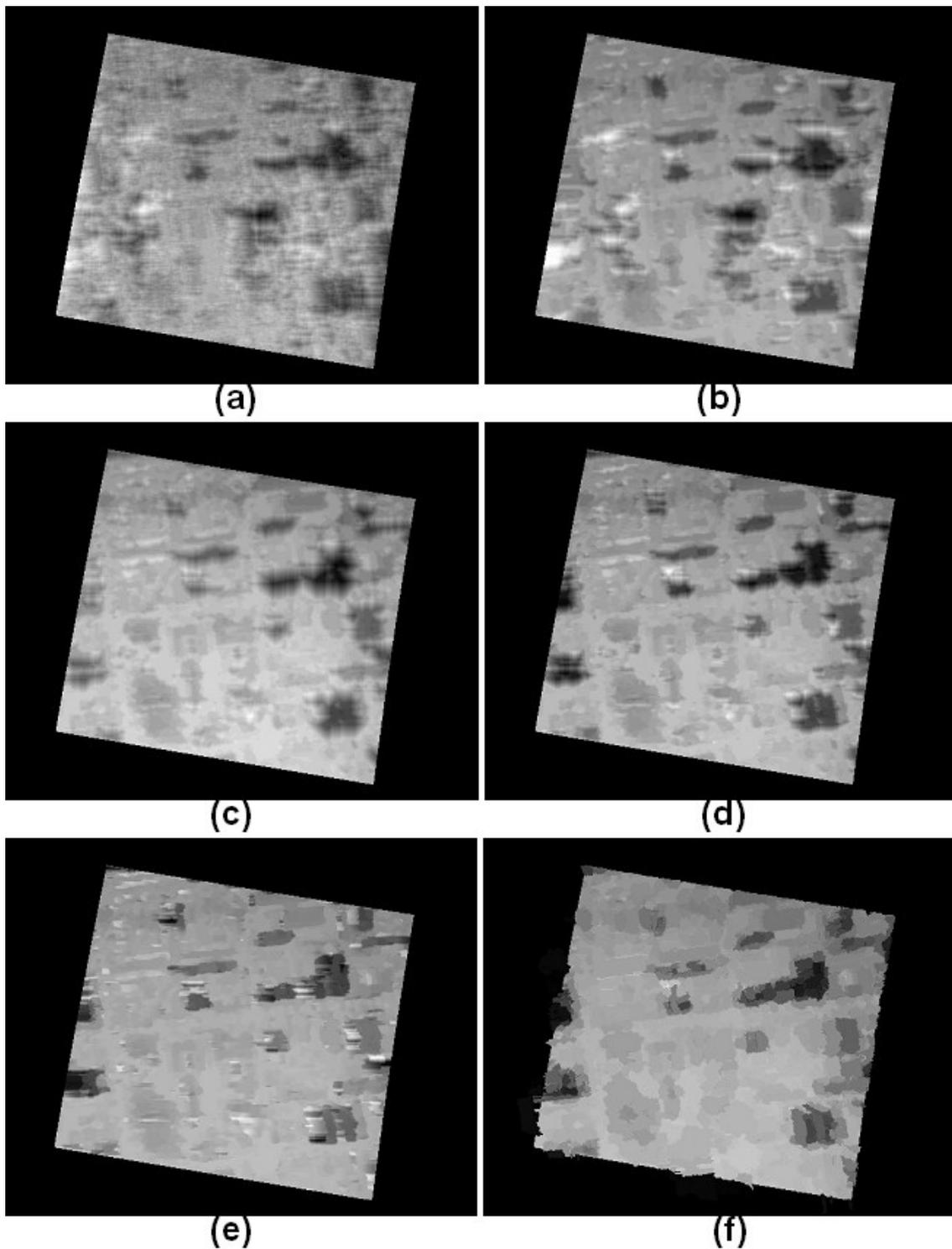
A análise dos resultados do processo de estereoscopia e calibração pela segmentação foi feita visualmente através da Figura 49. O primeiro resultado obtido (Figura 49-a) foi através de programação dinâmica com liberdade de apenas 1 degrau no caminho mínimo (Seção 4.3.1) e filtro de média unitário (Seção 4.3.2). A primeira atualização feita foi a utilização de um filtro de tamanho 9x9 (Seção 4.3.2) com distribuição progressiva de pesos (Figura 49-b). Visando reduzir a quantidade de erros do resultado, foi adicionado ao filtro uma equalização de intensidade (Figura 49-c). Neste ponto, a disparidade obtida possui poucos erros, porém não localiza bordas de edificações devido ao grande deslocamento necessário nas áreas de oclusão. Para resolver este problema, dobra-se a liberdade do caminho mínimo no mapa de disparidade (Figura 49-d). Mesmo neste caso, a borda continua com pouca definição. Para definitivamente localizar e demarcar as bordas da imagem, propõe-se dois métodos baseados na segmentação por *watershed colorido*, *on-the-fly* (Figura 49-e) e *a posteriori* (Figura 49-f); ambos os resultados realçaram definitivamente as bordas das edificações, porém o primeiro (49-e) apresentou alguns erros, a maioria em áreas de oclusão, e o segundo (49-f) eliminou pequenas características da disparidade, mantendo apenas as informações de bordas.



**Figura 48** – Resultado métrico dos algoritmos de *watershed*. Color representa o *watershed colorido*, Vol-Color a aplicação *a posteriori* do limiar de cor e Volume o CBMA.

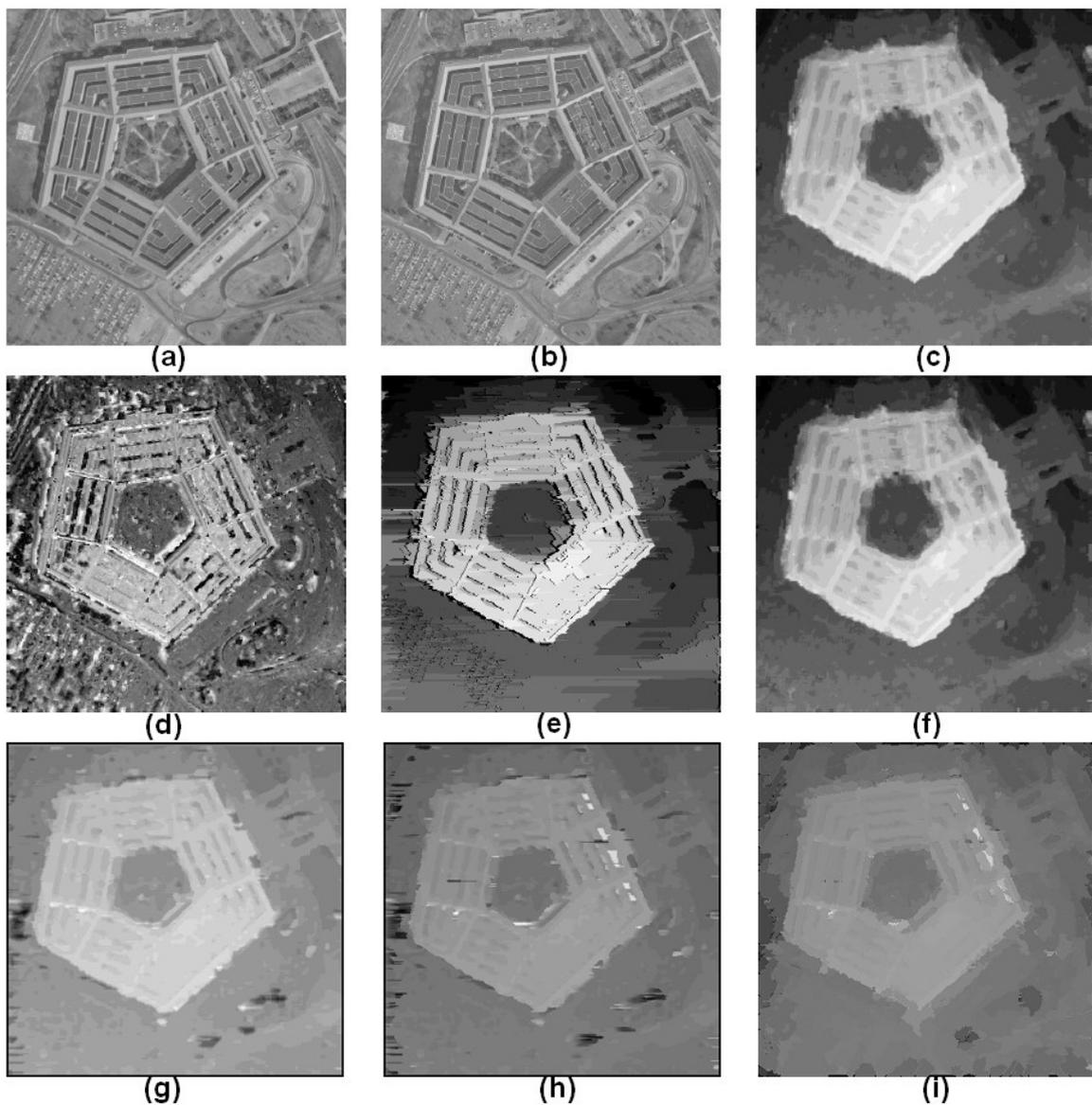
	Figura 20	Figura 22	Figura 21	Figura 23	Figura 24	Result
Color	0.62	0.39	0.42	0.42	0.30	0.43
Vol-Color	0.47	0.39	0.36	0.31	0.32	0.37
Volume	0.10	0.12	0.16	0.24	0.14	0.15

**Planilha 1** – Percentagem de acerto dos algoritmos de *watershed* para as imagens testadas.



**Figura 49** – Resultados de estereoscopia, (a) filtro unitário, (b) filtro 9x9, (c) filtro 9x9 com equalização, (d) filtro 9x9 com equalização e liberdade dupla na programação dinâmica, (e) filtro 9x9 com equalização e calibração *on-the-fly*, (f) filtro 9x9 com equalização e calibração *a posteriori*.

Para demonstrar que os métodos desta dissertação não se aplicam somente a regiões com grandes disparidades, compara-se as três variações propostas com quatro métodos existentes na literatura. Na Figura 50, utiliza-se um par estéreo bastante comum nas publicações de visão computacional. Em (a) e (b), tem-se o par original, na Figura 50-c tem-se o resultado obtido por Changming Sun (2002) [39], em (d) a imagem referente a Roy (1999) [32], em (e) tem-se o método de Cox (1996) [6] e em (f) o resultado obtido por Sun (1997) [38]. Em (g) tem-se o resultado do método de caminho mínimo com deslocamento duplo, em (h) o resultado da estereoscopia com calibração *on-the-fly* e em (i) o resultado *a posteriori*.



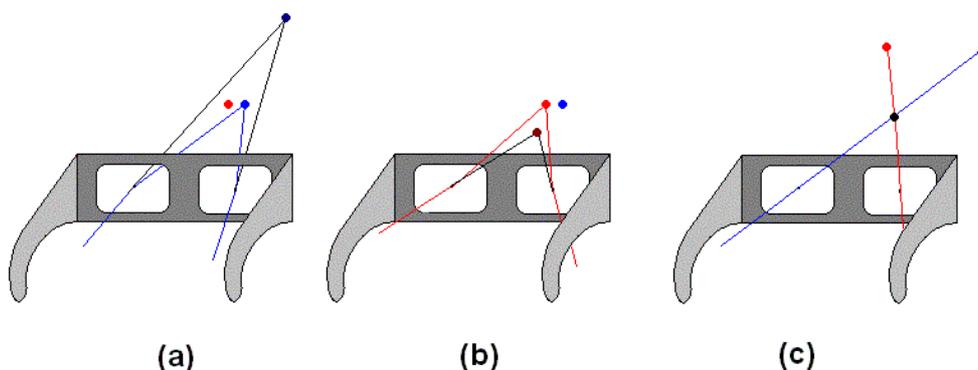
**Figura 50** – Resultados de estereoscopia, (a) e (b) par estereo, (c) (Sun, 2002) [39] (d) (Roy, 1999) [32], (e) (Cox, 1996) [33], (f) (Sun, 1997) [38], (g) duplo deslocamento, (h) calibração *on-the-fly*, (i) calibração *a posteriori*.

O método proposto por (Roy, 1999) [32], nitidamente mantém informações de cor da imagem original na imagem de disparidade gerada, além de apresentar muito ruído. O resultado de (Cox, 1996) [6], localizou bem as bordas da edificação, porém apresentou ruídos e sombreamento das bordas no mapa de disparidade, resultando em depressões não existentes. Os resultados propostos por Sun (2002 [39], 1997 [38]) quase não diferem entre si, apresentaram pouco ruído na imagem de disparidade, porém não preservaram as bordas das edificações. Os resultados propostos neste projeto respeitaram mais os limites dos artefatos que os resultados de Sun, sendo que as imagens (Figura 50 - g,h) apresentaram mais ruídos que as letras (Figura 50 - c,f - Sun).

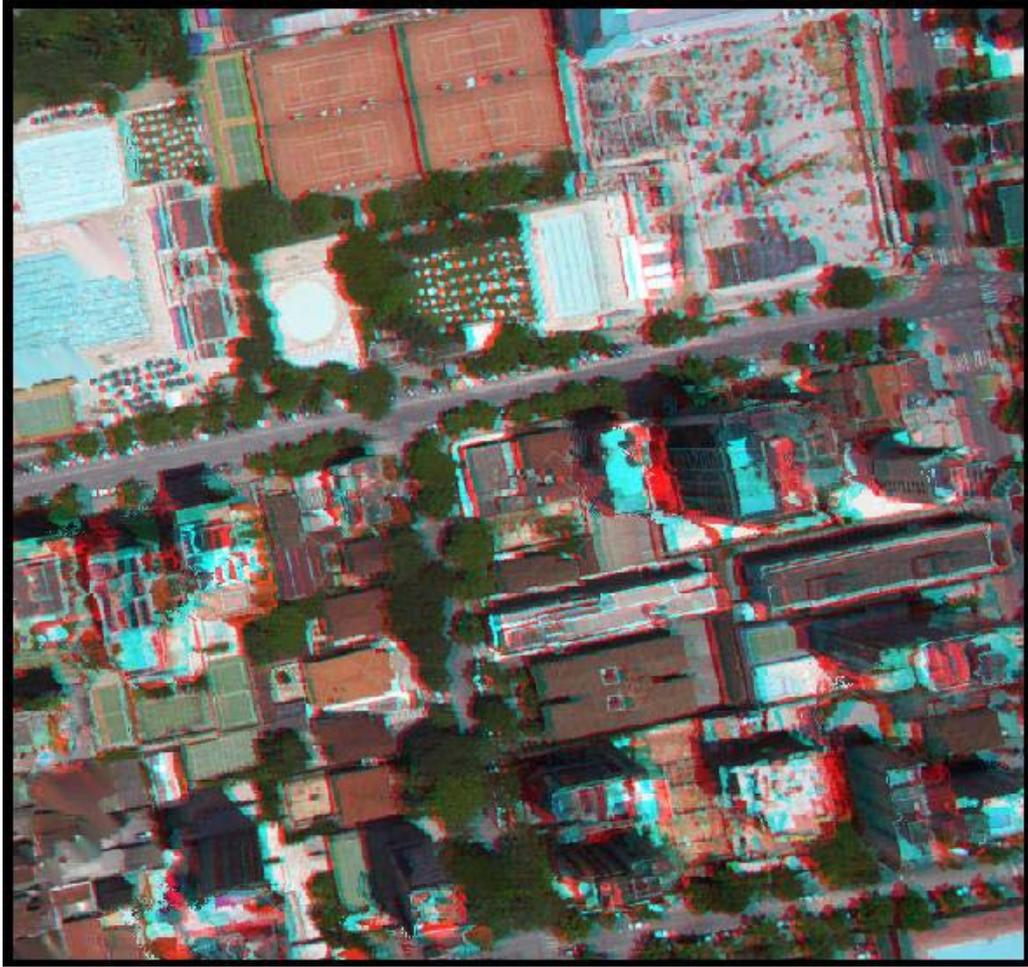
## 5.2 – VISUALIZAÇÃO DOS RESULTADOS

Nesta seção, apresenta-se algumas imagens sobre as quais foram testados os algoritmos. A visualização dos resultados é feita através de sobreposição dos mesmos às imagens originais e do uso de simulação 3D através de óculos *chromadepth* e *anaglyphy*. A metodologia do *chromadepth* se baseia em lentes que alteram a direção dos raios de luz em diferentes inclinações, dependendo da frequência do mesmo (Figura 51-a,b), já o óculos *anaglyphy* se baseia em uma filtragem de raios de determinadas frequências (Figura 51-c). Ambos os óculos fazem com que os olhos se cruzem em posições desfocadas com a posição do monitor, garantindo a simulação da visão de 3 dimensões. Apesar do efeito alcançado pelo *chromadepth* ser muito menor que o do *anaglyphy*, ele é útil no sentido de que permite uma análise da imagem sem um óculos específico, o que não ocorre com o *anaglyphy*, o qual embaça totalmente a imagem original.

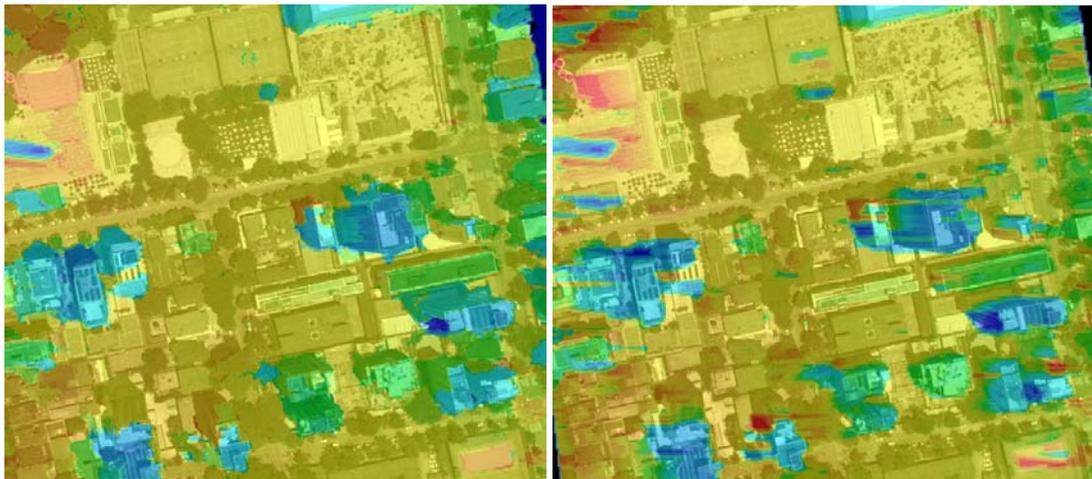
A visualização das imagens em *anaglyphy* (Figura 52 e 54) deverá ser feita com óculos de lentes azul e vermelha, orientadas de forma às edificações saltarem sobre o solo das imagens. Pode-se perceber pelas imagens *chromadepth* (Figura 53 e 55) aquilo que já havia sido informado nas Seções 4.4 e 5.1, o resultado do método de calibração *on-the-fly* preserva mais as informações de disparidade, porém apresenta mais ruídos que o método *a posteriori*, sendo que eles se concentram nas áreas de oclusão.



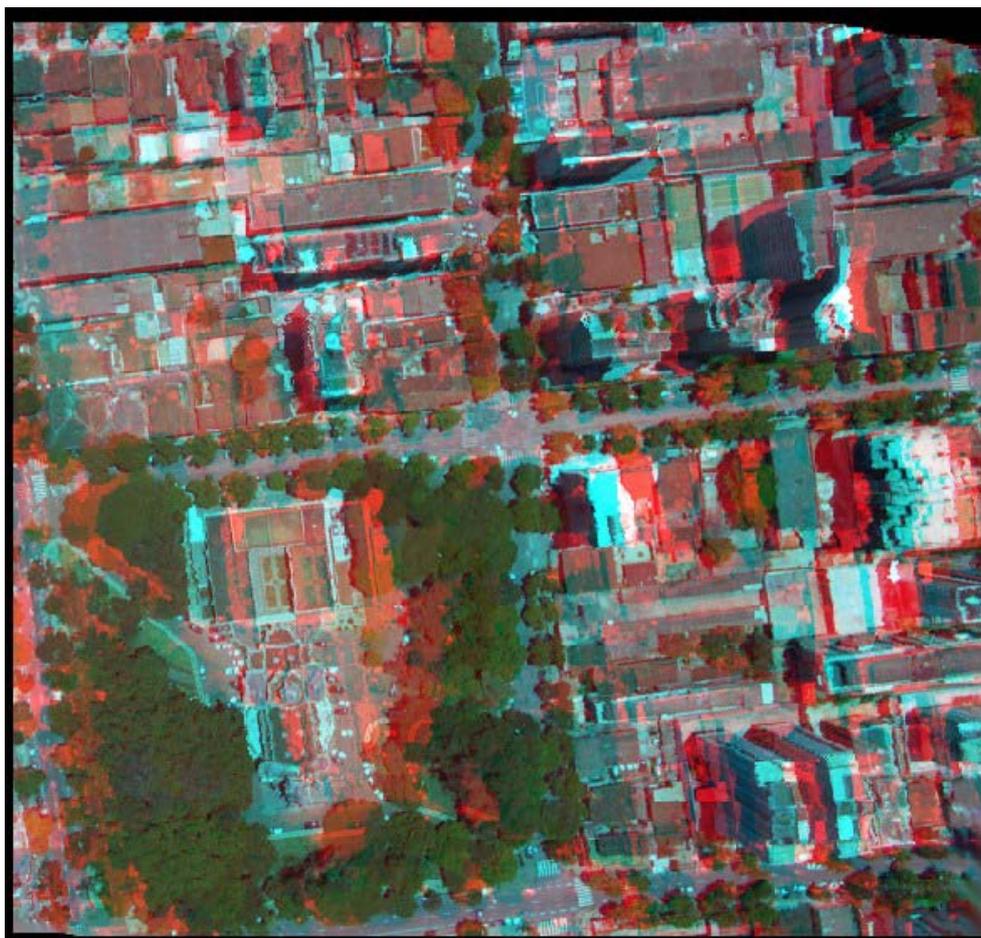
**Figura 51** – Óculos 3D, (a) e (b) distorção na cor azul e vermelha com o óculos *chromadepth*, (c) filtragem das cores azul e vermelha no *anaglyphy*.



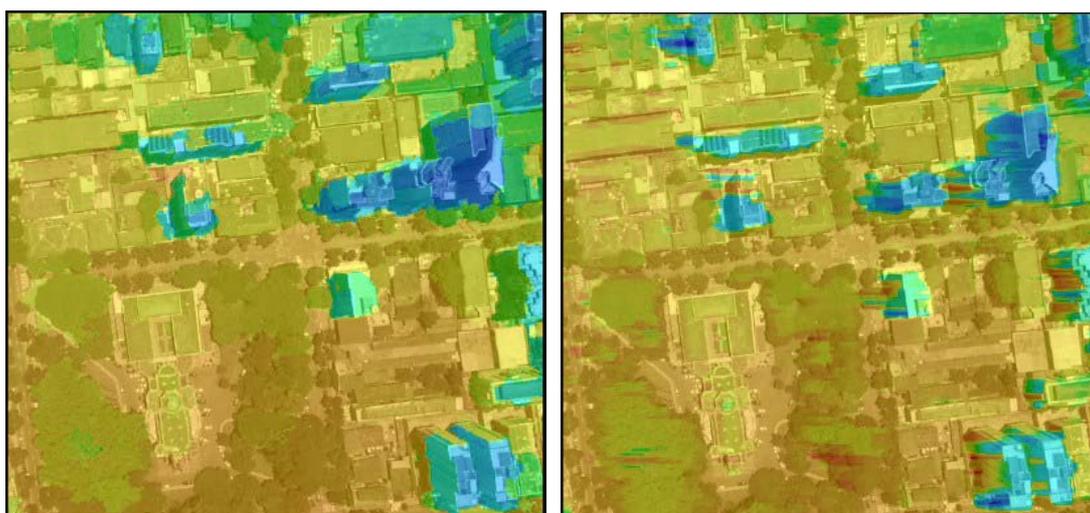
**Figura 52** – Visualização *anaglyphy* com resultado de calibração *a posteriori*.



**Figura 53** – Visualização *chromadepth*, com resultado *a posteriori*, à esquerda, e *on-the-fly*, à direita.



**Figura 54** – Visualização *anaglyphy* com resultado de calibração *a posteriori*.



**Figura 55** – Visualização *chromadepth*, com resultado *a posteriori*, à esquerda, e *on-the-fly*, à direita.

## CAPÍTULO 6 – CONCLUSÃO

---

O objetivo desta dissertação é fornecer uma metodologia para a reconstrução de cenas urbanas, através da utilização de características específicas de fotos aéreas visando solucionar problemas realçados pelas técnicas utilizadas. Propõe-se uma sequência de passos relacionados com áreas de processamento digital de imagens e visão computacional. Em todas as fases do processo foram adotados métodos que utilizam-se características das imagens, como cor, bordas e intensidade, para aprimorar o resultado final. Para calibrar o processo de estereoscopia, utiliza-se informações de bordas da imagem, localizadas através da informação de cor (*watershed colorido*); ainda, utilizam-se filtros com controle de intensidade, devido ao fato de as sombras se alterarem muito em pares estereo aéreos. Os resultados obtidos (Figura 49) reconheceram bem as edificações e foram capazes de preservar suas bordas. Mesmo em casos de pouca oclusão, o algoritmo apresentou um resultado bastante positivo, quando comparado com técnicas publicadas na literatura de visão computacional (Figura 50).

Mesmo sem a calibração pela segmentação, o resultado obtido com maior liberdade na programação dinâmica foi bastante satisfatório em áreas de pouca oclusão (Figura 50-g), porém borrando ligeiramente as bordas em imagens com grandes áreas ocultas (Figura 49-d). Os resultados de calibração *a posteriori* e *on-the-fly* preservaram as bordas mesmo em áreas com grandes oclusões, porém o resultado do método *a posteriori* eliminou pequenas características de disparidade e o método *on-the-fly* apresentou alguns erros no resultado. Apesar de os resultados terem sido bastante satisfatórios, propomos como trabalho futuro uma correção posterior à aplicação do método *on-the-fly* para eliminar as imperfeições resultantes deste processo; isto poderá ser feito utilizando inclusive o resultado obtido pelo *watershed colorido*.

O processo apresentado nesta tese pode ser utilizado em diversas aplicações relacionadas com urbanismo, agricultura, mineração e meio ambiente. Pode-se citar diversas aplicações, incluindo:

- restituição auto ou semi-automática de feições urbanas
- montagem de DTM, DEM e DSM urbanos
- cálculo de área verde real
- indicador sócio-econômico
- controle sobre crescimento vertical
- controle sobre fluxo de ar em regiões centrais
- controle sobre densidade populacional
- controle sobre desenvolvimento das plantações
- controle de pragas
- controle sobre cavas de mineração
- controle sobre impacto ambiental

## REFERÊNCIAS BIBLIOGRÁFICAS

---

- [1] A. Broadhurst, T. Drummond. A probabilistic framework for space carving. In Proc 8<sup>th</sup> Int. Conf. On Computer Vision, 2001.
- [2] A. Mehnert et, al., An improved seeded region growing algorithm. Pattern Recognition Letters, 1997.
- [4] A. Huguet, R. Carceroni and A. Albuquerque. Towards Automatic 3D Reconstruction of Urban Scenes from Low-Altitude Aerial Images. ICIAP (2003).
- [5] Berthold K.P. Horn. Recovering Baseline and Orientation from Essential Matrix. January 1990.
- [6] Cox, I., Hingorani, S., Rao, S., and Maggs, B. (1996). A maximum likelihood stereo algorithm. Computer Vision and Image Understanding, 63(3): 542-567.
- [7] C. Silva and J. Santos Victor. Intrinsic images for dense stereo matching with occlusions. In Proc. 6<sup>th</sup> European Conference on Computer Vision, 2000.
- [8] D. Liebowitz and A. Zisserman. Metric Rectification for Perspective Images of Planes. In Proceedings of the IEEE Conf. On Comp. Vision and Pat. Recog. Pages 482-488, Santa Barbara, California, June 1998. IEEE Computer Society.
- [9] F. Meyer. Un algorithme optimal de ligne de partage des eaux. VIII Congres de Reconnaissance de Forme et d'Intelligence Artificielle. Lyon, France, 1991.
- [10] F. Meyer and S. Beucher, Morphological segmentation. Journal of Visual Communication and Image Representation, 1990.
- [11] H.C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. Nature, 293:133 – 135, Sept 1981.
- [12] H. Tek et. Al. Volumetric segmentation of medical images by three-dimensional bubbles. Computer Vision and Machine Understanding, 1997.
- [13] J. Isaac et. Al. Sorting by Address Calculation. Journal of the ACM, 1994.
- [14] K. N. Kutulakos. Approximate N-View stereo. In Proc. European Conf on Computer Vision. 2000.
- [15] K. N. Kutulakos. A theory of shape by space carving. Int. J. Computer Vision.2000 Marr Prize Special Issue.

- [16] L. Vincent, P. Soile. Watershed on Digital Spaces: An Efficient Solution Based on Immersion Simulation. *IEEE Trans. Of Pattern Analysis and Machine Intelligence*. V13, n. 6, p. 593-598, 1991.
- [17] M. C. Andrade. Um algoritmo topológico de segmentação de imagens por atributos. Tese de Doutorado, Departamento de Ciência da Computação – DCC, Universidade Federal de Minas Gerais – UFMG, 1998.
- [18] M. C. Andrade et. Al. Segmentation of microscopic images by flooding simulation: a catchment basins merging algorithm. *Proceedings of SPIE Nonlinear Image Processing VIII*, San Jose, USA, 1997.
- [19] M. Kass et. Al. Snakes Active Contour models. *International Journal of Computer Vision*, 1988.
- [20] O. D. Faugeras and R. Keriven. Variational principles, surface evolution, pde's, level set methods and the stereo problem. *IEEE Trans. Image Processing*, 1998.
- [21] O. Faugeras. *Three-Dimensional Computer Vision: a Geometric Viewpoint*. MIT Press, 1993.
- [23] P. E. Debevec, C. J. Taylor. Modeling and rendering architecture from photographs. In *Proc SIGGRAPH 1996*.
- [24] P. Fua. From multiple stereo views to multiple 3-D surfaces. *Int. J. Computer Vision*, 1997.
- [25] P. N. Belhumeur. A bayesian approach to binocular stereopsis, *Int. J. on Computer Vision*, 1996.
- [26] Q. Chen and G. Medioni. A volumetric stereo matching methodÇ Application to image-based modeling. In *Proc. Computer Vision and Pattern Recognition Conf.* 1999.
- [27] Q. T. Luong and O. Faugeras. Self-calibration of a moving camera from point correspondences and fundamental matrices. *The International Journal of Computer Vision*, 1997.
- [28] R. Adams et. Al. Seeded region growind. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 1994.
- [29] R. Azeliski, S. Avidan, and P. Anandan. Layer extraction from multiple images containing reflections and transparency. In *Proc. Computer Vision and Pattern Recognition Conf.* 2000.
- [30] R.Y. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4):323-344, Aug. 1987.

- [31] Richard I. Hartley. In Defence of the 8-point Algorithm. GE-Corporate Research and Development, Schenectady, NY, 12309.
- [32] Roy, S. (1999). Stereo without epipolar lines: A maximum-flow formulation. *International Journal of Computer Vision*, 34(2/3):147-161.
- [33] Roy, S. and Cox, I. J. (1998). A maximum-flow formulation of the N-camera stereo correspondence problem. In *Proceedings of International Conference on Computer Vision*, pages 492-499, Bombay, India. IEEE.
- [34] Roy Sebastien, Cox J. Ingemar and Jean Meunier. Cylindrical Retification to Minimize Epipolar Distortion. *IEEE. Proc. of Int. Conf. On Computer Vision and Pattern Recognition*. Puerto Rico, June 1997, p. 393-399.
- [35] S. Bognoux. From projective to euclidean space under any practical situation, a criticism of self-calibration. In *Proceedings of the 6<sup>th</sup> International Conference on Computer Vision*, pp 790-796, Jan. 1998.
- [36] S. J. Maybank and O. D. Faugeras. A theory of self-calibration of a moving camera. *The International Journal of Computer Vision*, 1992.
- [37] S. M. Seitz and C. R. Dyer. Photorealistic scene reconstruction by voxel coloring. *The International Journal of Computer Vision*, 1999.
- [38] Sun, C. (1997). A fast stereo matching method. In *Digital Image Computing: Techniques and Application*, pages 95-100, Massey University, Auckland, New Zealand.
- [39] Sun, C. Fast Stereo Matching Using Rectangular Subregioning and 3D Maximum-Surface Techniques. *International Journal of Computer Vision*. Vol. 47, no 1/2/3, pp 99-117, May 2002.
- [40] Y. Yu and J. Malik. Recovering photometric properties of architectural scenes from photographs. In *Proc. SIGGRAPH 1998*.
- [41] [www.istar.com](http://www.istar.com)
- [42] [www.belohorizonte.com.br](http://www.belohorizonte.com.br), [www.infostrata.com.br](http://www.infostrata.com.br)
- [43] [www.enri.com](http://www.enri.com)
- [44] [www.erdas.com](http://www.erdas.com), [www.img.com.br](http://www.img.com.br)
- [45] [www.leicageosystems.com](http://www.leicageosystems.com)
- [46] [www.orbisat.com.br](http://www.orbisat.com.br)

[47] [www.earthgis.com](http://www.earthgis.com)

[48] William H. Press, Brian P. Flannery, Saul A. Teukolsky and William T. Vetterling. Numerical Recipes in C: The Art of Scientific Computing. Cambridge University Press, 1988.

[49] Zhengyou Zhang. A Flexible New Technique for Camera Calibration. Dec, 1998.