UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
COORDENAÇÃO DE PÓS-GRADUAÇÃO EM INFORMÁTICA

Dissertação de Mestrado

Algoritmos para Composição Automática de Fotografias

Claudio S. Vasconcelos da C. Cavalcanti

Campina Grande
Julho de 2007

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
COORDENAÇÃO DE PÓS-GRADUAÇÃO EM INFORMÁTICA

Algoritmos para Composição Automática de Fotografias

Claudio S. Vasconcelos da C. Cavalcanti

Dissertação submetida à Coordenação do Curso de Pós-Graduação em Ciência da Computação do Centro de Engenharia Elétrica e Informática da Universidade Federal de Campina Grande – Campus I como parte dos requisitos necessários para obtenção do grau de Mestre em Ciência da Computação (MSc).

Área de Concentração: Ciência da Computação
Linha de Pesquisa: Modelos Computacionais e Cognitivos

Herman Martins Gomes
Orientador

Campina Grande
Julho de 2007

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
COORDENAÇÃO DE PÓS-GRADUAÇÃO EM INFORMÁTICA

Algoritmos para Composição Automática de Fotografias

Claudio S. Vasconcelos da C. Cavalcanti

Dissertação submetida à Coordenação do Curso de Pós-Graduação em Ciência da Computação do Centro de Engenharia Elétrica e Informática da Universidade Federal de Campina Grande – Campus I como parte dos requisitos necessários para obtenção do grau de Mestre em Ciência da Computação (MSc).

Área de Concentração: Ciência da Computação
Linha de Pesquisa: Modelos Computacionais e Cognitivos

Herman Martins Gomes
Orientador

Campina Grande
Julho de 2007

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL DA UFCG

C376a Cavalcanti, Claudio S. Vasconcelos da C.
2007 Algoritmos para composição automática de fotografias /
Claudio S. Vasconcelos da C. Cavalcanti. – Campina Grande, 2007
137fs.: il. color.

Dissertação (Mestrado em Ciência da Computação) -
Universidade Federal de Campina Grande,
Centro de Engenharia Elétrica e Informática.

Referências

Orientador: Herman Martins Gomes.

1. Processamento de Imagens 2. Composição de Fotografias 3.
Automação de Processos 4. Visão Computacional I–
Título CDU
004.383.5(043)

"ALGORITMOS PARA COMPOSIÇÃO AUTOMÁTICA DE FOTOGRAFIAS"

CLÁUDIO SEBASTIÃO VASCONCELOS DA CUNHA CAVALCANTI

DISSERTAÇÃO APROVADA EM 30.07.2007

PROF. HERMAN MARTINS GOMES, Ph.D
Orientador

PROF. JOSÉ EUSTÁQUIO RANGEL DE QUEIROZ , D.Sc
Examinador

PROF. GEORGE DARMITON DA CUNHA CAVALCANTI, Dr.
Examinador

CAMPINA GRANDE – PB

Dissertação de Mestrado sob o título "*Algoritmos para Composição Automática de Fotografias*", defendida por Claudio S. Vasconcelos da C. Cavalcanti e aprovada em Julho de 2007, em Campina Grande, Estado da Paraíba, pela banca examinadora constituída pelos doutores:

Prof. Ph.D. Herman Martins Gomes

DSC / CEEI / UFCG

Orientador

Prof. Dr. José Eustáquio Rangel de Queiroz

DSC / CEEI / UFCG

Examinador

Prof. Dr. George Darmiton da Cunha Cavalcanti

CIn / UFPE

Examinador

Agradecimentos

A Deus, por ter me dado forças para realizar este trabalhoo alem de ter me tirado do fundo do rio!

À minha família: pais, avós, irmão e irmã, cunhada, cunhados e sobrinhos por terem me dado todo o apoio que precisei enquanto em Campina Grande, me ensinando a focar em solucoes, nao em problemas.

À Laura, por ter aceitado enfrentar comigo não só este como todos os outros desafios ainda porvir o qual espero comemorar sempre ao seu lado.

Aos amigos de Recife, por sempre torcerem por mim na distância e sempre me acolherem com a mesma alegria quando do retorno. Aos professores e colegas da UPE/Poli, em particular o Prof. Carlos, por terem sempre torcido por mim neste desafio e fazendo-me acreditar no sucesso. Aos colegas da IDM, em particular Diogo e Dickson, pelo apoio dado nos primeiros meses em Campina Grande.

Ao professor Herman, que orientou este trabalho com paciência e dedicação.

Aos amigos de Campina Grande, especialmente os integrantes e ex-integrantes do projeto iPhotoBot (Em particular Eanes, o mestre Latex, Luana, profa. Luciana e os outros que omitirei mas devido à falta de espaço) pela ajuda nas implementações, troca de idéias, incentivo e momentos de descontração.

Aos membros da banca examinadora, pelas críticas e sugestões que contribuíram para o enriquecimento deste trabalho.

A todos que fazem a COPIN, dentre professores e funcionários, os quais nunca negaram auxílio mesmo em momentos difíceis.s. A HP, que apoiou financeiramente parte deste trabalho.

Enfim, para *.* que de alguma forma contribuíram para a conclusão deste trabalho.

O meu muito obrigado!

Resumo

Além de ser uma das mais populares formas de arte, a fotografia também é uma forma de lazer e ferramenta de trabalho. Com a redução dos preços e a conseqüente popularização dos equipamentos e acessórios necessários à fotografia, especialmente o preço das câmeras digitais, é crescente o interesse por novos algoritmos e ferramentas que favoreçam a captura de imagens com maior qualidade. Diante do exposto, a presente dissertação objetivou a proposição e desenvolvimento de algoritmos capazes de detectar e corrigir falhas na composição fotográfica. As regras de composição fotográfica, em geral, são heurísticas utilizadas por fotógrafos que se difundiram a ponto de serem denominadas de "regras". Mesmo não sendo consenso entre os fotógrafos, é possível que a implementação destas regras possa levar um fotógrafo amador, sem conhecimento prévio de fotografia, a produzir fotografias de alta qualidade e teor profissional. Neste trabalho são propostas duas alternativas para a correção da composição: um método para correção *on-line*, no qual a foto final só é obtida após satisfeitas algumas condições de qualidade, e outro para a correção *off-line*, o qual classifica (ou modifica) a imagem *a posteriori*. Para tanto, são utilizados algoritmos destinados à detecção e correção de problemas no posicionamento do tema. Os resultados foram avaliados em dois experimentos. No primeiro experimento, os usuários concordaram em até 65% com os resultados obtidos pelo sistema, através de uma análise subjetiva. No segundo experimento, foi mostrado como é possível, utilizando-se apenas uma câmera *Pan-Tilt-Zoom* (câmera dotada de três graus de liberdade sendo dois de rotação e um do campo de visão), localizar e fotografar pessoas em um determinado ambiente a partir das regras de composição desenvolvidas.

Abstract

Besides being one of the most popular forms of art, photography is often used for a wide variety of purposes, including professional and entertainment ones. Nowadays, since cameras (specially digital ones) are less expensive and more popular, there is an increasing need for tools to help photographers (both amateurs and professionals) to obtain photographs of better quality. Within this context, algorithms for detection and correction of errors on the composition of a photograph are proposed in this work. Photographic composition rules are heuristics used by photographers, which became so widespread that they are now also known as "rules". Photographers, however, are not unanimous about the use of some of those rules. Despite of that it is possible that the use of photographic composition rules can improve the quality of amateur photographs, leveling them to a professional standard. In this dissertation, two approaches are proposed to automate composition rules: an on-line method, in which a picture is only taken when a number of conditions is satisfied; and an off-line method, which classifiers or corrects the image after it has been acquired. Hence, algorithms for detecting and correcting problems on subject positioning are used. Two experiments were used to evaluate the performance of the system. The first one shows that users agree with the correction performed on 65% of the photographs, through a subjective analysis. By using only a Pan-Tilt-Zoom camera and the composition rules implemented in this work, the second experiment shows how to locate and photograph human subjects in a given environment.

Conteúdo

Lista de Tabelas

Lista de Figuras

Lista de Algoritmos

Lista de Siglas e Abreviaturas

- AI: *Artificial Intelligence*
- API: *Application Programming Interface*
- CCD: *Charge-Coupled Device*
- JPEG: *Joint Photographic Experts Groups*
- HTTP: *Hyper-Text Transfer Protocol*
- LCD: *Liquid Cristal Display*
- MB: *Mega Bytes = 1 milhão de Bytes*
- MPEG: *Moving Picture Experts Group*
- NN: *Neural Networks*
- PTZ: *Pan-Tilt-Zoom*
- RAM: *Random Access Memory*
- SLR: *Single-lens Reflex*
- SOM: *Self-Organizing Maps*
- SVM: *Support Vector Machines*
- USB: *Universal Serial Bus*

Capítulo 1

Introdução

Existem tarefas que, mesmo realizadas com rapidez por qualquer ser humano, possuem um alto grau de complexidade quando da execução por uma máquina. O reconhecimento de objetos ou pessoas através de uma única imagem é um exemplo de uma tarefa com estas características. Nesta dissertação, é discutida uma proposta em que é feita a aplicação automática de regras de Composição Fotográfica exclusivamente a partir de imagens. As regras de Composição são um conjunto de regras utilizadas por fotógrafos com a intenção de melhorar a qualidade de uma fotografia através, dentre outros aspectos, da ênfase ao seu(s) tema(s), o tema fotográfico, portanto, é o elemento que se deseja enfatizar em uma imagem.

A avaliação da qualidade de uma obra-de-arte é altamente subjetiva, pois a avaliação certamente variará entre diversos observadores e, possivelmente, também variará entre os julgamentos de um mesmo observador, obtidos em instantes diferentes. Isto ocorre devido ao fato de que o "gosto" por uma determinada arte ou o quão agradável uma arte pode parecer diz respeito ao sentimento que é causado por sua apreciação, sentimento este que é influenciado pelo meio e pela atual condição emocional do observador [Kant, 1993]. De acordo com Kant: "O juízo de gosto funda-se sobre um conceito (de um fundamento em geral da conformidade a fins subjectiva da natureza para a faculdade do juízo), a partir do qual porém nada pode ser conhecido e provado acerca do objecto, porque esse conceito é em si indeterminável e inapropriado para o conhecimento".

A fotografia, como uma forma de arte, também segue este princípio, não sendo possível fazer-se um juízo sobre quais fotografias são melhores. Entretanto, quando se trata da fotografia como ferramenta de trabalho ou lazer, tem-se um objetivo mais bem definido -

uma pessoa, uma personalidade, um acontecimento, um lugar. Graças a isso, pode-se avaliar a qualidade de uma fotografia pela forma utilizada por um determinado fotógrafo para que esse objetivo - a fotografia que retrate bem uma dada cena, pessoa dentre outros alvos - pudesse ser atingido. Conseqüentemente, nesta dissertação estarão sendo tratadas apenas as fotografias cujos objetivos sejam claros, ou seja, as fotografias com um tema bem definido. Para simplificar, o tema fotográfico abordado nesta dissertação serão sempre pessoas.

Mesmo em iguais condições, a diferença entre a qualidade de uma fotografia obtida por um fotógrafo profissional e um amador pode ser significativa, pois o próprio material utilizado pode ser um diferenciador na qualidade do produto final. Contudo, as ações tomadas pelo fotógrafo profissional podem não ter sido muito mais complexas. O profissional de fotografia, em geral, para ter uma foto de qualidade superior, além da experiência (item este que não pode ser medido *a priori*) apenas seguiu algumas regras, em alguns casos até mesmo de forma inconsciente, que certamente o amador nem sequer concorda que são corretas, apenas concordando que a qualidade final da fotografia foi superior.

As regras citadas no parágrafo anterior são chamadas de Regras de Composição Fotográfica. Para que uma foto seja considerada boa, não é preciso apenas que as pessoas-alvo da fotografia estejam enquadradas - como a grande maioria das pessoas imagina. Entretanto, uma fotografia, ainda que não seja considerada tecnicamente boa, pode ser considerada excelente se estiver dotada de um valor sentimental para o observador que a julga (e.g.; um observador qualquer provavelmente preferirá a fotografia de seu filho à fotografia de outra criança, ainda que a outra fotografia seja tecnicamente melhor). Tal valor não é tratado por nenhuma técnica de composição existente. Por outro lado, existem diversas dessas regras que, se seguidas corretamente, podem melhorar sensivelmente a qualidade da fotografia.

As regras de composição podem ser consideradas, em grande parte dos casos, como processos empíricos usados pelos profissionais da área e que, de acordo com um consenso quase geral, se seguidas convenientemente podem produzir fotos com maior qualidade do que as fotos obtidas sem a utilização de regras. Daí sua proliferação ao longo dos anos, principalmente no meio profissional. Há fotógrafos, contudo, que defendem que muitas dessas regras podem e devem ser quebradas em situações as mais diversas, produzindo fotografias com qualidade igual ou superior àquelas fotos que seguem as regras, o que, no entanto, exige uma maior experiência e percepção visual do profissional [Grill and Scanlon, 1990].

Saber quebrar uma regra de composição pode ser tão (ou mais) importante do que saber usá-la. A aplicação da regra traz benefícios à fotografia, já a não-aplicação da regra pode (ou não) implicar melhorias na qualidade estética da fotografia. Em se tratando de fotógrafos amadores, a não aplicação das regras usualmente não produz bons resultados. Entretanto, da mesma forma que, se um fotógrafo seguir a risca todas as regras, não há garantia da obtenção de uma boa foto, saber quebrar regras não é requisito essencial de um bom fotógrafo. Portanto, as regras não devem ser encaradas como "regras" no seu sentido mais amplo, e sim como guias para melhoria das fotos.

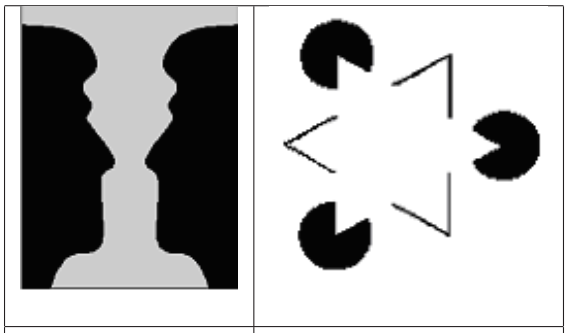
As regras se consolidaram na medida em que esses padrões eram percebidos em um número grande de fotos consideradas muito boas. Daí, o grande objetivo de fazer com que quaisquer pessoas pudessem obter uma significativa melhoria na qualidade da fotografia ao seguir esses padrões de qualidade. Portanto, se a composição for usada para fortalecer a idéia que o fotógrafo deseja passar, ela deve ser usada. Todavia, se a utilização da regra tirar a concentração do fotógrafo, a regra não deve ser usada [Grill and Scanlon, 1990].

As regras são bem mais antigas do que o ato de fotografar. Há centenas de anos, os pintores já utilizavam algumas das técnicas para desenvolver seus quadros, inclusive os mais consagrados pintores tais como Michelangelo e Leonardo da Vinci, dentre outros artistas [Ramalho and Palacin, 2004].

Basicamente, as regras objetivam direcionar a atenção do observador para um ponto específico. Deseja-se que o observador aprecie a fotografia por uma quantidade maior de tempo, o que pode ser uma medida da qualidade da fotografia. Portanto, uma forma de se alcançar isto é manter a atenção do observador. A atenção humana direciona-se em decorrência de diversos mecanismos de "baixo nível" que atuam concorrentemente no cérebro.

Dentre os mecanismos de obtenção de padrões da visão humana, podem-se destacar os descritos pela teoria de Gestalt [Gomes Filho, 2000]. A palavra *Gestalt* tem origem alemã, mas não possui uma tradução precisa para o Português, se aproximando de "forma" ou "formato do conjunto". As leis de Gestalt mostram como a visão humana se comporta em determinadas situações. A Figura 1.1 apresenta um exemplo explicado pela teoria de Gestalt. O cérebro busca por padrões que são percebidos mesmo sem estarem presentes fisicamente na imagem. Esta teoria pode ser resumida como sendo a busca por estas formas, e como o cérebro as procura. A Teoria de Gestalt, por ser ligada a imagens e como as partes de um

grupo se comportam é, portanto, muito utilizada também na fotografia [Linhares, 2004].



(a) (b)

Figura 1.1: Exemplos visuais da Teoria de Gestalt. Em (a) não é claro se a figura representa duas faces ou se representa um cálice. Em (b) a ausência das linhas não impede de reconhecer a forma. O triângulo branco não possui linhas, mas o cérebro humano as vê.

Com base em todas estas informações, é proposta nesta dissertação uma abordagem para que um subconjunto dessas regras de composição possa ser automatizado a fim de que sejam produzidas fotografias de melhor qualidade. A forma como essas regras são obedecidas é que varia do processamento de uma imagem estática para a análise de uma cena dinâmica.

Esta dissertação está inserida no contexto do projeto iPhotoBot iniciado no ano de 2005, desenvolvido em colaboração com a HP-Brasil. Dentre os objetivos do projeto para aquele ano, estava a elaboração de um protótipo para composição automática de fotografias, o qual foi implementado pelo autor deste trabalho. Na presente dissertação, busca-se melhor elaborar este método e validar estatisticamente os resultados através da análise subjetiva de uma quantidade maior de pessoas.

1.1 Motivação

A Fotografia é uma atividade que vem crescendo recentemente em número de adeptos, de equipamentos e de investimentos de empresas privadas. Estimativas sugerem que a quantidade de informação fotográfica produzida em 2002 foi de 440 *petabytes* (10^{15} *bytes*) [Lyman and Varian, 2003]. Este crescimento se deve, especialmente, ao surgimento das câme-

ras fotográficas digitais e as recentes inovações que as mesmas vem trazendo nos últimos anos [photography.com, 2006].

Como conseqüência, tem-se que o ato de adquirir uma foto está mais barato (uma vez que não é necessário revelar todas as fotos obtidas), tem mais qualidade (dado que alguns ajustes automáticos evitam a deterioração de muitas fotografias) e é mais rápido, pois os resultados da fotografia podem ser vistos de imediato no visor LCD da câmera ou em qualquer computador que possua porta USB, não sendo mais necessário esperar o processo de revelação da fotografia para visualizá-la. Este é um dos fatores que mais atrai os fotógrafos (amadores ou profissionais) a preterirem sua predecessora - a câmera analógica, que utiliza filme ao invés de um sensor CCD para capturar a luz. Há ainda a possibilidade de imprimir a foto, processo que pode ser realizado usualmente nas mesmas casas comerciais que outrora revelaram os negativos dos filmes.

Com o armazenamento em dispositivos eletrônicos, ao invés do filme fotossensível, ganha-se em velocidade (não se perde tempo com trocas de filmes), segurança (não há perigo de se queimar todo o filme por abrir-se a câmera - acidente comum), reprodutibilidade (a cópia digital é reproduzida rapidamente), previsibilidade (o fotógrafo pode prever o resultado) e reciclabilidade (o fotógrafo pode apagar uma fotografia que considerar ruim, reutilizando de imediato o espaço da memória recuperado).

Entretanto, existem algumas desvantagens do uso das câmeras digitais. A primeira desvantagem é quanto à qualidade da fotografia. Estima-se que as câmeras digitais precisariam ter sensores capazes de capturar e armazenar 14 *Mega Pixels* [Clark, 2001] para que a qualidade da fotografia digital fosse equivalente à fotografia de um filme de 35mm.

As câmeras digitais vêm trazendo, desde seu advento, avanços na solução de problemas que, até então, eram grandes desafios para a popularização da fotografia. Estes problemas são, normalmente, decorrentes da necessidade de configurações na câmera e requererem maior conhecimento e experiência do usuário pois cada fotografia exigia uma nova análise do ambiente.

Logo, nas câmeras (de filme) mais comuns, essas configurações possuem um valor padronizado de forma a tentar atingir um maior número de pessoas, o que gerava problemas na qualidade da fotografia. O foco da câmera, por exemplo, era configurado no infinito (∞) por esta configuração produzir fotos de qualidade regular em grande parte das situações. Por

vezes, a não configuração destes valores era confundida com a falta de qualidade da câmera.

Neste contexto, podem-se citar alguns avanços integrados às câmeras digitais a escolha automática dos valores de abertura do diafragma e o valor para a velocidade do obturador, calculados com base na iluminação capturada pela máquina. Estas escolhas são, em geral, independentes, podendo o fotógrafo fixar o valor de uma para obter um dado efeito visual e apenas modificar o valor da outra. A má escolha da combinação destes dois valores pode resultar em sub ou superexposição (respectivamente uma fotografia muito escura ou muito clara), fotografias borradas, dentre outros problemas comuns. Dada a complexidade para a escolha manual destes valores, estas configurações padrão já figuram em muitos modelos profissionais de câmeras digitais e analógicas.

Da mesma forma que sistemas mais simples estão sendo integrados ao processamento da máquina sem perda significativa de desempenho, mas com ganho na qualidade final da imagem, espera-se que seja possível o desenvolvimento de outros sistemas que se ocupem de antigas preocupações do fotógrafo, transformando o ato de fotografar em um ato simples, prazeroso e que produza resultados de alta qualidade.

Entretanto, há de se convir que, na busca por um efeito artístico, um fotógrafo possa abdicar de algumas correções. Não seria viável reunir todas as possíveis situações em um sistema computacional, de forma que este fosse capaz de ser executado em um equipamento como uma máquina digital (levando-se em consideração as tecnologias atuais).

Em virtude do exposto, serão tratados apenas as imperfeições mais comuns nas fotografias de amadores ou profissionais, imperfeições estas que são encontradas quando as câmeras fotográficas são utilizadas em situações corriqueiras nas quais o fotógrafo define pessoas como sendo o seu alvo.

1.2 Objetivos e Relevância

Nesta dissertação, são propostas abordagens que objetivam o aumento da qualidade de uma fotografia digital. As correções podem tanto ser realizadas diretamente, como passando pelo crivo final do fotógrafo. No caso de solução supostamente passível de incorporação em câmeras, deve-se avaliar a complexidade da solução de forma que seja factível utilizá-la em um processador de baixa capacidade de processamento, como é o caso dos processadores de

câmeras digitais.

Como objetivos específicos, tem-se as seguintes propostas:

- Desenvolver um sistema autônomo para efetuar ajustes considerados benéficos à fotografia. Algumas teorias podem dar suporte a algumas leis estéticas tal como a Teoria de Gestalt [Desolneux et al., 2004], entretanto boa parte das regras não dispõe previamente de suporte matemático. Devido a esta dificuldade, foram focadas apenas as regras que possuíssem consenso na literatura de Fotografia, minimizando, assim, a necessidade de validação subjetiva do ajuste. Neste objetivo, os ajustes foram realizados após a fotografia ter sido obtida, desde que alguma face fosse detectada.

- Desenvolver um sistema que utilize uma câmera *Pan-Tilt-Zoom* (câmera dotada de motores para rotação e ajuste do ângulo de visão) para localizar, enquadrar, fotografar e analisar as fotografias quanto à composição fotográfica. Este é um problema pouco tratado na literatura. Além disso, os trabalhos existentes normalmente utilizam as informações oriundas de sensores ou câmeras extras. Nesta dissertação discute-se como, utilizando uma única câmera, é possível realizar estas tarefas. Assim, é mostrado como pode ser feita a correção antes mesmo da fotografia ser obtida (método on-line de correção).

1.3 Estrutura da Dissertação

Esta dissertação está dividida em seis capítulos.

No Capítulo 2, mostra-se e discute-se o estado-da-arte do problema da composição fotográfica. De que forma e com qual eficácia atuam os principais métodos existentes. A revisão bibliográfica se inicia procurando na própria fotografia quais os principais fatores e características que podem ser encontrados nas fotografias consideradas boas por uma ampla audiência. Em seguida foi realizado um amplo estudo na literatura em que as correções são feitas de forma off-line na imagem e em propostas para algoritmos que utilizam robótica para solução do problema. Por fim, um estudo realizado em áreas correlatas a este trabalho.

No Capítulo 3, descrevem-se algoritmos para composição automática de fotografias e para análise da qualidade de uma fotografia, ambos desenvolvidos no decorrer deste traba-

lho. São apresentados a inspiração, os trabalhos relacionados e o algoritmo em si, ilustrando como o trabalho descrito pode solucionar o problema. Os algoritmos de composição desenvolvidos são suportados por um módulo inteligente que escolhe dinamicamente qual regra de composição deve ser utilizada. Este capítulo também descreve a extração de características com o fim de se obter informações relevantes em uma fotografia assim como as regras comumente utilizadas por um dado conjunto de fotógrafos e qual a influência da utilização de uma dada regra no resultado final segundo a opinião do fotógrafo e de uma audiência imparcial.

O Capítulo 4 contém recomendações de como usar as regras de composição descritas nos capítulos anteriores para guiar um sistema de correção on-line, ilustrado pelo uso de uma câmera *Pan-Tilt-Zoom*, para que as regras de composição sejam aplicadas. Antes de realizar a composição, contudo, é preciso localizar o alvo no ambiente de uma cena. Neste capítulo, é apresentado um algoritmo para movimentar a câmera em busca de pessoas e, a partir da experiência aprendida com o passar do tempo, aumentar o número de procuras em posições de maior probabilidade de se encontrarem as mesmas pessoas e variar de acordo com essa experiência em busca de pessoas ainda não localizadas.

O Capítulo 5 reúne os experimentos realizados na integração das técnicas discutidas nos Capítulos 3 e 4. É realizada, portanto, uma análise quantitativa e qualitativa dos algoritmos de forma a determinar o quão promissor pode ser a utilização da composição fotográfica na melhoria da qualidade de uma fotografia.

No Capítulo 6, apresentam-se as conclusões obtidas a partir deste estudo, as principais contribuições e os trabalhos futuros que podem ser derivados a partir do exposto ao longo desta dissertação.

Por fim, o Apêndice A traz conceitos sobre fotografia e sobre regras de composição, sob o ponto de vista de profissionais. Os Apêndices C e D contêm amostras de imagens obtidas, resultantes dos experimentos descritos no Capítulo 5.

Capítulo 2

Trabalhos Relacionados

O tema da automatização de ajustes em fotografias digitais vem recebendo cada vez mais novas publicações [Datta et al., 2006; Santella et al., 2006; Zhang et al., 2005; Banerjee and Evans, 2004; Byers et al., 2004; Byers et al., 2003]. Entretanto, ainda é reduzido o número de trabalhos dedicados a este assunto, quando comparado ao número de trabalhos em processamento de imagens em geral. Isto pode ser justificado em decorrência de alguns fatores específicos. Em primeiro lugar, trata-se de um problema novo, considerando que a tecnologia da fotografia digital é recente e vem constantemente sendo modificada.

Em segundo lugar, têm-se também as dificuldades inerentes ao problema, como a metodologia de validação dos resultados e a construção da base de testes. Geralmente, a avaliação da qualidade da fotografia tem um alto teor de subjetividade envolvido, portanto é difícil analisar se a melhora estatística produzida por um algoritmo proposto: (a) se deve a coincidências existentes em uma base de imagens (que podem ocorrer, por exemplo, devido à utilização de um mesmo modelo de câmera fotográfica) ou (b) se deve à avaliação de um dado grupo de juízes que possua mesma opinião sobre o assunto.

Enfim, trata-se de um problema desafiador e repleto de incertezas, a começar pela dificuldade em encontrar literatura técnica especializada na área de análise automática de fotografias. Este capítulo visa à descrição dos trabalhos mais relevantes relacionados ao tema pesquisado. Para uma melhor apresentação, os artigos estão divididos em três seções: técnicas auxiliares, composição automática de fotografias e detecção e enquadramento de temas fotográficos utilizando uma câmera *Pan-Tilt-Zoom*.

Na primeira seção, serão revisados artigos cujos conteúdos não são diretamente relacio-

nados à dissertação aqui apresentada, porém apresentam idéias que puderam ser utilizadas no desenvolvimento de experimentos. A seção seguinte trata de composição automática de fotografias, na qual serão descritos artigos que têm por finalidade principal a composição automática de imagens, seja qual for o contexto da aplicação proposta. Por fim, são apresentados artigos que tratam de câmeras *Pan-Tilt-Zoom*, mas sem o objetivo específico da composição fotográfica, pois artigos sobre este assunto específicos ainda são escasos.

2.1 Técnicas Auxiliares

Nesta seção, são apresentados trabalhos que são ou podem ser utilizados para resolver problemas descritos nesta dissertação, como a localização do tema em uma imagem que, mesmo não sendo o foco deste trabalho, é parte essencial da abordagem proposta merecendo, assim, destaque pois pode influir no comportamento do restante do sistema.

Dentre as formas de detecção e rastreamento (do inglês *tracking*) de pessoas, destacam-se a detecção de faces a partir de Redes Neurais [Rowley et al., 1998a; Rowley et al., 1998b] ou a partir do uso da imagem integral (do inglês *integral image*) e do algoritmo Adaboost [Viola and Jones, 2001].

Em uma fotografia, contudo, muitas vezes as pessoas aparecem de corpo inteiro. É fundamental a análise da pose das pessoas também como um elemento determinante da qualidade da imagem. Vários trabalhos têm a detecção da pose como finalidade. Dentre estes trabalhos, Takahashi e Sugakawa [Takahashi and Sugakawa, 2004] utilizam redes SOM para a detecção de imagens com o *background* subtraído com a utilização de uma *Lookup Table*. Cavalcanti e Gomes [Cavalcanti and Gomes, 2005] utilizam filtros de pele aliados a cálculos das dimensões das áreas detectadas para, de forma heurística, inferir se a área é passível de ser uma parte do corpo humano.

Sprague et al. [Sprague and Luo, 2002] utilizam segmentação de cores para lidar com o problema de detecção de partes do corpo de pessoas vestidas. Após segmentadas as regiões, é construída uma árvore para aumentar a eficiência da busca para, em seguida, classificar utilizando posição, tamanho, cor, etc. relativas entre duas partes para avaliar a possibilidade de pertencerem ao mesmo corpo. No trabalho de Yamada et al. [Yamada et al., 1998], é proposta uma detecção de silhueta com base na redução do *background*, a qual pode ser

realizada analisando-se o valor médio do *background* e comparando-se ao mesmo após a oclusão.

Em uma outra abordagem, Hu et al. [Hu et al., 2000] usam um método estatístico para, a partir de uma câmera fixa, extrair o *background* para, em seguida, utilizar Algoritmos Genéticos para associar a silhueta obtida no primeiro passo a um modelo. Por fim, Ozer et al. [Ozer and Wolf, 2002] apresentam um sistema para detecção de atividade/gestos em um domínio de compressão (usando o MPEG) para, em seguida, aplicá-lo em um sistema sem compressão.

Na detecção de objetos (sem que necessariamente sejam pessoas), existem as abordagens propostas por Luo et al. [Luo et al., 2004], Li et al. [Li et al., 1999] e Berg et al. [Berg et al., 2005]. Na primeira abordagem, a detecção de objetos é realizada através de segmentação de imagens com base na saliência e em um treinamento utilizando redes de Bayes. Já na segunda abordagem, utiliza-se o campo de profundidade (semelhante ao descrito acima para o trabalho de Banerjee et al. [Banerjee and Evans, 2004]) para, a partir de imagens nas quais as bordas de um tema estão em destaque, efetuar a segmentação do objeto de interesse. Por fim, na terceira abordagem, é descrita uma abordagem utilizando casamento deformável (do inglês *deformable shape matching*). Apesar deste último trabalho ser destinado, *a priori*, ao casamento de padrões semelhantes, a estratégia também pode ser útil na procura de pontos importantes em uma imagem.

2.2 Composição Automática de Fotografias

O problema estudado, composição automática de fotografias, é ainda pouco explorado na literatura atual. Neste parágrafo, são enumerados alguns dos motivos que influenciam este ainda pequeno interesse pelo assunto. Em primeiro lugar, este é um problema muito recente, que surgiu paralelamente ao avanço dos aparelhos de digitalização de imagens e ao barateamento das câmeras fotográficas digitais. Com as câmeras fotográficas tradicionais (cujo processo de captura, armazenamento e exibição das fotografias e suas diferenças para uma máquina digital serão tratados em capítulos posteriores), não é possível realizar diretamente nenhuma alteração na fotografia. As únicas ações que um fotógrafo pode realizar são filtragens de cor (através de filtros colocados sobre a lente objetiva antes da fotografia ser obtida)

e alterações posteriores feitas através de recortes e mudanças nas características do papel fotográfico. Estas modificações não podem ser desfeitas e um erro pode comprometer todo o trabalho do fotógrafo. Nesta seção, são apresentadas abordagens existentes na literatura, as quais propõem dar um passo adiante na solução do problema da composição automática de fotografias.

Byers et al. [Byers et al., 2003; Byers et al., 2004] apresentam uma proposta para a construção de um robô fotógrafo, capaz de se movimentar em um ambiente plano em busca de pessoas e, em seguida, fotografá-las. As pessoas são encontradas por um sensor laser e, em seguida, via um filtro da tonalidade da pele. Os autores dizem não usar detecção de faces em virtude do alto consumo de energia necessário para isso. Após detectadas as pessoas, para se locomover no ambiente, o robô cria um mapa do local considerando as pessoas detectadas como objetivos a serem calculados por funções. Definida a rota, inicia-se o movimento na direção deste objetivo. São utilizadas duas câmeras: uma de alta velocidade e outra de alta definição. A de alta velocidade provê frames constantemente, em busca de um alvo. Já a câmera de alta definição é utilizada apenas para a obtenção de fotografias de pessoas detectadas pelo sistema como um todo (câmera e sensores).

Quando de frente ao tema, a câmera principal é movimentada (junto ao robô) com a intenção de obedecer a algumas regras de composição fotográfica. O experimento é testado em uma situação real (um evento científico e um casamento) e o resultado é avaliado pela taxa de fotos que foram requeridas pelos fotografados como também por uma votação feita posteriormente com 2000 imagens, na qual os votantes escolhiam uma dentre 5 gradações de qualidade. O experimento obteve 9% de fotografias consideradas muito boas e 20% de fotografias consideradas boas.

O trabalho de Byers et al. diferencia-se desta dissertação em alguns aspectos de muita relevância. Em primeiro lugar, pela localização das pessoas, já que o autor utiliza sensores *laser*, enquanto nesta dissertação apenas as imagens capturadas na cena são utilizadas. Em utilizando apenas imagens oriundas da câmera, tem-se uma abordagem mais simples, devido à menor quantidade de informação a ser tratada. Um segundo aspecto diz respeito à detecção de pessoas, a qual é feita através de filtros de pele na abordagem proposta por Byers, enquanto um detector de faces é utilizado nesta dissertação. De fato, a detecção de faces é mais cara, computacionalmente falando, porém mais precisa e objetiva do que unicamente a

detecção de pele.

A etapa de composição fotográfica é semelhante, mudando apenas a metodologia da composição, mas no trabalho de Byers et al. são utilizadas regras de composição bastante similares aos desta dissertação. A avaliação do resultado também é realizada de forma similar. No trabalho de Byers et al., entretanto, pressupõe-se que haverá interação entre as pessoas e o robô, enquanto no trabalho aqui apresentado não há essa obrigação, podendo as fotografias serem obtidas sem a expectativa da pose.

O trabalho de Banerjee et al. [Banerjee and Evans, 2004] trata especificamente de uma regra de composição, a regra-dos-terços. Utilizando uma máquina fotográfica com grande abertura do diafragma, é possível, através do foco seletivo, o qual é obtido com baixos campos de profundidade, enfatizar as bordas do tema. Desta forma, é possível detectar o objeto da fotografia e efetuar sua transposição (utilizando recortes ou redimensionamentos) para um dos pontos dos terços. O processamento é realizado de forma rápida e não há a necessidade de detectores de alvos pré-treinados, sendo a informação da imagem suficiente para a composição.

A abordagem descrita no supra-citado trabalho, tem como característica positiva a não necessidade de algoritmos de detecção e a eficiência do seu código, o qual pode realizar o processamento de forma mais rápida do que o algoritmo aqui apresentado. Além disso, qualquer outro objeto que, porventura, esteja visível na imagem e esteja a uma mesma distância do objeto de interesse é detectado como objeto e sua composição é realizada, ao contrário do desejado.

O ponto fraco desta abordagem, entretanto, é o requerimento de uma pré-configuração da máquina que, usualmente, não consegue ser bem reproduzida em uma máquina digital comum, por requerer grande abertura do diafragma da câmera. A abertura do diafragma produz um efeito melhor visto em equipamentos com maior distância focal. As câmeras digitais mais populares, entretanto, são fabricadas com menor distância focal com o objetivo de serem mais compactas, uma vez que precisariam de uma grande lente para aumentar a distância focal. Por fim, a abordagem apresentada não seria útil em imagens já obtidas sem esta configuração.

No artigo *Auto Croping for Digital Photographs*, Zhang et al. [Zhang et al., 2005] apresentam uma abordagem similar à desenvolvida nesta dissertação. O trabalho gira em torno do

problema de *auto cropping*, aqui batizado de regra-do-zoom, o qual usa uma função de energia que avalia três modelos: o modelo de composição, o modelo de penalidade e o modelo conservativo.

Os dois primeiros trabalhos citados no parágrafo anterior, assemelham-se às regras aqui definidas como regra-dos-terços e regra-da-integridade, respectivamente. Para detecção do tema, é utilizada detecção de faces aliada a modelos de atenção visual, para auxiliar na detecção de elementos não encontrados pelas detecções de faces. Este trabalho também é voltado para a composição de imagens já obtidas, porém não dá suporte à obtenção de novas fotografias.

Algumas diferenças cruciais entre a abordagem de Zhang et al. [Zhang et al., 2005] e a presente dissertação podem ser apontadas. A primeira é a quantidade de imagens utilizadas nos testes, 100 contra 1327 utilizadas nesta dissertação. Em seguida, o número de padrões de zoom considerados 14, contra 4 nesta dissertação (ver Capítulo 3 para mais detalhes).

Por fim, a votação utilizada para validar o algoritmo utilizado na primeira abordagem apresenta 3 opções aos votantes: boa, aceitável e ruim, sendo as duas primeiras favoráveis à concordância, enquanto a abordagem aqui proposta apresenta 4 opções: ótima, boa, ruim e péssima, favorecendo uma melhor separação entre as opiniões positivas e negativas. Na primeira abordagem, em se contabilizando apenas as votações consideradas boas, houve uma concordância da audiência de 41% ou, somando-se as duas probabilidades, têm-se uma aceitação de 84% por parte da audiência, contra os 65% obtidos na segunda abordagem (ver Capítulo 3, para maiores detalhes a respeito destes resultados).

Datta et al. [Datta et al., 2006] apresentam um estudo tratando da avaliação da estética de uma imagem utilizando uma abordagem computacional. São propostas 56 métricas das quais, experimentalmente, extraiu-se as 15 que obtiveram melhor combinação. A abordagem atinge uma taxa média de 70,12% de acerto ao utilizar SVM - *Support Vector Machines* (ou Máquinas de Vetores Suporte) [Vapnik, 1999] para efetuar a classificação a partir destas 15 métricas.

A abordagem desenvolvida nesta dissertação é semelhante àquela apresentada por Datta et al. [Datta et al., 2006] ao utilizar algumas métricas para classificar uma fotografia quanto a sua qualidade. O treinamento utilizado em algumas regras, contudo, não deixa transparecer a metodologia e quais seus defeitos e suas qualidades, logo, não é possível saber como

as 15 regras escolhidas se comportaram nas diferentes situações, o que não permite prever quais destas utilizar para a extração de características em um sistema similar como o aqui proposto. Adicionalmente, o trabalho apresentado é voltado para a classificação de imagens, não tendo soluções para corrigir uma imagem que poderia ser boa caso pequenos ajustes fossem realizados. Apesar destas observações, é um excelente trabalho, o qual pode ser utilizado posteriormente com outras finalidades.

Por fim, em um recente trabalho de Santella et al. [Santella et al., 2006], é enfatizada a regra-do-zoom, descrita em seu trabalho como *cropping*, utilizando direção do olhar (do ingles *gaze direction*) para decidir quais são os elementos principais da fotografia. De acordo com o experimento realizado nesta dissertação, que utiliza uma votação para validar o experimento, é obtida uma taxa de aceitação de 58% quando o trabalho é comparado ao método que usa detecção dos pontos de maior saliência da imagem.

No mesmo supra-citado trabalho, ao utilizar apenas a atenção visual como informação de posicionamento do alvo, possibilita que erros sejam cometidos ao permitir a ênfase de pontos de atenção que podem ser destacados por características particulares da imagem. A união desta informação com outras informações reduziria a incidência deste problema. Além da localização do alvo da fotografia, a abordagem difere desta dissertação no tocante ao recorte proposto, já que no trabalho de Santella et al. não há restrições quanto às dimensões da imagem final, podendo o recorte assumir quaisquer dimensões, enquanto nesta dissertação o recorte deve ter proporções idênticas as da imagem original.

2.3 Detecção e Enquadramento de Temas Utilizando Câmera *Pan-Tilt-Zoom*

Nesta seção, serão apresentados os principais trabalhos estudados que se referem ao controle de uma câmera PTZ (*Pan-Tilt-Zoom*) de forma que esta possa interagir com um determinado cenário. Até onde se pôde apurar, não existem artigos que tratem especificamente da composição automática de fotografias através da utilização de uma câmera motorizada. Portanto, é feita nesta seção uma descrição mais abrangente dos trabalhos, de forma a levantar-se diversas técnicas e algoritmos que possam ser utilizados.

2.3.1 Enquadramento de Temas

Sinha e Pollefeys [Sinha and Pollefeys, 2004] apresentam uma abordagem para, utilizando apenas uma câmera, efetuar a calibração e a fotografia de uma cena panorâmica sem dispor de nenhuma informação da estrutura da cena. Os parâmetros intrínsecos da câmera são estimados calculando as homografias das imagens capturadas a partir de diferentes ângulos de rotação e níveis de zoom. Os parâmetros extrínsecos são calculados a partir de pares de imagens panorâmicas, obtidas em se fazendo um mosaico a partir de imagens obtidas em um nível fixo de zoom.

A vantagem da abordagem é a qualidade final da imagem panorâmica e a possibilidade de se efetuar a calibração utilizando-se apenas uma câmera PTZ, sem nenhuma outra informação. A principal desvantagem é o tempo necessário para fazê-lo. Leva-se em torno de 25 minutos para obtenção de uma fotografia panorâmica calibrada. Este trabalho possui o diferencial de utilizar apenas a câmera PTZ tanto para efetuar a calibração como para se obter a imagem panorâmica.

O trabalho de Clady et al. [Clady et al., 2001] apresenta um sistema voltado para a assistência de direção de carros. O objetivo é o desenvolvimento de sensores visuais com a utilização de uma câmera PTZ e de uma câmera normal para efetuar o rastreamento (do inglês *tracking*) de veículos frontais ao veículo no qual estão posicionadas as câmeras.

O algoritmo proposto pelos supra-citados autores, segundo eles próprios, é veloz e robusto, funcionando a uma velocidade de 20ms e sendo capaz de localizar o tema também em baixa iluminação. O processo se dá pela captura de imagens através da câmera PTZ e, em seguida, pela utilização de um algoritmo de rastreamento para detectar a posição e tamanho do objeto a ser procurado e a movimentação necessária à câmera. O algoritmo de rastreamento segue a abordagem de Hager e Belhumeur [Hager and Belhumeur, 1998]. A câmera auxiliar é utilizada para continuar a localização quando a câmera PTZ tiver seu ângulo de visão reduzido, ao focar em um dado objeto.

A abordagem de Clady et al. foi estudada, de forma a se entender como poderia ser utilizada uma câmera auxiliar à PTZ, a fim de que esta pudesse fornecer uma visão mais ampla da cena, já que tem grande ângulo de visão. Entretanto, para que a solução descrita nesta dissertação pudesse utilizar tal abordagem, seria necessário mais de uma câmera semelhante à câmera sugerida por Clady et al., de forma que fosse possível se ter uma visão mais ampla

da cena, já que o espaço de busca é bem maior do que o exemplo descrito no artigo, o qual engloba apenas a visão frontal do veículo.

Um outro trabalho, desenvolvido por Senior e Hampapur [Senior and Hampapur, 2005], apresenta uma abordagem para obtenção de imagens em multi-escala através do controle de uma câmera PTZ calibrada automaticamente a partir de outras câmeras. Para um controle autônomo, podem-se definir marcações especiais feitas pelo usuário, as quais indicam na imagem oriunda da câmera PTZ qual o conjunto de coordenadas que deve ser utilizado para o correto enquadramento de um dado ponto através de uma *Lookup Table*.

Pode-se também utilizar mais de uma câmera para obter a correta calibração da imagem de forma automática, em se apontando ambas as câmeras para uma região comum favorável, na qual movimente-se um objeto. Assim como no trabalho de Sinha e Pollefeys, é utilizada a homografia para calibração das imagens.

Novamente o método de calibração proposto, desta vez por Senior e Hampapur [Senior and Hampapur, 2005], exige uma segunda câmera, pré-condição que se deseja evitar. A abordagem utilizando *Lookup Table* foi testada nesta dissertação, não se obtendo o devido sucesso (maiores detalhes podem ser obtidos no Capítulo 4).

A abordagem proposta por Funahashi et al. [Funahasahi et al., 2004] busca uma localização e *tracking* das partes de uma face em se utilizando câmeras PTZ. O sistema proposto utiliza um par de câmeras, sendo uma PTZ e a outra uma câmera CCD fixa para, de forma hierárquica, detectar e fotografar partes da face de uma pessoa.

2.3.2 Escolha do Percurso

O sistema proposto nesta dissertação procura por temas de interesse em um ambiente do qual *a priori* não se tem informação. Dado que o ambiente pode ser planificado, através das fotografias que podem ser obtidas pela câmera PTZ, pode-se associar o problema estudado nesta dissertação ao problema de escolha de percursos pesquisado na Robótica, pois o objetivo da câmera pode ser resumido à procura, em um plano, por objetos de interesse através da movimentação de um agente (a própria câmera). Portanto, as estratégias de busca implementadas para robôs interagirem com uma cena na qual precisam caminhar em um plano, podem ser utilizadas por esta abordagem.

O primeiro trabalho estudado é o de Tomono [Tomono, 2003]. Este trabalho propõe um

sistema para construção automática do percurso de um agente, através das detecções feitas no ambiente. Também é realizado um mapa probabilístico do ambiente, no qual são simuladas as modificações que podem existir dentro de um raio de tolerância de movimentação. O objetivo é criar um mapa durante a execução, caso sejam detectados alvos ou obstáculos.

Este trabalho foge um pouco ao escopo de investigação desta dissertação, uma vez que contrói um mapa com base em movimentação limitada por um raio de tolerância, condição esta que poderia ter sido utilizada nesta dissertação, porém não é desejável pois limitaria o problema a uma dada situação.

No segundo trabalho, de Stack e Smith [Stack and Smith, 2003], propõe-se um algoritmo para busca de minas submersas, o qual utiliza o padrão de busca linear com variações randômicas e alteradas pelo histórico para localizar minas, já que estas geralmente estão igualmente espaçadas a fim de que o navio que as dispersou possa conhecer um caminho de retorno.

Apesar do segundo trabalho acima não ser fortemente relacionado a esta dissertação, este é utilizado como uma inspiração para o tipo de busca realizado pela câmera PTZ. Maiores detalhes podem ser vistos no Capítulo 4.

Mais dois trabalhos, semelhantes entre si, propõem a busca de alvos por uma região. O primeiro por Yang e Luo [Yang and Luo, 2004] e o segundo por Qiu et al. [Qiu et al., 2006]. Nestes trabalhos, um robô movimenta-se pela cena e, de acordo com o que é detectado, calcula o percurso de forma dinâmica através do treinamento de uma Rede Neural [Haykin, 1999] para aprender a alterar sua trajetória quando da detecção de obstáculos em uma cena.

Apesar de semelhantes, o segundo algoritmo apresenta o diferencial de utilizar uma janela de predição a qual, após prever uma situação de bloqueio, adapta o sistema para a nova realidade diminuindo, portanto, a quantidade de rotações necessárias, o que agiliza a busca. Apesar das diferenças, ambos os artigos são preparados para tratar de objetos que se movem no ambiente.

As duas abordagens poderiam ser utilizadas como inspiração para o desenvolvimento desta dissertação. Entretanto, a segunda abordagem, descrita anteriormente, é mais complexa que a primeira, em virtude da necessidade de um treinamento extra, o qual não faria muito sentido no problema aqui apresentado, uma vez que não existem obstáculos no sistema. Portanto, seria caro, computacionalmente falando, um treinamento que é realizado

para prever obstáculos que não existirão. Trabalhos futuros onde existam regiões no ambiente nos quais a câmera não possa procurar por assuntos, pode utilizar esta abordagem para uma melhor busca e construção do caminho de procura.

2.4 Considerações Finais

Este capítulo apresentou um estudo sobre os métodos atualmente utilizados para a composição automática de fotografias como também de técnicas auxiliares utilizadas nesta dissertação, essenciais a um sistema para automação da composição fotográfica como um todo.

A revisão bibliográfica foi dividida em três seções, sendo a primeira uma revisão das técnicas auxiliares, nas quais se destacou a detecção de um tema em uma cena, através de diversas abordagens tais como a filtragem da cor da pele, detecção de movimento e rastreamento de partes do corpo humano de pessoas envolvidas na cena.

A segunda parte contou com uma revisão bibliográfica sobre regras de composição, a qual apresentou os algoritmos utilizados para que seja feita composição automática de fotografias, seja em imagens já obtidas ou ainda a serem adquiridas, as quais podem ser ajustados com a análise da informação detectada na cena.

Por fim, uma parte dedicada a algoritmos de movimentação de câmeras *Pan-Tilt-Zoom* nas mais diversas finalidades: partindo de *tracking* de automóveis até planejamento da rota ideal a ser percorrida por robôs.

Estas três seções objetivaram dar suporte a toda a dissertação, de forma a se obter um trabalho que reúna um pouco de cada característica e realize a tarefa específica da composição automática de fotografias em qualquer que seja a situação, antes ou apos a fotografia ser obtida ou ainda analisando as imagens para uma correta obtenção ou classificação da imagem.

O estudo dos trabalhos relacionados à abordagem apresentada nesta dissertação mostra que as técnicas utilizadas são promissoras e, em virtude de suas datas de publicação, refletem o atual estado-da-arte do problema. No próximo capítulo, serão abordadas técnicas para composição automática de fotografias realizadas em imagens estáticas.

Capítulo 3

Algoritmos para Composição Fotográfica e Extração de Características

Neste capítulo, é apresentada uma abordagem para corrigir imagens no tocante à composição fotográfica, de forma a melhorar sua qualidade visual. Algoritmos implementando três regras de composição são aplicados em imagens já obtidas gerando novas imagens. Um conjunto de voluntários opina quanto à melhoria destas últimas para avaliar o desempenho dos algoritmos. Também é descrito um algoritmo para extração de características o qual pode ser utilizado para classificar as imagens quanto à qualidade, tendo por base a conformidade às regras de composição.

3.1 Introdução

A fotografia artística é uma das mais conhecidas e praticadas formas contemporâneas de arte [Hedgecoe, 2003]. Apesar de suas técnicas e princípios terem variado pouco ao longo dos anos, ainda não é de amplo domínio dos fotógrafos amadores o conhecimento destas técnicas. Por outro lado, fotografar nos dias de hoje é mais fácil e prazeroso em decorrência de avanços da tecnologia utilizada nas câmeras. Entretando, ainda existe uma quantidade incalculável de tarefas que poderiam ser realizadas de forma automática pela câmera, facilitando o trabalho do fotógrafo e, conseqüentemete, aumentando a qualidade final da fotografia, dando um aspecto profissional à mesma independentemente do conhecimento das técnicas de fotografia do operador. A composição fotográfica pode ser uma destas tarefas.

As regras de composição fotográfica podem ser vistas, de maneira geral, como heurísticas utilizadas há anos por fotógrafos experientes, que se popularizaram ao ponto de serem consideradas regras. Apesar de grande parte das regras de composição fotográfica terem sido desenvolvidas empiricamente, algumas são sustentadas por algumas teorias da psicologia e do funcionamento do sistema visual humano. Uma destas teorias é a Teoria de Gestalt [Koffka, 1955], que explica como o cérebro humano interpreta alguns padrões visualizados pelo olho. Esses padrões influenciam a percepção pelo cérebro da cena, podendo então serem utilizados para guiar a atenção do observador para um determinado ponto de interesse.

Logo, as regras visam a enfatizar o tema da fotografia [Grill and Scanlon, 1990]. O tema (também denominado alvo ou assunto, também sendo utilizadas ambas as denominações ao longo desta dissertação) da fotografia é "o quê" o fotógrafo deseja mostrar a um dado observador sobre a cena fotografada, ainda que este observador não possua conhecimento técnico sobre fotografia ou arte. Qualquer elemento pode ser o tema da fotografia, inclusive entes abstratos, por exemplo, pessoas, um objeto, um animal até mesmo o sentimento entre duas pessoas. O que importa é que este tema seja claro ao observador e enfatizado na fotografia. Por ser bastante amplo e abstrato, o tema é restrito e bem definido nesta dissertação, evitando, assim, divergências entre os observadores. Desta forma, toda fotografia utilizada tem, obrigatoriamente, um tema o qual deve ser pessoas - em conjunto ou não. Extensões desta dissertação poderão focar em um outro objeto ou outra forma de interpretar a cena.

Ao longo desta dissertação, é dado foco em algumas regras de composição fotográfica. Algumas heurísticas utilizadas por fotógrafos também serão implementadas nesta dissertação e, por serem amplamente conhecidas e utilizadas, a exemplo do ajuste do Zoom[1], será usada a denominação de regras-de-composição.

As três principais regras trabalhadas são a regra-dos-terços, regra-do-zoom e regra-da-integridade do alvo. Para localizar o alvo, pode-se utilizar um detector de faces como evidência da presença de pessoas. Conseqüentemente, em se assumindo que é conhecido o posicionamento e as dimensões das faces contidas em uma fotografia, pode-se deduzir o posicionamento e dimensão do alvo e usar estas informações para alterar a imagem, resultando

[1]Nome popular dado à mudança da distância focal da câmera. A conseqüência é a alteração gradual, dentro de um mesmo plano, do ângulo de visão. Chama-se zoom-in quando este diminui e zoom-out quando aumenta. Esta mudança é feita através da movimentação das lentes móveis da câmera.

sempre que possível em uma foto bem composta. Afora estas três regras, outras poderiam ser facilmente integradas ao sistema, contudo estas não foram incluídas nesta dissertação por questões de tempo, restrição de escopo e, adicionalmente, com a intenção de se facilitar a verificação da corretude do sistema a partir de poucos módulos, já que a modificação feita em uma mesma imagem por diversas regras de composição poderiam dificultar a visualização da correção.

A solução aqui descrita foi implementada em *software* e foram utilizadas funções de processamento de imagens (na maior parte dos casos, recortes e redimensionamentos) para atingir o resultado desejado. Por questões de simplicidade, os algoritmos a seguir são descritos para efetuar correções em imagens estáticas. Entretanto, os algoritmo desenvolvidos para aplicar estas regras, poderão fazê-lo também antes da fotografia ser obtida. Para que a fotografia seja corrigida antes mesmo de ser obtida, é preciso o uso de um agente externo, seja mecanismos de movimentação para a câmera, o próprio fotógrafo (que pode ser alertado através de indicações visuais ou na utilização de um sistema autônomo). No Capítulo 4, é apresentada uma discussão sobre a aplicação destes algoritmos em ambientes dinâmicos.

É preciso, também, que seja definido o escopo deste trabalho. O sistema descrito neste capítulo trata apenas de fotografias cujos temas possam ser automaticamente detectados por um agente externo. Como exemplo, foi utilizado um detector de faces sendo, portanto, expandido para pessoas através de regras de antropometria. Detectores de faces, entretanto, possuem limites no ângulo da face em relação à fotografia. Logo, muitas imagens podem conter faces que o detector utilizado não encontre e que venham a interferir na fotografia, uma vez que o sistema desconhece a existência de faces naquela região, podendo resultar em cortes não desejados. Portanto, faces não detectadas devem ser ignoradas na seção de votação da qualidade da fotografia. Com relação a antropometria, devem-se ignorar diferenças de proporção entre pessoas seja por sexo, idade ou raça. Estas questões podem ser tratadas posteriormente.

Este capítulo é dividido em 6 seções. Na Seção 3.2, é apresentada uma revisão bibliográfica sobre composição fotográfica a fim de identificar quais as principais regras de composição utilizadas, quais são as mais genéricas, os principais problemas encontrados pelos profissionais de fotografia que podem ser automatizados, os principais erros cometidos pelos fotógrafos amadores além de outras informações oriunda de profissionais que sejam determi-

nantes de como o sistema deve se portar e quais regras obedecer. Na Seção 3.3, é apresentada a abordagem para a solução dos problemas (ou de sua maioria) identificados na seção anterior. São descritos os módulos desenvolvidos e os componentes auxiliares como o detector de temas. A Seção 3.4 contém os experimentos realizados com a intenção de validar a abordagem descrita neste capítulo. A Seção 3.5 trata de algoritmos desenvolvidos para extrair informação de imagens sem nela efetuar mudanças com a intenção de avaliar uma fotografia com base nestas informações. Por fim, na Seção 3.6, são apresentadas algumas conclusões específicas deste capítulo.

3.2 Composição Fotográfica

Esta dissertação aborda técnicas de composição fotográfica que podem ser utilizadas com a finalidade de se melhorar uma fotografia. Na Seção 2.2, foram descritos os trabalhos mais relevantes publicados até o momento que tratam de composição automática de fotografias. Na presente seção, por outro lado, o foco está na área de Fotografia (não automática), de forma a identificar as regras de composição e heurísticas mais relevantes.

Vale salientar que realizar a revisão em Fotografia não foi uma tarefa simples nem os resultados finais foram os originalmente planejados. Isto se deveu principalmente ao fato de que a Fotografia explora o conceito de Arte, na qual não existem certezas e sim sentimentos, tudo é permitido, desde que atenda os ideais filosóficos do artista implicando um baixo número de livros sobre técnicas de fotografia pois as regras não podem atar a criatividade do artista. As regras de composição fotográfica em si, além de outros exemplos, podem ser vistos no Apêndice A.

O primeiro trabalho revisado é uma referência da área, datado de 1990 e escrito por Grill e Scanlon [Grill and Scanlon, 1990]. Neste livro, os autores defendem que um dos principais erros que fazem uma fotografia perder o seu valor de comunicação é que, sendo o objetivo da fotografia na maioria dos caso o de preservar a memória do fotógrafo, este esquece que a fotografia deve possuir conteúdo para que outros observadores também possam captar o sentimento obtido no momento da fotografia. Isto se dá pois ao não definir claramente o alvo de sua fotografia, o fotógrafo não trata a composição da fotografia desta forma, enfraquecendo a qualidade e a objetividade da mensagem que é passada em uma fotografia. O autor

também destaca a necessidade da utilização dos padrões que o cérebro humano insistentemente procura ao visualizar uma imagem. Logo, a sugestão para utilização de linhas, pontos e cores, claramente delineados ou a partir de outros objetos que formam intrinsecamente tais elementos.

Em seu livro "Photographing People", Michael Freeman [Freeman, 2004], faz algumas considerações subjetivas sobre fotografia. Segundo o autor, "A grande preocupação do retratista é revelar a natureza das pessoas" mas nunca abdicando da simplicidade. Ainda segundo o supra-citado autor, "A face humana já é suficientemente interessante sem técnicas complicadas ou composição incomum". Quanto às técnicas, é destacado o cuidado com objetos de distração, tais como cores, imagens e palavras, de forma a não adicionarem conteúdo sem criar distração. Na Figura 3.1, mostra-se uma imagem com e sem elemento de distração. Em se tratando da integridade do alvo, o autor reitera a necessidade de um cuidado maior no ajuste da câmera, para evitar cortes que possam gerar desconforto ao observador, assim como um corte da imagem na altura dos joelhos do tema fotografado.

Figura 3.1: Na imagem da esquerda as formas contribuíram para a distração do ponto de interesse, enquanto na imagem da direita as formas ajudaram a direcionar o olhar para o tema.

Com relação aos níveis de zoom, Freeman [Freeman, 2004] destaca características para cada um destes níveis. Para fotografias de cabeça e ombros, o corte na altura dos ombros deve parecer "quadrado" quando a intenção for criar uma aparência estática e formal. Outra importante recomendação é a de fotografar ao nível dos olhos do alvo. Para imagens de

corpo inteiro, pode-se descuidar um pouco da expressão e concentrar-se mais na expressão corporal. Porém, as grandes dificuldades são a falta de elementos para "preencher" a imagem e encontrar uma relação entre o tema e o plano-de-fundo.

Para fotografias de pequenos grupos, é sugerida uma composição em forma de triângulo, tentando criar uma relação entre as pessoas. Já para grandes grupos, sugere apenas quebrar a tradicional foto em forma de fila se estiver clara a intenção do fotógrafo e houver contribuição dos modelos. A Figura 3.2 apresenta exemplos de fotografias obtidas com as características discutidas acima, como a fotografia no mesmo nível dos olhos do alvo e a composição em forma de triângulo

(a) (b) (c)

Figura 3.2: Em (a) uma imagem obtida posicionando a câmera superiormente ao alvo, em (b) exemplo do ganho de qualidade quando a fotografia é obtida com a câmera na altura do olho da criança. Em (c) exemplo de composição em forma de triângulo.

Bill Hurter começa seu livro [Hurter, 2004] sugerindo que o que faz uma boa fotografia é a imaginação que o fotógrafo consegue fazer surgir do observador. O autor descreve o surgimento do ato de posar de forma a minimizar as perdas de transferir o ser humano de três dimensões para uma imagem em das dimensões.

Quanto às técnicas de composição, Hurter [Hurter, 2004] afirma que a composição é responsável por posicionar corretamente o tema dentro do quadro de fotografia e esta é a fonte da maioria dos erros, já que os fotógrafos menos experientes preferem colocar o tema no centro da imagem. Quanto à regra-dos-terços, o autor afirma que o objetivo é criar assimetria do tema na imagem.

Sobre os níveis de zoom, Hurter [Hurter, 2004] afirma que, em fotografias de cabeça e ombros, o olho da pessoa fotografada deve ser o ponto de interesse. Já para uma fotografia de média e longa distância (do inglês, *medium* e *long-shot*), a face é o centro de interesse. O

autor também destaca a necessidade de espaço (ou sala) para a borda na direção em que os olhos da pessoa fotografada apontam.

Ainda segundo o supra-citado autor, em fotografias contendo pequenos grupos a utilização da forma de triângulo, de maneira que a linha virtual que liga as faces da imagem descreva um triângulo. Para fotos de grupos, sua principal observação é quanto a proximidade entre as pessoas para que i) todos estejam dentro de um mesmo plano de foco e ii) próximos para representar cordialidade e distância para representar elegância. As faces também devem distar umas das outras de forma consistente. Em grandes diferenças de altura, considerar a possibilidade de uma foto em longa distância para amenizar o efeito da distância.

Pequenos grupos, tais como casais, devem sempre estar em alturas diferentes, aproximando-se a altura dos olhos para a altura da boca. Esta pequena diferença cria linhas virtuais entre os alvos, dando direção à fotografia. Segundo o autor, há linhas implícitas por toda a cena e é papel do fotógrafo utilizar estas linhas para dar direção e suavidade à imagem. Por exemplo em fotografias de grupos, deve-se observar a linha dos ombros e das faces. Em se tendo um conjunto de linhas cujas ligações são muito agudas, deve-se ajustar de forma que o polígono formado por estas linhas apresente angulações mais suaves.

Por fim, dois conselhos extraídos do livro de Bill Hurter [Hurter, 2004]. Em primeiro lugar, não capturar fotografias com número par de pessoas pois, segundo o autor, o cérebro e os olhos tendem a aceitar mais facilmente a desordem causada pelos números ímpares do que de objetos em quantidade par. Em segundo lugar, deve-se ter cuidado com o plano de fundo da imagem para que este não se confunda com o tema gerando efeitos desagradáveis podendo estragar, inclusive, belas composições.

A partir de outro trabalho [Hedgecoe, 2005] cujo autor é um fotógrafo reconhecido pelo número de livros voltados para todos os níveis de conhecimento sobre fotografia, podem-se destacar algumas observações.

Inicialmente, o fotógrafo destaca o recorte em uma imagem como uma forma de isolar o tema de outros elementos de distração, seja no plano-de-fundo seja no plano frontal da fotografia.

Para o autor, boas fotografias de pessoas devem não somente mostrar a aparência de pessoas, mas funcionar como uma espécie de biografia visual, capturando características marcantes e revelando sua personalidade. Em fotografias de curta distância (ou *close-ups*),

por exemplo, o foco nos olhos, ao invés de qualquer outra parte, pode ajudar na revelação da personalidade da pessoa.

Também é importante que o tema esteja ao nível da câmera, forçando o fotógrafo, portanto, a nivelar-se ao tema, ao invés de rotacionar a câmera como, por exemplo, em uma fotografia de um criança que seja ou esteja mais baixa do que o fotógrafo. Neste caso, o fotógrafo deve se abaixar para se igualar à altura da criança, ao invés de, de pé, fotografá-la apontando a câmera para baixo.

Em outra parte do texto, o autor apresenta os erros mais comuns na obtenção de fotografias. Um primeiro erro é a obstrução do alvo, quando o fotógrafo, por distração, deixa um dedo, a tira da câmera ou até mesmo parte de uma parede obstruindo o caminho entre a câmera e o alvo. Este problema é mais comum nas câmeras em que o visor é separado da lente objetiva da câmera.

Em seguida, erros na utilização do *flash*, resultado em sub-exposição, super-exposição ou um problema conhecido por vinheta. Erros relacionados ao *flash* normalmente são causados por má configuração da câmera ou por aproximação excessiva ou insuficiente do alvo da fotografia. Não diretamente relacionado ao *flash*, mas causado por ele, o efeito dos olhos vermelhos, quando as pessoas na fotografia ficam com olhos vermelhos. Este efeito ocorre mais comumente quando pessoas são fotografadas em ambientes escuros. Com a dilatação da pupila, causada pela escuridão do ambiente, o *flash* da câmera ilumina a retina do olho das pessoas diretamente e, por ser esta irrigada de vasos sanguíneos, apresenta a coloração vermelha, causando o efeito que é mais percebido em câmeras compactas, devido à proximidade do obturador para a luz do *flash*.

Por fim, a baixa qualidade da fotografia seja por erro no foco ou seja por trepidações na câmera. No primeiro caso, o problema é decorrente do mau ajuste do foco da câmera que pode ter fixado em outro objeto da cena; no segundo caso, a razão pode estar assiciada à configuração da câmera, a qual pode estar permitindo uma grande quantidade de tempo de captura. Na Figura 3.3, mostram-se exemplos de fotografias com erro no foco, olhos vemelhos e obstrução.

Muitas outras observações, regras e informações puderam ser extraídos dos trabalhos acima citados ou de outros não mencionados. Estas outras informações não são discutidas devido ao baixo consenso entre o material estudado ou necessidade de maior compreensão

(a) (b) (c)

Figura 3.3: Erros comuns: (a) Erro no foco, (b) Efeito dos olhos vermelhos e (c) obstrução.

do *significado* de elementros da fotografia. Portanto, com a finalidade de delimitar o escopo deste trabalho, serão considerados estas obras como ponto de partida, ignorando-se conceitos mais elaborados que venham a requerer um nível maior de interpretação da cena, requerimento este aquém das técnicas correntes de Processamento de Imagens e Visão Computacional. As regras aqui apresentadas, portanto, tentam ser mais fiéis à opinião dos autores destes trabalhos.

3.3 Abordagem Proposta

O sistema proposto pode ser dividido em três partes: captura da imagem, processamento da imagem e armazenamento da imagem processada. A captura da imagem pode ser feita por uma câmera ou até mesmo por uma imagem em disco, sendo externa ao sistema e o armazenamento da imagem deve ser feito em algum dispositivo. Na Figura 3.4, ilustra-se o sistema proposto. A imagem é passada para o detector de faces que retorna as coordenadas das faces detectadas na imagem (caso alguma face seja detectada). Maiores detalhes a respeito da detecção de faces são fornecidos na Seção 3.3.1. As coordenadas das faces e a imagem original (que pode ser necessária a algum módulo de composição para extração de característiacs adicionais) são transmitidas aos módulos de composição fotográfica que decidem pelo recorte que deve ser efetuado nas imagens para que as regras de composição sejam obedecidas. Por fim, o módulo de recorte utiliza esta informação para recortar a imagem original produzindo uma nova imagem.

A abordagem proposta para a composição fotográfica é composta por quatro módulos. O primeiro módulo é um controlador, destinado a gerenciar os outros três módulos que imple-

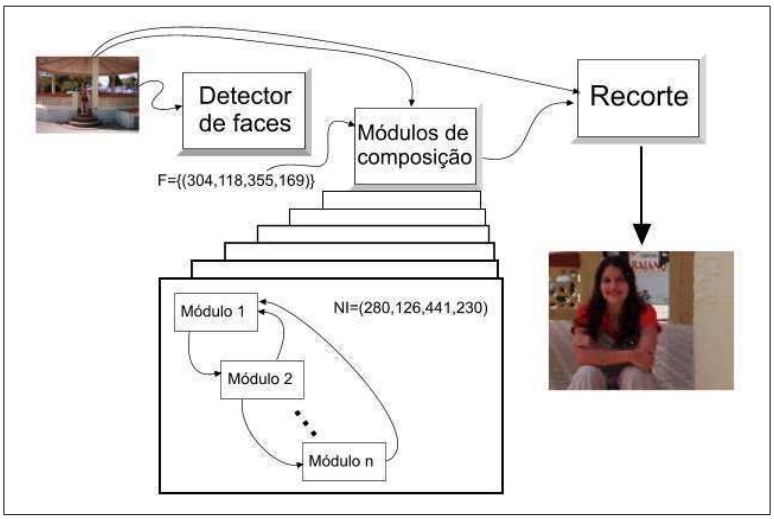


Figura 3.4: A partir da câmera ou disco uma imagem é processada pelo sistema de composição fotográfica sendo armazenado em disco.

mentam, cada um, uma regra de composição. O controlador, a partir das imagens capturadas da câmera, decide qual regra de composição utilizar (caso a aplicação de alguma regra seja necessária). A alteração pode tanto ser realizada na imagem obtida, gerando assim uma nova imagem, quanto diretamente na fonte (e.g., a câmera) para posterior requisição de uma nova imagem, situação tratada no Capítulo 4. Esta estratégia permite uma fácil adição ou remoção de outros módulos ao sistema pois cada módulo recebe uma imagem, processa-a e retorna uma nova imagem, não havendo um forte acoplamento entre os módulos, já que não deve existir dependência entre eles. A Figura 3.4 ilustra o relacionamento entre os módulos de composição.

Os módulos de composição são executados independentemente, embora seguindo uma ordem pré-selecionada. Cada módulo tem suas próprias especificações e sua forma de melhorar a qualidade da imagem. As especificações podem ser personalizadas, deixando a abordagem mais flexível. Por exemplo, um fotógrafo pode preferir uma dada configuração na qual o módulo do zoom tem prioridade quanto ao módulo dos terços, enquanto um outro fotógrafo pode escolher o caminho inverso. A ordem dos módulos altera o resultado, já que, em decorrência do que já foi dito, o módulo posterior processa a imagem que foi retornada pelo módulo anterior, conseqüentemente, já modificada.

É também tarefa do controlador, garantir que a alteração feita por um módulo não interfira em um outro módulo de hierarquia superior. Por exemplo, o módulo de zoom não deve

promover alterações na imagem que resultem em um corte no tema quando o módulo de integridade, que possui hierarquia mais alta, indicar um corte na região. Nesta abordagem, a hierarquia de controle foi definida através da precedência, logo, um dado módulo não pode efetuar correções que em hipótese alguma desfaçam ou contrariem as mudanças realizadas no módulo executado anteriormente. Caso a alteração proposta por um módulo venha a interferir no processamento realizado por algum módulo que o antecedeu, nenhuma alteração será realizada.

Após todas as mudanças terem sido realizadas, o passo final é redimensionar a imagem (caso mudanças em suas dimensões originais tenham sido processadas) mantendo a proporção original. Este passo final pode ser trocado pela alteração na própria câmera, em se tratando de um sistema dinâmico.

Os módulos de composição propostos possuem dois aspectos em comum. Em primeiro lugar, a imagem original não é afetada. Portanto, apenas sugestões de novas coordenadas, características, etc são informadas, deixando a decisão final ao módulo controlador. Em segundo lugar, todos usam as coordenadas da face como informação primordial para realizar as modificações.

3.3.1 Detecção do Tema da Fotografia

Para que seja feita a composição fotográfica, o elemento mais importante a ser definido é o tema. Em se encontrando/definindo o tema da fotografia, todas as modificações serão realizadas utilizando-o como base.

A detecção de alvos, independente de qual for usada, deve ser executada logo após a fotografia ser obtida e unicamente neste momento, dado que usualmente trata-se de um algoritmo computacionalmente caro. Em conseqüência, deve-se guardar a informação e re-processar a posição a cada alteração na imagem sem que seja executado novamente.

As seções a seguir descrevem como, a partir do detector de faces, são inferidos onde estão e qual espaço ocupam as pessoas de uma cena. Na ausência de um detector mais robusto, capaz de encontrar as pessoas por completo e suas respectivas poses, o detector de faces é a melhor abordagem.

Detector de Faces

O detector de faces provê a informação principal nesta abordagem aqui apresentada. Assim, é utilizado um detector de faces inspirado na abordagem de Viola & Jones [Viola and Jones, 2001]. Apesar da importância do algoritmo de detecção de faces para o sistema, outros algoritmos de detecção de faces podem ser utilizados sem grandes prejuízos ao sistema como um todo, desde que sejam observados dois fatores: (a) o desempenho do algoritmo - dado que o detector de faces é aplicado nos quadros capturados pela câmera, a velocidade do detector influi na taxa de quadros por segundo que o sistema vai ser capaz de lidar; e (b) a precisão do detector de faces, dado que, como descrito na Seção 3.3.1, nossos algoritmos utilizam proporções antropométricas sendo assim, a imprecisão na detecção do tamanho da face pode levar a uma inferência incorreta das outras dimensões do corpo. Outros dois detectores de faces foram testados, produzindo resultados similares e mostrando que a diferença de precisão pode ser, se não desprezível, controlável: o detector de faces desenvolvido pelo grupo liderado por Kanade [Rowley et al., 1998a] e o detector de faces da OpenCV [OpenCV, 2005]. A Figura 3.5 ilustra a comparação desses três detectores.

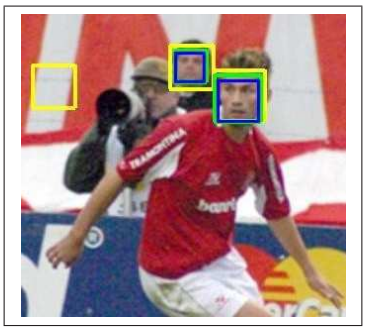


Figura 3.5: Comparação entre diferentes detectores de face: amarelo corresponde ao OpenCV, verde ao proposto por Viola & Jones e azul o detector do grupo liderado por Kanade.

A diferença entre os detectores de face, entretanto, não costuma ser significativa, uma vez que a amplitude de possíveis definições para face não é grande. A maior fonte de erros é relativa a falsas detecções ou falsas rejeições. Estima-se, com base na quantidade de imagens que produziram erros em decorrência de problemas na detecção de faces, que se o detector em questão não produzisse erros, o ganho de aceitação com relação aos votantes seria de mais de 10%, sendo este um importante fator a ser considerado em versões futuras.

Medidas Antropométricas

Devido à necessidade de se conhecer o espaço que cada tema ocupa, seja para evitar um corte indesejável, seja para utilizar estas linhas como guia, é desejável a utilização de bibliotecas para detecção de pessoas, tais como aquelas propostas nos artigos descritos no Capítulo 2 [Takahashi and Sugakawa, 2004; Ozer and Wolf, 2002; Sprague and Luo, 2002; Hu et al., 2000; Yamada et al., 1998]. Todas possuem custo computacional elevado que, devido ao tempo já gasto para detecção de faces, poderia inviabilizar o sistema devido às requisições de velocidade no processamento para atingir uma taxa de quadros por segundo adequada ao problema enfrentado.

Alternativamente, é proposto nesta dissertação um algoritmo para estimar o espaço ocupado por pessoas a partir de relações entre as partes do corpo, em particular entre as dimensões da face e do corpo humano. Desta forma, partindo do presuposto de que se sabe a posição e dimensões das faces, pode-se inferir, de forma aproximada, qual o espaço possivelmente ocupado pelo corpo do tema. Estas relações, nomeadas de Antropometria, são conhecidas na Anatomia e nas artes plásticas (e.g., Homem Vitruviano de Leonardo da Vinci, ilustrado na Figura 3.6).

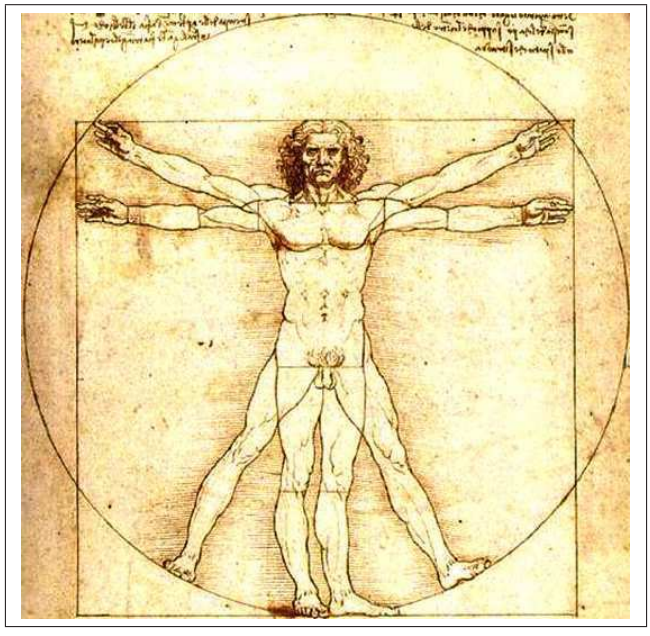


Figura 3.6: Homem de Vitrúvio mostra um exemplo do estudo antropométrico aplicado às artes plásticas.

As relações utilizadas, contudo, são válidas apenas para corpos de pessoas adultas. As proporções não são válidas para crianças, pois a cabeça, ao nascer, tem proporção aproximada de metade do tamanho do corpo da criança, tem seu crescimento desacelerado com a idade até atingir a fase adulta [Dace, 2006]. Em se tratando de adultos, embora possa existir variação nas proporções quando da aplicação em um grupo heterogêneo, esta diferença não é significativa. Algumas relações são mostradas abaixo:

$$TamanhoDoCorpo = 7 \times TamanhoDaCabeca; \tag{3.1}$$

$$LinhaDaCintura = 4 \times TamanhoDaCabeca; \tag{3.2}$$

$$LarguraDosOmbros = 3 \times LarguraDaCabeca; \tag{3.3}$$

3.3.2 Regra-dos-Terços

A regra-dos-terços é uma das mais antigas e mais conhecidas regras de composição. A regra sugere que o tema da fotografia seja posicionado em uma posição específica da imagem. Esta posição é escolhida de acordo com um dos quatro pontos de interseção das retas verticais e horizontais que dividem a imagem em nove partes de mesmas dimensões. Para tanto, duas retas são posicionadas verticalmente e duas horizontalmente, sendo elas equidistantes-entre-si (considerando-se o mesmo conjunto de mesma orientação). O nome da regra vem do fato de as retas dividirem a imagem em três partes iguais. Esta técnica, contudo, é antiga e bem anterior ao advento das câmeras fotográficas, tendo sido utilizada nas pinturas feitas por grandes nomes, tais como Leonardo da Vinci e Michelangelo [Ramalho and Palacin, 2004].

Muitos fatores biológicos (quanto ao observador) sustentam esta regra e explicam o porque do posicionamento do tema afastado do centro, uma posição que pode ser mais natural para a composição. O erro mais comum de um fotógrafo amador é utilizar a popular regra do "centralize e fotografe". A principal teoria é de que a descentralização do alvo é importante pois o mecanismo de atenção visual humano foca no alvo e, quando centralizado, ignora o resto da imagem, o que resulta em uma fotografia estática [Hurter, 2004]. O deslocamento do alvo para o terço força o observador a procurar mais entes na cena. Apesar disto, nada impede que o tema da fotografia seja centralizado quando há o interesse do fotógrafo de enfatizá-lo. Um exemplo da regra é mostrado na Figura 3.7. Neste exemplo, o tema da

imagem vai além do rosto da modelo enfatizando seu olhar ao invés da face como um todo.

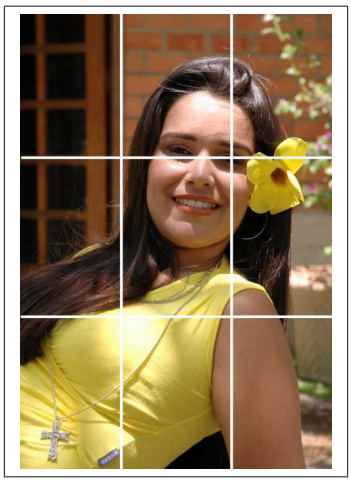


Figura 3.7: Exemplo da regra-dos-terços. O alvo - olho da modelo - está no ponto dos terços.

A abordagem proposta nesta dissertação para a implementação da regra-dos-terços se utiliza das coordenadas das faces presentes em uma imagem provida pelo detector de faces para, em seguida, posicionar estas faces detectadas no ponto de terço mais próximo, adotando o centro da face como ponto de referência. Isto restringe o problema para fotografias contendo pessoas cujas faces possam ser detectadas por algum detector automático de faces. Outras abordagens, tais como a de Banerjee e Evans [Banerjee and Evans, 2004], sugerem a detecção de outras informações, tais como bordas ou intensidades e utilizar o centro-de-massa destas informações como centro de referência.

A escolha do ponto do terço para qual a face deve ser movida é uma etapa trivial dado que se deseja corrigir a fotografia com um mínimo de atuação, o ponto escolhido deve ser o mais próximo ao centro de massa da face. O ponto mais próximo é calculado através de distância Euclideana entre dois pontos, como mostra a equação a seguir. Sendo x e y as coordenadas do centro da face; l a largura e a a altura da imagem, $m = tercoMaisProximo(x,y)$ o ponto de terço mais próximo horizontal e sendo min(x,y) definido como segue:

$$min(x,y) = \begin{cases} 1 & \text{se } x \leq y \\ 2 & \text{se } x > y \end{cases} \tag{3.4}$$

tem-se que:

$$tercoMaisProximo(x,y) = min\left(\sqrt{(\frac{l}{3}-x)^2+(\frac{a}{3}-y)^2}, \sqrt{(\frac{2l}{3}-x)^2+(\frac{a}{3}-y)^2}\right) \quad (3.5)$$

Logo, a coordenada do terço mais próximo é dada por $(\frac{m\times l}{3}, \frac{a}{3})$. Nesse caso, sempre será escolhido o terço superior entre os terços verticais. A escolha pelo terço superior se dá pelo fato de que o posicionamento do tema no terço inferior é mais indicado para situações nas quais a paisagem faz parte da composição, situação esta não considerada nesta dissertação. Pode-se fazer o mesmo cálculo também para o terço inferior ou também o terço mais próximo.

A direção da linha (vertical ou horizontal), que representa a linha-dos-terços, depende da orientação da fotografia (retrato ou paisagem). Este problema é tratado nesta dissertação a partir do uso da informação que já se dispõe sobre a orientação das faces inferindo-se, assim, a direção da fotografia.

Com relação à aplicabilidade desta regra, sendo o tema pessoas e tendo-se as faces detectadas, existe a restrição para se considerar um único tema, o que implica fotografias de apenas uma pessoa. Para fotografias de mais de uma pessoa, caso do centro de massa do grupo ao extremo do grupo diste menos que um terço da fotografia (nas ocasiões em que as pessoas estão próximas umas das outras), pode-se também aplicar a regra acima descrita. Portanto, o funcionamento da regra traduz-se no cálculo do centro-de-massa do tema para, em seguida, calcular-se o desvio necessário para que o tema passe a ocupar o terço da imagem.

A nova largura da imagem é calculada pela seguinte expressão:

$$LarguraDaImagem = 3 \times \frac{LarguraOriginal - XCentroFace}{2} \quad (3.6)$$

em que *LarguraOriginal* é a largura da imagem antes de ser processada e *XCentroFace* é a coordenada x do centro da face encontrada na imagem. Vale salientar que as novas dimensões da imagem são obtidas por meio de recortes e não de redimensionamento.

A nova altura da imagem é calculada de forma semelhante, mas mantendo a proporção original:

$$AlturaDaImagem = LarguraDaImagem \times \frac{AlturaOriginal}{LarguraOriginal} \quad (3.7)$$

Opcionalmente, pode-se ignorar a proporção original, produzindo assim imagens cujas proporções não são iguais a da imagem original. Em se mantendo as proporções iniciais, o próximo passo é o redimensionamento da imagem de maneira que a imagem retorne às

dimensões originais. Este redimensionamento tem como objetivo o de simular a modificação prévia, podendo não ser utilizado.

Para efetuar o redimensionamento indicado no parágrafo anterior, foi utilizado o algoritmo de interpolação por bloco, algoritmo de interpolação padrão da biblioteca CImg [CImg, 2005]. Outros algoritmos de interpolação podem ser utilizados para a obtenção de imagens com melhor qualidade especialmente nas situações nas quais a imagem recortada tem dimensões muito menores que as originais.

Outros temas podem ser utilizados, por exemplo, o olho da pessoa fotografada ao invés do centro da face. Para isso, é necessária a utilização de outros algoritmos de detecção que sejam capazes de detectar outros elementos como, no exemplo, um detector de olhos.

3.3.3 Regra-do-Zoom

Um bom retrato, em geral, requer que as pessoas e lugares possam ser reconhecidos. A distância do fotógrafo ao alvo, portanto, influencia neste reconhecimento, pois quanto mais distante estiver o alvo do fotógrafo, menor é a face na imagem. Isto pode ser evitado corrigindo-se o zoom através das lentes ópticas.

A regra-do-zoom (também chamada de recorte por muitos autores de técnicas automáticas) trabalha na ênfase das áreas mais importantes da fotografia em detrimento do descarte de outras regiões menos relevantes. O ajuste do zoom pode ser realizado de forma que o alvo não esteja excessivamente longe nem perto.

Em algumas situações há o desejo em se mostrar intencionalmente a paisagem de forma a não se levar em consideração um alvo humano. Já em outras situações, há a intenção de se aproximar excessivamente a fotografia do alvo em busca de efeitos artísticos. Ambas estas situações estão fora do escopo deste trabalho.

Visando a obtenção de uma fotografia agradável, é desejável que se permita um espaço entre o alvo e as bordas da imagem [Hurter, 2004]. Este espaço é comumente chamado de *room*. Similarmente, tem-se *headroom* como sendo o espaço existente entre a cabeça e as bordas da imagem [Busselle, 1999].

Fotografias cujas faces estão muito próximas às bordas passam a impressão de "choque" da cabeça na borda. Por outro lado, se as cabeças presentes na imagem estiverem muito longe da borda superior porém extremamente perto da borda inferior, tem-se o efeito de "degola".

Algumas heurísticas são definidas por fotógrafos profissionais com o objetivo de decidir o quão perto do alvo a câmera deve estar [Busselle, 1999]. Assim como acontece com a maioria das outras regras de composição, contudo, essas heurísticas não são unanimemente aceitas. A partir de uma análise de alguns trabalhos nas áreas de fotografia e de cinema [Hedgecoe, 2005; Hurter, 2004; Tremblay, 2003; Okazaki, 1998], nesta dissertação chegou-se à definição de seis padrões de distância para o tema de uma fotografia. Estes seis padrões estão listados a seguir:

- XLS (*extreme wide/long shot*): O alvo está tão distante que se mistura ao plano-de-fundo;
- LS (*long-shot*): O alvo é capturado da cabeça aos pés;
- WMS (*wide medium-shot*) Apenas $\frac{3}{4}$ do corpo do alvo é visto (a partir da cabeça);
- MS (*medium-shot*) A fotografia compreende da cabeça à linha da cintura;
- CUP (*close-up*): A fotografia inclui apenas cabeça e ombros;
- XCUP (*extreme close-up*): Apenas um conjunto de características faz parte da composição (olhos, boca, nariz, etc).

O objetivo desta seção é o desenvolvimento de algoritmos para decidir qual o fator de zoom a ser aplicado na imagem, seja de forma automática (através de uma câmera com interface de zoom controlável via software) ou manual (através de informação visual passada pelo *software* ao usuário através do visor LCD da câmera). Por questões de escopo, são consideradas nessa dissertação apenas as distâncias LS, MS, CUP e XCUP descritas acima, dado que são as distâncias mais comumente usadas por fotógrafos profissionais. Essas regras, contudo, aplicam-se apenas para fotografias contendo uma única pessoa.

A distância do tema (no caso, a face da pessoa fotografada) às bordas da imagem foram definidas arbitrariamente de forma a respeitar o *headroom*. As proporções utilizadas são mostradas na Tabela 3.1. Os valores indicam a distância até as bordas em pixels, medida em função da altura da face (para as bordas Superior e Inferior) e a largura da face (bordas laterais). Ou seja, da lateral da face até a borda esquerda em uma fotografia que tenha

Tabela 3.1: Parâmetros para a regra-do-zoom. Os valores representam a distância em função do tamanho da face.

Level	**Superior**	**Inferior**	**Esquerda**	**Direita**
Long Shot	1,0	8,0	1,0	1,0
Medium Shot	0,7	3,0	0,8	0,8
Close-Up	0,6	0,6	0,4	0,4
Extreme Close-Up	0,08	0,6	0,3	0,3

como meta um *close-up*, a distância mínima necessária é $0,8 \times largura\_da\_face$, sendo $largura\_da\_face$ o tamanho horizontal da face.

Usando as proporções definidas na Seção 3.3.1, foi desenvolvido um algoritmo para a regra-do-zoom. Para testes e ilustrações, a imagem de entrada é recortada e redimensionada para voltar a suas dimensões iniciais. Entretanto, o objetivo é usar este algoritmo em câmeras com controle do zoom.

Caso incorporada a uma câmera digital, a regra-do-zoom poderia funcionar seguindo-se os passos (i) o usuário escolhe uma das configurações (LS, MS, CUP ou XCUP) (ii) o usuário aponta a câmera em direção ao alvo e pressiona pela metade o botão de captura; (iii) o algoritmo embarcado calcula o fator de zoom e recorte necessários para satisfazer a configuração escolhida, considerando as coordenadas detectadas para a face sendo fotografada); se o zoom atual é satisfatório e não há necessidade de modificações, a câmera sinaliza ao usuário para que a fotografia seja obtida; caso contrário, a câmera pode sugerir modificações na fotografia; (iv) quando a câmera está pronta e o usuário pressiona o botão por completo, o zoom é ajustado e a fotografia é obtida. Os níveis de zoom estão ilustrados na Figura 3.8.

3.3.4 Regra-da-Integridade

Esta regra verifica se alguma parte do alvo está sendo "cortada" na fotografia. Dado que a única informação que se possui *a priori* da imagem é a posição das faces e suas dimensões, foi definida uma heurística para inferir se alguma parte do alvo, no caso do corpo da pessoa, estaria além das dimensões da imagem. Esta heurística é embasada nas proporções da Antropometria que inferem dimensões de partes do corpo a partir de outras partes. Logo,

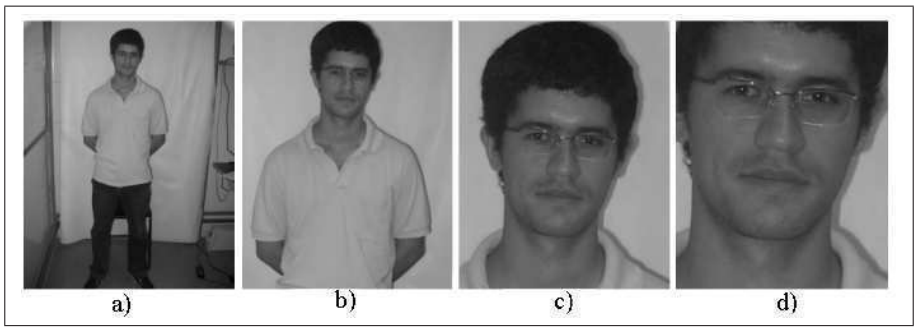


Figura 3.8: Ilustração dos níveis de zoom: (a) *Long-Shot*; (b) *Medium-Shot*; (c) *Close-Up*; (d) *Extreme Close-Up*.

sabendo-se as dimensões da face, pode-se calcular uma aproximação da altura e largura de uma pessoa. A partir da Equação 3.3 e dos valores definidos pela Tabela 3.1, a regra-da-integridade é calculada. Portanto se todas as pessoas da cena, detectadas a partir das razões antropométricas não pertencerem aos limites estabelecidos, considera-se que a regra está sendo infringida.

Evidentemente, um cálculo tão simples pode levar a erros, especialmente quando as pessoas não estão posando de forma ereta, porém este problema poderia ser resolvido com algoritmos mais refinados de detecção sendo deixados, portanto, para trabalhos futuros. Uma possível abordagem pode utilizar um filtro para tonalidade da pele como indício da parte do corpo [Cavalcanti and Gomes, 2005] como também a informação obtida pelo foco da câmera. Adicionalmente, também existem estudos sobre a fotogenia da pose, o que descartaria de foram subjetiva boa parte das poses que o método heurístico falharia.

3.4 Experimentos

Os experimentos têm como objetivo avaliar através da opinião de pessoas o grau de consistência dos algoritmos propostos. Devido à diversidade de opiniões que podem surgir em um experimento como este, não existe uma métrica que possa ser usada e que represente unanimamente as opiniões.

Para tentar reduzir essa ausência de métricas, dois experimentos foram realizados: o primeiro baseado em comparação de imagens antes e depois da aplicação dos algoritmos; e o segundo baseado numa análise técnica de todas as imagens, objetivando checar a corretude

dos algoritmos e quais são as principais causas que impedem que boas fotos sejam obtidas dentro do presente contexto.

Para estes experimentos, foram utilizadas 1327 imagens escolhidas arbitrariamente contendo pessoas em diversas poses. As imagens foram obtidas utilizando um rastreador *web* o qual procura imagens em sítios destinados a postagem de fotografias como por exemplo: www.fotolog.net, www.flogao.com.br, etc.

Após obtidas as imagens desses sítios, um detector de faces foi utilizado para descartar as imagens sem faces detectáveis. Este conjunto de fotografias não foi filtrado por pessoas. O objetivo desta heterogeneidade foi avaliar a robustez do algoritmo em situações inesperadas. Obviamente, algumas fotografias não podem ser corrigidas nem através de software nem manualmente, devido aos graves erros cometidos já no momento da fotografia.

No primeiro experimento, foi arbitrariamente escolhida a distância *Close Up* como padrão para o módulo de zoom. Esta escolha foi feita dado que em maximizando-se as diferenças entre a imagem original e a modificada, é melhor para o observador perceber as diferenças e decidir por qual imagem optar. Esta escolha, contudo, pode ser prejudicial nos casos em que na imagem não se aplicava este nível de zoom.

Um sistema *online* de votação foi desenvolvido e alguns voluntários foram convidados a votar. Não havia requisitos quanto ao conhecimento em fotografia. Um total de cinco voluntários votaram em todas as fotografias. O número de votantes é pequeno devido às características do problema, no qual desejava-se que cada pessoa votasse em todas as fotos. Devido à grande quantidade de imagens, nem todas as pessoas que inicialmente se voluntariaram vieram a concluir a votação.

Um outro ponto a levar-se em consideração é a ausência de profissionais da área no grupo de votantes. Existem alguns fatores positivos e negativos quanto a isso. Como pontos positivos, pode-se citar o fato de os fotógrafos normalmente seguirem linhas artísticas bem caracteristicas de seus trabalhos, o que forçaria a votação de um número muito grande de profissionais para que se pudesse ter uma linha bem definida de opiniões. Outro fator positivo é que a avaliação feita por um fotógrafo muitas vezes é empírica e não corresponde às espectativas de seu cliente este, que por sua vez, usualmente é um fotógrafo amador e é, de fato, público-alvo do sistema aqui proposto.

Como pontos negativos, deve-se enfatizar a grande quantidade de informação que poderia

ter sido obtida com os profissionais e a ausência da validação dos resultados deste trabalho por um grupo seleto e experiente. Esta escolha, contudo, não invalida este trabalho, uma vez que o estado-da-arte da área apresenta artigos com número e nível de conhecimento semelhantes aos votantes convidados para a validação do trabalho proposto nesta dissertacão, tal como os trabalhos de Byers et al. [Byers et al., 2003], Zhang et al. [Zhang et al., 2005] e Santella et al. [Santella et al., 2006].

A cada votante foi solicitado que, ao ver a imagem original à esquerda e a modificada à direira, escolhesse dentre cinco opções. As opções eram: (1) A fotografia modificada é melhor; (2) Não há diferença entre as imagens e não acho que devesse existir; (3) Não há diferença entre as imagens, mas acho que deveria existir; (4) A fotografia modificada é pior; (5) Imprópria.

Em princípio bastariam duas opções: (1) a fotografia está melhor e (2) a fotografia está pior. Entretanto, em alguns casos, a imagem modificada é semelhante a imagem original, não sendo possível diferenciá-las com facilidade. Isso explica as duas opções em que o votante opina que não percebe diferença entre as imagens. Caso o algoritmo não modifique uma imagem originalmente boa, considera-se que o procedimento correto foi tomado. Entretanto, caso tenha sido ignorada alguma correção, considera-se que o procedimento foi incorreto. Já a última opção, impropria, é necessária dado que, como não existiu filtragem manual de conteúdo, podem ter restado resíduos de desenhos, pinturas ou outras imagens cujas detecções de faces foram realizadas incorretamente.

A ordem das imagens foi alterada randomicamente para garantir a coerência do votante, ou seja, a imagem indicada como original e a indicada como modificada podem, eventualmente, estar invertidas. A Figura 3.9 apresenta a tela do navegador capturada no momento da votação.

O objetivo principal deste experimento foi avaliar se o algoritmo processou a imagem de acordo com o esperado. O votante concorda que o algoritmo cumpriu seu papel corretamente quando escolhe as opções 1 ou 3, já que optou pela melhoria de uma imagem ou não-modificação de uma imagem já considerada boa. O votante discorda do algoritmo ao escolher a opção 2 ou 4, indicando que a alteração realizada pelo algoritmo foi inapropriada ou o mesmo deixou de corrigir uma imagem de baixa qualidade. A opção 5 sendo escolhida por um dos votantes invalida os votos dos demais votantes, dado que se uma fotografia pare-

Figura 3.9: Interface utilizada para votação da qualidade da composição efetuada pelo algoritmo proposto.

ceu imprópria para um dado votante, esta mesma fotografia pode confundir outros votantes e gerar um resultado não confiável.

Ao final da votação, as imagens impróprias foram descartadas. Para a contagem foi utilizado o critério de maioria simples, logo bastava que uma das opções tivesse um voto a mais para ser considerada vitoriosa. O resultado final foi: 65% das fotografias que foram processadas pelo algoritmo foram consideradas melhores que suas respectivas imagens de entrada, enquanto 34,7% foram consideradas piores que as originais. As imagens que, porventura, tivessem empate na votação seriam contadas à parte. Uma vez que a quantidade de votantes que opinaram em todas as imagens foi ímpar (cinco votantes), não houve empate. O número de imagens classificadas como impróprias é 374.

A Figura 3.10 apresenta exemplos de imagens as quais os votantes consideraram melhor após a modificação realizada pelo algoritmo e algumas imagens consideradas piores também por ação do algoritmo de composição.

Esta pesquisa objetivava avaliar se, em geral, o algoritmo produzia modificações desejáveis. Não foi possível, contudo, avaliar o quão melhor a imagem é, especialmente em se considerando a alta subjetividade da tarefa. A votação também não considerou fatores externos tais como questões culturais e raciais, os quais podem ser avaliados em trabalhos

futuros.

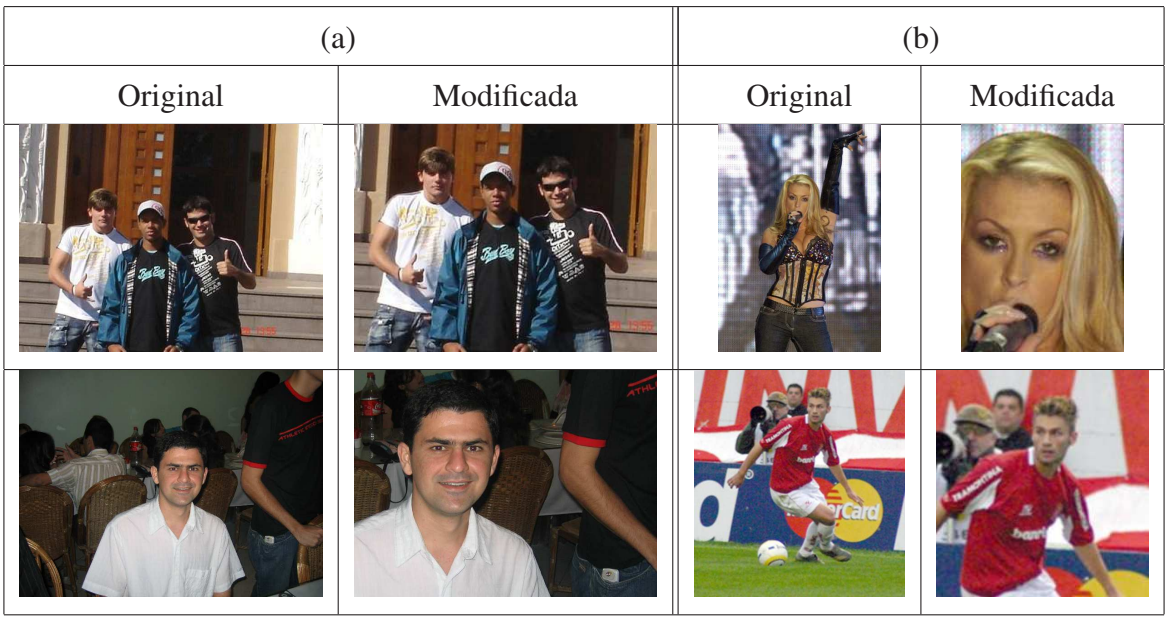


Figura 3.10: Alguns exemplos de fotografias votadas como: (a) A imagem modificada parece melhor e (b) A imagem modificada parece pior.

O alto número de imagens consideradas piores, gerou a demanda de um outro experimento para investigar quais problemas foram determinantes para que a imagem gerada pelo algoritmo proposto nesta dissertação fosse considerada pior que a imagem original.

O segundo experimento, utilizou 378 imagens escolhidas arbitrariamente, mas diferentes das primeiras 1327. O objetivo de se ter utilizado imagens diferentes era verificar o comportamento do algoritmo em um subconjunto menor de imagens. Assim poder-se-ia verificar a qualidade das imagens uma-a-uma no conjunto completo.

Ao contrário do primeiro conjunto, este foi filtrado por humanos, descartando a maioria das imagens impróprias. Os experimentos tiveram também o objetivo de analisar se os módulos aplicaram corretamente suas correções.

As imagens foram divididas em dois grupos de fotografias, sendo o primeiro de fotografias contendo apenas uma pessoa e o segundo contendo grupos de pessoas. Os resultados precisam ser interpretados de forma independente, pois as regras aplicadas também o são.

Os níveis de zoom diferem pois para fotografias de grupos é considerado a região mínima de zoom na qual todo o grupo está contido, ou seja, no nível de zoom *close-up*, a região

mínima deve conter todas as faces e os pescoços, enquanto na fotografia de *medium-shot* é requerido que a linha da cintura esteja visível. O tamanho do grupo varia. O número máximo depende da distância e da lente utilizada na câmera. Evidentemente que, quão mais distante estiver o alvo, mais difícil é a detecção de faces.

A partir do momento que as imagens são estáticas e extraídas da Internet, não há como modificá-las quando a decisão for além de suas fronteiras (como fazer um "*zoom out*" além do tamanho da imagem). Estas imagens irão refletir o atual funcionamento de uma câmera, na qual o zoom não pode ser diminuído além de seu nível mínimo.

A Tabela 3.2 apresenta os principais problemas encontrados neste experimento. A primeira linha da tabela mostra o número de imagens em cada sub-grupo, seguida pela direção do olhar, o que implica a análise da regra-dos-terços, a qual exige espaço na direção do olhar. Oclusão parcial significa que existe alguma parte do tema que está sendo escondida (ou apenas uma pequena parte está sendo mostrada). Poses incomuns, indica quantas imagens tiveram pessoas posando de forma não-convencional. Por fim, fotografia original conta quantas imagens tinham problemas impossíveis de serem solucionados já na imagem original, impedindo que qualquer modificação surtisse algum efeito positivo.

Tabela 3.2: Percentual dos principais problemas encontrados no experimento de composição automática de fotografias.

	Uma pessoa	**Grupo**
Fotografia original	10%	8%
Direção do olhar	2%	-
Oclusão parcial	5%	12%
Erro no *headroom*	5%	4%
Poses incomuns	2%	2%
Total de imagens	152	226

Pode-se perceber ao analisar a tabela que a maioria dos erros deu-se já na fotografia original, não havendo correção plausível. Também que a regra-do-zoom é a que mais permite erros, seja por cortar um alvo de forma incorreta seja por permitir um *headroom* incorreto. Isso pode ser explicado pela imperfeição da detecção de faces ou da inferência antropométrica.

Os resultados de ambos os experimentos podem ser melhorados caso o detector de faces utilizado possa sempre retornar corretamente todas as faces contidas na imagem e movimento na câmera fosse permitido, ou seja, outra fotografia pudesse ser obtida de um outro ângulo ou com alguma outra característica alterada.

Com relação ao desempenho do algoritmo, este foi testado em um Athlon XP 2600+ com 512MB de memória RAM, processando imagens com dimensões de 375x500. Os testes de desempenho foram executados em 14 imagens diferentes 10 vezes para cada imagem. O tempo de processamento variou entre 7 e 45ms. Esta variação se dá devido ao número de faces encontradas e devido ao processamento extra necessário para cada uma destas faces. A este tempo ainda é preciso adicionar-se o tempo de detecção de faces. O tempo normal de processamento para um detector de faces é entre 10 e 30ms, permitindo uma taxa de processamento de aproximadamente 7 quadros por segundo. Este protótipo ainda pode ser otimizado visando a melhores taxas.

Algumas imagens utilizadas nos experimentos acima descritos podem ser vistas no Apêndice B.

3.5 Extração de Características

Uma outra abordagem testada com relação à composição fotográfica foi, ao invés de se tentar corrigir uma imagem a partir da informação do posicionamento das pessoas, analizar a qualidade da referida imagem a partir de outras características que pudessem ser extraídas da imagem.

O objetivo é poder classificar uma imagem como sendo boa, aceitável ou ruim com base apenas nestas informações extraídas. Para tanto, foram utilizados alguns algoritmos inspirados no trabalho de Datta et al. [Datta et al., 2006], os quais extraem informação de baixo nível da imagem, como também a abordagem apresentada neste capítulo que utiliza informação de alto nível da imagem, tratando os dados obtidos como informação que pode ser traduzida na qualidade ao invés de um guia para modificação da imagem. Este cálculo requer a localização das faces detectadas, considerada aqui como pré-requisito.

Após os dados serem organizados e normalizados, pode ser possível treinar algum sistema de aprendizagem estatística de forma a fazer automaticamente esta classificação. Esta

seção visa a descrever cada um destes algoritmos de extração de características. Em princípio, não há limite para a quantidade de algoritmos de extração de características. Entretanto, para uma aplicação "*on-line*", é desejável uma quantidade reduzida de processamento. O Apêndice D apresenta uma amostra de imagens e os valores de cada uma das características extraídas de acordo com o descrito nesta seção.

3.5.1 Conformidade à Regra-dos-Terços

A primeira característica a ser analisada é a conformidade do tema fotografado à regra-dos-terços. Como descrito anteriormente, a regra-dos-terços é respeitada quando o tema está posicionado nas retas que dividem a imagem em 3 partes tanto horizontalmente quanto verticalmente e nos pontos de intersecção, pontos estes preferíveis do que apenas posicionar o alvo por sobre as retas.

Com a intenção de considerar a regra de forma mais abrangente, são propostos três medidas, sendo dois primeiros os que calculam a distância às retas horizontais e verticais dos terços e, o terceiro, calcula a distância do alvo ao ponto do terço mais próximo. Foi definido que para fotografias de grupos, será calculada a média das distâncias para os três casos propostos.

As Equações 3.8 e 3.9 mostram como é feito o cálculo da distância da face às retas dos terços para uma ou mais faces. Partindo da Equação 3.5, sendo x e y as coordenadas da face em questão, l a largura da imagem, a a altura da imagem, *primFace* a primeira face, *ultimaFace* a última face da seqüência e $m = tercoMaisProximo(x,y)$, tem-se que:

$$DistanciaMediaTercosH = \begin{cases} | x - \frac{m \times l}{3} | & \text{se } numeroDeFaces == 1 \\ \sum_{face=primFace}^{ultimaFace} \frac{| x - \frac{m \times l}{3} |}{numeroDeFaces} & \text{se } numeroDeFaces > 1 \end{cases} \tag{3.8}$$

O mesmo vale para a Equação 3.9 apenas mudando a orientação da reta a ser buscada.

$$DistanciaMediaTercosV = \begin{cases} | \, y - \frac{m \times a}{3} \, | & \text{se } numeroDeFaces == 1 \\ \sum_{face=primFace}^{ultimaFace} \frac{| \, y - \frac{m \times a}{3} \, |}{numeroDeFaces} & \text{se } numeroDeFaces > 1 \end{cases} \tag{3.9}$$

No qual o terço mais próximo vertical é o terço superior pelos mesmo motivos já explicados anteriormente.

A outra medida que pode ser utilizada é a distância do tema ao ponto dos terços, não importando se ele está, por exemplo, por sobre a reta. O cálculo desta medida também é simples e é dado pela Equação 3.11. Para facilitar a representação, também se define na Equação 3.10 a função pontoMaisPróximo que retorna as coordenadas do ponto de terço mais próximo. Sendo l e a a largura e a altura da imagem respectivamente, tem-se que:

$$PMP(m,n) = pontoMaisProximo(m,n) = (\frac{m \times l}{3}, \frac{n \times a}{3}) \tag{3.10}$$

logo,

$$DistanciaPontoTercos = \begin{cases} d((x,y), PMP(m,1)) & \text{se } numeroDeFaces == 1 \\ \sum_{face=primFace}^{ultimaFace} \frac{d((x,y), PMP(m,1))}{numeroDeFaces} & \text{se } numeroDeFaces > 1 \end{cases} \tag{3.11}$$

Na equação acima, está sendo assumido que a linha horizontal escolhida será sempre a linha horizontal superior. Esta escolha se dá pois quando o tema fotográfico é posicionado na linha horizontal inferior, pode haver um excessivo destaque ao plano-de-fundo. Neste cenário, foge-se do escopo de aplicação desta dissertação que é destinado a imagens que possuem apenas pessoas como alvo.

3.5.2 Conformidade à Regra-do-Zoom

Ao contrário da regra-dos-terços, que possui uma clara meta a ser atingida (que o tema esteja em qualquer um dos pontos dos terços), a regra-do-zoom não possui uma única posição válida. Logo, a menos que se defina qual distância do zoom será utilizada, não é possível avaliar se a regra foi atingida ou não.

Para contornar esse problema e uma vez que o tamanho da face das pessoas é um importante aspecto a ser considerado, optou-se pela alternativa de se extrair o tamanho médio das faces e, a partir daí, verifica-se ou não a conformidade a uma das regras escolhidas. Portanto, sabendo-se as coordenadas superior e inferior de cada face dadas por $y_{superior}$ e $y_{inferior}$ respectivamente, o tamanho da cabeça (*tamanhoDaCabeca*) é dado por:

$$tamanhoDaCabeca = \mid y_{superior} - y_{inferior} \mid \tag{3.12}$$

E, partindo da Equação 3.12, o tamanho médio das faces encontradas, aqui denominado *ZoomMedio*, é facilmente calculado por:

$$ZoomMedio = \sum_{face=primFace}^{ultimaFace} \frac{tamanhoDaCabeca}{numeroDeFaces} \tag{3.13}$$

3.5.3 Conformidade à Regra-da-Integridade

A característica da regra-da-integridade é representada pelo percentual de conformidade a esta regra. Ou seja, tendo-se as coordenadas e dimensões das faces e utilizando-se de medidas antropométricas, pode-se calcular em quais imagens o tema pode estar cortado. Entretanto, não é desejável utilizar esta regra de forma binária, o que implica um cálculo percentual de conformidade, a partir do qual pequenos erros de posicionamento que levem a cortes serão minimamente penalizado enquanto cortes mais significativos serão duramente penalizados.

A análise da conformidade à regra-da-integridade é dividida em vertical e horizontal. A vertical observa se a distância da cabeça aos extremos respeitou um dos padrões de zoom. Caso contrário, um escore (inicialmente com valor máximo, ou seja 1) é decrementado, para cada face da imagem, do percentual da diferença em relação ao nível de zoom mais próximo. Portanto, o cálculo de *Integridade* é efetuado a partir da Equação 3.14.

$$Integridade = 1 - \sum_{face=primFace}^{ultimaFace} \frac{zoomMaisProximo\,() - linhaDaCabeca}{numeroDeFaces} \tag{3.14}$$

Onde *zoomMaisProximo* representa a função que retorna o nível de zoom mais próximo do nível de zoom da face e *linhaDaCabeca* a coordenada inferior da face.

A regra-da-integridade horizontal, similarmente, verifica se alguma das faces não respeitou o limite mínimo de $1,5 \times LarguraDaFace$ entre o canto da face e a borda lateral da imagem. Para cada face que infringir a regra, é diminuído o percentual de não-conformidade. O Algoritmo 3.1 ilustra a regra:

Algoritmo 3.1 Cálculo da conformidade à regra-da-integridade horizontal.

Sendo $x_{esquerdo}$ a coordenada esquerda da face, $x_{direito}$ a coordenada direita da face, $larguraDaFace = | x_{direito} - x_{esquerdo} |$, l a largura da imagem e $excesso_{direito}$ e $excesso_{esquerdo}$ o percentual de espaço insuficiente nas bordas direita e esquerda respectivamente,

para todo $x|x$ é coordenada de face $\in facesNaImagem$

se $(x_{esquerdo} - 1,5 \times larguraDaFace < 0) \vee (x_{direito} + 1,5 \times larguraDaFace > l)$, **então**

Integridade = 1 - $(excesso_{direito} + excesso_{esquerdo})$

fim se

fim para todo

3.5.4 Conformidade à Regra-do-Espaço

Esta regra indica se uma determinada face atinge a distância mínima (semelhante à regra-da-integridade) e não ultrapassa uma distância máxima até as bordas superior e inferior. Desta forma, ela age igualmente à regra-da-integridade. Para simplificar, é indicado, quando há falhas, apenas o espaçamento em excesso. Ou seja, sendo a a altura da imagem, $topoDaCabeca$ a coordenada superior da face, $cantoDaCabeca$ a coordenada inferior da face e o tamanho da cabeça representado por $tamanhoDaCabeca$ (como calculado na Equação 3.12), o espaçamento superior e inferior só estará errado se não forem satisfeitas as Equações 3.15 e 3.16, respectivamente:

$$\frac{(tamanhoDaImagem - topoDaCabeca)}{tamanhoDaCabeca} \geq 1,5 \tag{3.15}$$

$$\frac{(tamanhoDaImagem - cantoDaCabeca)}{tamanhoDaCabeca} \geq 10 \tag{3.16}$$

A taxa de conformidade é medida, portanto, através do cálculo do percentual que foi ultrapassado na regra.

3.5.5 Fotogenia das Faces

Por fim, este algoritmo pode ser utilizado para determinar se as faces detectadas são fotogênicas ou não, de acordo com o trabalho de Batista et al. [Batista et al., 2006]. Nesse artigo, o autor treina uma Rede Neural utilizando informações extraídas por um extrator de características, PCA - Análise dos Componentes Principais (*Principal Component Analysis*), os quais são capazes de determinar, utilizando apenas a imagem da boca, a fotogenia da face. A regra utilizada, a qual indica o estado da face, pode retornar três valores:

$$Fotogenia(face) = \begin{cases} 0 & \text{se face é não-fotogênica} \\ 1 & \text{se face é fotogênica} \\ 2 & \text{se não é possível determinar} \end{cases} \tag{3.17}$$

Para a execução deste algoritmo, é preciso, além do detector de faces comum a todo o sistema, a utilização de um detector de olhos, portanto quando os olhos não são encontrados, a função não é capaz de determinar a fotogenia da face retornando 2.

3.5.6 Classificação Automática de Fotografias

Durante o desenvolvimento das regras de extração de características, foi testada a abordagem de treinar uma Rede Neural a partir dos dados das imagens e de uma série de imagens rotuladas por voluntários como sendo Boas ou Ruins. A intenção da abordagem era verificar se, para um dado conjunto de imagens rotuladas, seria possível indicar quais características são mais determinantes. Desta forma, após treinada uma rede utilizando-se apenas as mais discriminantes das características, esta seria capaz de classificar automaticamente uma fotografia, dentre os rótulos acima apresentados.

Após o estudo do rotulamento a partir das características, pôde-se perceber que a homogeneidade intra-classe não acontece para todas as classes. Este fato pode ser um indício de que os dados não são separáveis, mesmo através do uso de Redes Neurais impedindo a

convergência da mesma. Pôde-se chegar a esta conclusão através do cálculo do Coeficiente de Variação que dá uma medida de como variam os dados intra-classe.

A Equação 3.18 mostra como é calculado o Coeficiente de Variação e a Figura 3.11 o gráfico representando os valores calculados para cada característica nas classes Boas e Ruins, aqui representadas pelas cores azul e vermelha respectivamente. Os dados utilizados são oriundos de alguns extratores de características descritos ao longo desta seção e os demais do trabalho de Datta et al. [Datta et al., 2006].

A mesma Figura 3.11 ainda indica fortes candidatos a serem descartados como a Integridade Vertical (IntVertic), a Integridade Horizontal (IntHoriz) e o valor médio da distância ao ponto dos terços (ThirdsM). Por outro lado, as caracteristicas espaçamento ao topo (RoomT), conformidade à regra-dos-terços horizontal (ThirdsHor) e conformidade à regra dos terços vertical (ThirdsVer) apresentaram uma taxa de homogeneidade maior do que as demais características, indicando possíveis candidatas a formarem um conjunto de treinamento. Os dados são interpretados para cada classe separadamente.

$$CV = \frac{\sigma}{\mu} \tag{3.18}$$

onde σ é o desvio padrão e μ a média dos dados.

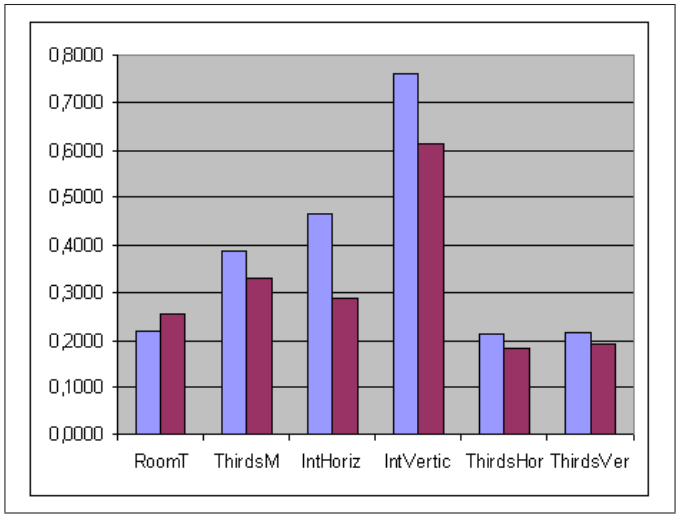


Figura 3.11: Coeficiente de Variação de cada uma das 6 características extraídas para cada uma das duas classes classes (sendo as classes boa e ruim representadas pelas cores azul e vermelha respectivamentes).

A partir deste gráfico é possível perceber-se que a variação intra-classe é pequena. Para

cada uma das características, esperava-se uma maior variação em uma das classes de forma que fosse viável a separação da informação. Variando desta forma, ambas as classes terão comportamento semelhante não havendo características determinantes para a separação dos dados.

Uma outra medida que pode ser utilizada para mostrar como os dados se relacionam é o Coeficiente de Correlação. Assim é possível analisar o quão similares são os dados. Sendo COV calculado como mostra a Equação 3.19, o Coeficiente de Correlação é calculado através da Equação 3.20. A Figura 3.3 mostra a correlação entre os dados.

Na Tabela 3.3, RoomTop corresponde aos dados obtidos pela regra-do-espaçamento, ThirdsM o cálculo do valor médio dos terços, IntegHoriz e IntegVerti correspondem respectivamente à integridade horizontal e a integridade vertical e ThirdsHor e ThirdsVer, correspondem respectivamente aos valores obtidos na análise da conformidade às regras-do-terço considerando as linhas de terço horizontais e verticais.

Os dados apresentados mostram que há uma forte correlação entre, dentre outros, ThirdsM e RoomTop, o que mostra que estes dois dados se assemelham bastante não sendo facilmente separáveis. Poder-se-ia, por exemplo, descartar uma das duas fontes de dado.

$$COV(X,Y) = \frac{1}{n}\sum_{i-1}^{n}(X_i - \mu_x) \times (Y_i - \mu_y) \tag{3.19}$$

$$\rho = \frac{COV(X,Y)}{\sigma_x \times \sigma_y} \tag{3.20}$$

Tabela 3.3: Coeficiente de Correlação entre as características.

	RoomTop	ThirdsM	IntegHoriz	IntegVerti	ThirdsHor	ThirdsVer
RoomTop	1					
ThirdsM	0,18261	1				
IntegHoriz	0,012492	0,090322	1			
IntegVerti	-0,0245	0,018206	0,25867	1		
ThirdsHor	0,122674	-0,48116	0,091371	0,087707	1	
ThirdsVer	0,052869	-0,57745	0,044387	0,025942	0,172451	1

Outras abordagens foram testadas, dentre elas o ID3 [Mitchell, 1999], com a qual se conseguiram melhores resultados, na ordem de 78% de classificação correta. Entretanto, a profundidade da árvore resultante e a variação nos resultados dos treinamentos faz concluir que a generalização foi baixa, não sendo esta abordagem capaz de satisfazer os inúmeros requisitos.

Apesar destes insucessos, esta abordagem não deve ser descartada, apenas sendo necessário um estudo mais aprofundado sobre a metodologia do rotulamento, o qual foi realizado sem delimitação, assim cada pessoa votava de acordo com seu juízo de valor se a fotografia era boa ou não. O problema pode estar neste passo, já que no passo seguinte, a classificação é feita por regras bem definidas as quais os votantes não necessariamente levaram em consideração.

3.6 Considerações Finais

Os resultados parecem promissores, apesar da simplicidade utilizada para executar as regras de composição propostas nesta dissertação. Alguns melhoramentos são necessários para resultados mais rápidos e melhores, especialmente no tocante à detecção e interpretação da ação do alvo, dado que a variação do contexto implica drásticas diferenças no julgamento.

Neste capítulo, foi descrito como, a partir de uma imagem já obtida, pode ser melhorada uma fotografia quanto a sua qualidade visual subjetiva.

Foram apresentados três módulos de composição, regra-dos-terços, integridade e regra-do-zoom e um módulo para controle dos três além de solução de conflitos.

Os experimentos foram realizados com a intenção de validar a robustez do sistema. Para isto, o sistema foi executado em dois conjuntos de 1327 e 378 imagens respectivamente, com o fim de verificar se a qualidade é mantida para um conjunto de imagens de boa qualidade, como também se a qualidade pode ser elevada, quando se dispõe de imagens de baixa qualidade. Como resultado, tem-se que 65% dos voluntários votantes, concordaram que as modificações feitas pelo algoritmo surtiram efeito positivo nas imagens.

Por fim, foram apresentados algoritmos para extração de características, os quais podem efetuar cálculos sobre o comportamento de uma determinada regra para uma imagem, dispondo apenas de sua face.

A aplicação destas regras em um ambiente dinâmico no qual é possível interagir com o cenário é mostrado no capítulo seguinte no qual será mostrado como realizar o mesmo trabalho de composição fotográfica a partir de uma câmera móvel modificando o ambiente antes da fotografia ser obtida.

Capítulo 4

Proposta de uma Plataforma para Fotografia Autônoma

4.1 Introdução

Neste capítulo, é proposto um sistema capaz de controlar remotamente uma câmera com o objetivo de localizar e fotografar um determinado tema disposto em um ambiente não-controlado. Igualmente ao restante desta dissertação, o tema em questão sempre são pessoas. A utilização de uma câmera controlável dá a possibilidade de correção prévia da composição fotográfica podendo, inclusive, ajudar a evitar erros que não seriam passíveis de correção após a fotografia ter sido obtida.

No Capítulo 3, foram propostas algumas correções na composição que podem ser realizadas após a fotografia ser obtida. Entretanto, algumas outras correções poderiam também ser realizadas, caso fossem permitidas alterações antes mesmo da fotografia ser obtida. Como exemplo, pode-se citar o corte de pessoas. Após a fotografia ser obtida, apenas pode-se modificar o corte a fim de que este pareça menos agressivo, contudo não é possível obter a parte da imagem cortada.

Caso a detecção fosse feita antes da fotografia, seria possível mover a câmera de forma que este corte fosse evitado. Dentro do presente contexto, é apresentada neste capítulo uma abordagem que permite que erros sejam detectados antes mesmo da fotografia ser obtida para, sempre que possível (e desejável) estes erros possam ser evitados.

Para isso, é preciso um hardware de fotografia específico, capaz de se movimentar de

acordo com o erro detectado, a fim de corrigi-lo. No protótipo desenvolvido, é utilizada uma câmera *Pan-Tilt-Zoom* (PTZ), ou seja, uma câmera que possui mecanismo que lhe dá três graus de liberdade, permitindo rotação horizontal e vertical além da mudança do ângulo de visão. Com esta câmera, é possível corrigir alguns dos problemas mais comuns encontrados nas fotografias obtidas por fotógrafos inexperientes.

O sistema assim proposto pode ser aplicado em diversas ocasiões, conforme discutido a seguir. A primeira é a de um ponto fixo de fotografia capaz de fotografar um evento, à medida que ele ocorre, mesmo que as pessoas participantes deste evento não estejam posando para esta fotografia.

A segunda ocasião possível é a busca por elementos detectados visualmente em um ambiente fixo, como uma busca por uma logomarca em um espaço.

Como uma terceira aplicação, existe a opção da pose intencional, na qual a câmera fotografa pessoas que estão posando para uma fotografia. Além das três citadas anteriormente, inúmeras outras aplicações semelhantes, ou a partir de pequenas variações da abordagem proposta, podem surgir.

No problema aqui apresentado, serão feitas restrições que devem ser observadas, a fim de se delimitar o escopo desta abordagem. Em primeiro lugar, todas as restrições feitas no Capítulo 3 continuam válidas. O problema é semelhante, apenas sendo tratado de uma forma diferente. Em seguida, deve-se delimitar que a câmera não é dotada de movimento espacial, ou seja, ela não pode locomover-se pelo ambiente na qual ela é colocada, mas apenas girar em torno de seu próprio eixo. Logo, assume-se que a câmera é posicionada na direção que as pessoas a serem fotografadas estarão em uma distância que seu zoom seja capaz de alcançar.

Também não é informado ao sistema o posicionamento do alvo, logo a câmera precisa encontrar o alvo por si só (sendo satisfeita a condição anterior). A localização não é feita a partir de nenhuma interação entre pessoas e a câmera, sendo essa localização realizada pela detecção de faces. Dadas as limitações, portanto, pessoas não detectadas serão tratadas como plano-de-fundo.

Para a simulação e experimentos, os quais serão fiéis às três aplicações aqui propostas, considera-se que o ambiente é um salão qualquer (sala de aula, auditório, etc) contendo pessoas ou imagens visíveis de pessoas, representando os alvos da fotografia que o sistema busca. Prioritariamente, deseja-se a correção do posicionamento da câmera. Entretanto,

outros indicadores de qualidade podem ser agregados. O objetivo, portanto, é obter uma grande quantidade de imagens maximizando-se a qualidade, a qual é medida de acordo com as métricas utilizadas pelos módulos agregados ao sistema.

A próxima seção contém uma descrição da câmera utilizada e como esta é controlada. Na Seção 4.3, são apresentados os principais problemas relacionados à localização de temas em uma cena. Posteriormente, na Seção 4.4 é mostrado como, a partir da detecção de pessoas na cena, estas podem ser enquadradas e fotografadas. Na Seção 4.5, é descrito como foram realizados os experimentos que validaram o sistema proposto. Em seguida, na Seção 4.6, são descritas algumas propostas de aplicação para o sistema descrito e, por fim, na Seção 4.7, apresentam-se as conclusões deste capítulo.

4.2 Funcionamento da Câmera

Para o controle da câmera, foi desenvolvida uma API (do inglês *Application Programming Interface*) em C++, capaz de simplificar a utilização das principais funcionalidades da câmera, tais como obtenção e ajuste dos graus de rotação (vertical e o horizontal), obtenção e ajuste do ângulo de visão, captura de imagens da câmera, etc.

O modelo VB-C50i da marca Canon, utilizado neste experimento, é capaz de fornecer imagens a uma taxa aproximada de 30 quadros por segundo, sendo apresentado na Figura 4.1. Essa taxa varia de acordo com a resolução da imagem (que pode ter as dimensões 160x120, 320x240 e 640x480) e com a utilização de algoritmos de detecção de faces e processamento de imagens. Devido ao fato de as imagens serem transmitidas via Ethernet, essa taxa pode variar em decorrência ao tráfego da rede.

Há também a necessidade de conversão dos dados transmitidos para um formato manipulável. Neste caso, a imagem é transmitida no formato JPEG. Portanto, é necessário um algoritmo que possa realizar a decodificação da imagem e a conversão para o formato de imagem utilizada pelo sistema proposto, o qual utiliza a biblioteca CImg [CImg, 2005]. Para este fim, foi utilizada a biblioteca libjpeg [IJG, 2005] para abrir a imagem diretamente após feita a cópia da câmera para o disco. Com relação ao ângulo de visão deste modelo, sua maior abertura possui 41,26° e a menor 1,97°.

A comunicação com a câmera é realizada através de chamadas a funções, enviadas como

Figura 4.1: Modelo de câmera utilizado - Canon VBC-50i.

requisições HTTP. Entretanto, o envio de comandos para câmera opera de modo não bloqueante, desta forma, logo após o envio de uma requisição para a câmera, o sistema continua em operação independentemente do tempo que a câmera leva para receber a requisição e para realizar a ação. Contudo, o módulo de movimentação e o módulo de captura da câmera operam de forma paralela, podendo haver captura enquanto há movimentação e vice-versa. Isso permite que o sistema continue capturando imagens na medida em que o movimento é processado.

A desvantagem, contudo, é que esta característica pode atrapalhar a detecção de objetos pois eventualmente serão capturadas imagens apresentando uma imagem "borrada" em decorrência da captura da imagem ter sido realizada com o obturador aberto por tempo suficiente para capturar a movimentação da câmera, também conhecido como *motion blur* [Parker, 1997]. Para resolver este problema pode-se usar estratégias de espera de forma que o sistema fique em espera ocupada durante o tempo estimado (dado que o modelo de câmera utilizado não possui sinalização para fim de movimento) enquanto a câmera conclui sua movimentação.

Outra desvantagem desta metodologia de captura de imagens é que o hardware da câmera não é programado para atualizar o valor de seu posicionamento enquanto está em movimento. Portanto, se durante o movimento algum objeto de interesse é detectado, não é possível saber com exatidão qual sua posição nem interromper o movimento. Se a câmera recebeu a ordem, por exemplo, de partir de sua origem, rotacionar 20° (em qualquer direção) e, durante a rotação, alguma face for detectada nas imagens capturadas enquanto ainda ocorria a rotação da câmera, não é possível interromper a rotação ou obter a posição em que a câmera estava quando a imagem foi capturada por limitações do próprio equipamento.

4.3 Localização de Alvos na Cena

A localização do tema consiste em, posicionada a câmera em um ambiente, delimitar o espaço de busca e encontrar um tema de interesse para enquadramento e posterior captura de imagens utilizando uma maior resolução do sensor de captura.

A busca por um alvo, contudo, depende imensamente de qual cenário o sistema está imerso, informação esta que muitas vezes não é passível de auto-detecção, sendo uma escolha que cabe ao usuário.

Não há uma forma de se procurar um alvo em uma imagem que seja mais lógica do que as outras quando não se tem informação alguma sobre aquela região. Logo, não há no início da execução do sistema, nenhuma estratégia que possa ser adotada para aumentar a quantidade de alvos encontrados. Portanto, se nenhuma restrição for feita, pode-se afirmar que em todo o espaço coberto pelo campo de visão da câmera podem haver pessoas.

Dentro do presente contexto, em se tratando de um sistema autônomo de procura de pessoas, em primeiro lugar deve-se procurar por pessoas para, em seguida e em se tendo os resultados desta pesquisa, restringir a área de busca, dando mais prioridade a regiões nas quais os temas foram encontrados enquanto as regiões procuradas sem sucesso são desencorajadas a novas buscas. Essa alternativa, entretanto, é arbitrária, pois em não se conhecendo o cenário, não é possível afirmar que a cena não irá mudar do conhecimento já obtido. O mais interessante é que o usuário possa definir qual a faixa de atuação do sistema e qual a freqüência de atualização da cena.

A Figura 4.2 ilustra o funcionamento do módulo Pan-Tilt-Zoom. Apos inicializar a câmera em uma posição, ajusta-se a mesma para cada face encontrada de forma que as pessoas encontradas, estejam em conformidade com as regras de composição de fotografias implementadas nos módulos.

A seguir, nas próximas três sub-seções, são detalhados os possíveis cenários nos quais o sistema pode ser imerso, como pode ser feita a detecção do tema e, posteriormente à detecção, como calcular a movimentação necessária para um enquadramento apropriado.

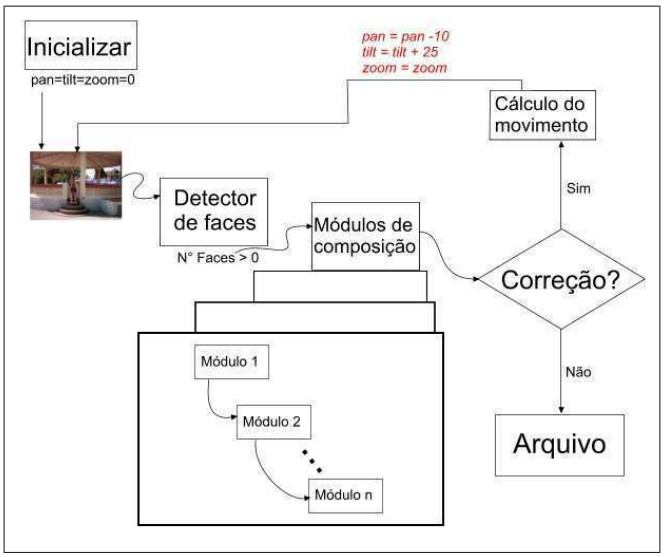


Figura 4.2: Diagrama do funcionamento do módulo PTZ.

4.3.1 Possíveis Cenários

Um passo importante para o funcionamento adequado do sistema é se ter ciência de qual cenário o sistema está interagindo. Poder-se-ia fazer uma interpretação automática do cenário, contudo esta é uma área de pesquisa ainda sendo estudada, dada a sua complexidade decorrente da quantidade de situações possíveis, ficando esta parte do sistema como trabalho futuro.

Para efeitos desta pesquisa, foi decidido pela fixação de três possíveis cenários. No primeiro cenário, chamado estático, considera-se que no ambiente não haverá mudanças nem na quantidade de alvos nem em suas localizações durante a execução do programa. Já em um segundo cenário, semi-dinâmico, pode-se permitir mudanças de pequena amplitude, entretanto, considera-se que o número de alvos não variará durante a execução do sistema. Nestes dois primeiros cenários, considera-se que onde um tema foi procurado e não foi encontrado, este não será encontrado no futuro.

Por fim, na proposta de um cenário dinâmico, há de se esperar quaisquer tipos de mudanças. Para tanto, é necessária a utilização de um fator de desatualização, ou seja, por quanto tempo - ou passos do sistema - uma dada região pode ser considerada atualizada, ou seja, por quantos passos, considera-se que a informação de presença ou não de temas é precisa.

Se a escolha for por um cenário estático: (i) não haverá desatualização, ou seja, a informação descoberta na procura em uma cena perpetua-se como correta durante a execução do

sistema; (ii) o alvo não se move; (iii) a quantidade de alvos permanece inalterada dentro do ângulo de visão da câmera.

Em um cenário semi-dinâmico: (i) não haverá desatualização e (ii) o alvo pode mover-se, desde que dentro do ângulo de visão da câmera na posição onde ele foi inicialmente detectado; (iii) o número de alvos na cena não variará.

Já em um cenário dinâmico, está implícito que: (i) mudanças podem ocorrer no cenário, apenas sendo necessária a indicação do tempo em que o cenário se modifica, (ii) o alvo pode mover-se livremente, inclusive para fora do campo de visão da câmera e (iii) podem também surgir novos alvos.

Após definido o cenário, pode-se, portanto, partir para a etapa de detecção do tema já que agora sabe-se como ele deve se comportar na cena.

4.3.2 Detecção do Tema

Neste sistema, para que um tema possa ser localizado é necessário que sejam utilizadas apenas as imagens capturadas pelo dispositivo em questão. No sistema exposto, a localização do tema é feita a partir da detecção de faces, por ser esta uma evidência da presença de pessoas em uma cena.

Além da detecção de faces, pode também ser utilizada a evidência do movimento através da detecção de movimentos [Latzel et al., 2005], a exemplo da implementação disponível no pacote da OpenCV [OpenCV, 2005]. Este, contudo, é opcional, pois apenas a detecção de faces por si só já é uma evidência da presença de pessoas.

Para a detecção de faces, qualquer detector de faces pode ser utilizado. O sistema proposto nesta dissertação foi testado com dois detectores de faces diferentes: o detector de faces Rowley-Kanade [Rowley et al., 1998a; Rowley et al., 1998b] e o detector de faces OpenCV [OpenCV, 2005].

Ambos variam em portabilidade, quantidade de falsos positivos, quantidade de falsos negativos e velocidade da detecção. Em geral, o Rowley-Kanade Face Detector é o que possui melhor desempenho considerando-se essas quatro métricas de qualidade.

Já o OpenCV, apesar de ser o mais portável (considerando-se a facilidade para se integrar a um sistema linux) e um pouco mais veloz (conseguindo taxas de quadros por segundo maiores que as do Rowley-Kanade) detecta muitos falsos positivos e a variação das dimensões

de uma mesma face detectada é grande, enquanto o concorrente é mais estável.

Uma outra vantagem do detector da OpenCV é que este dispõe de um módulo de treinamento utilizando o algoritmo de busca pela *integral image* e classificador *Adaboost* [Viola and Jones, 2001], no qual é possível indicar qual alvo deseja-se localizar, sendo possível, por exemplo, treinar o detector para ao invés de detectar faces, efetuar a detecção de outros objetos.

Apesar de ser um grande indicativo da presença de pessoas, a detecção de faces possui desvantagens com relação a outros métodos. Em primeiro lugar, devido ao fato de a detecção de faces ser lenta em relação à velocidade de movimentação de pessoas. Em segundo lugar, devido à dificuldade deste detector localizar pessoas caso a face não esteja completamente visível, seja por uma oclusão causada por outro elemento da imagem, seja devido à rotação da pessoa na cena. Portanto, este método é mais indicado para ambientes nos quais, sendo o tema pessoas, elas movimentem-se pouco ou lentamente e a possibilidade de oclusão seja reduzida, e.g.: salas de aula, auditórios, teatros, etc. pois, normalmente, os espectadores estão olhando para um ponto fixo, não havendo significativa movimentação ou objetos que possam causar oclusão.

Outras abordagens já estão em desenvolvimento para lidar com estes problemas, tal como a utilização de detectores de movimento (disponível no pacote OpenCV[OpenCV, 2005]) e a utilização de detecção de pontos de atenção visual [Itti et al., 2005; Pereira and Gomes, 2006; Pereira et al., 2006], os quais podem ser utilizados como um guia, apesar de que, em algumas ocasiões, os algoritmos de atenção podem marcar pontos longe do alvo, pelo fato deste não ser o elemento de maior compiscuidade na cena, resultando em instabilidade no sistema.

Em alguns cenários, as pessoas poderiam ser fotografadas pela câmera, se esta tivesse grau de rotação maior ou possibilidade de movimentação. Byers et al. [Byers et al., 2003] propõem um sistema capaz de localizar pessoas em um cenário utilizando sensores infravermelhos e informação de tonalidade de pele para, em seguida, movimentar-se no cenário a fim de obter uma fotografia frontal.

Portanto, a abordagem aqui apresentada considera que há apenas um detector de faces a ser aplicado em imagens obtidas por uma única câmera PTZ, sem ajuda de sensores externos, outras câmeras ou de possibilidade de movimentação pelo ambiente.

4.3.3 Representação do Conhecimento

Para representar a informação sobre a cena, como os lugares recentemente procurados, as áreas mais desatualizadas, a localização das pessoas, dentre outras informações passíveis de serem encontradas, foi criada uma estrutura de representação. Para essa estrutura de representação, utiliza-se uma imagem cujas dimensões são fiéis à área de busca, sendo cada 0,5 grau representado por um *pixel* da imagem, na qual a coordenada da câmera é simbolizada pelo centro da imagem que esta captura. Dado que a câmera pode girar 200° e 120° horizontalmente e verticalmente respectivamente, a imagem deveria ter dimensões de 400 x 240. Entretanto, quando o centro da câmera aponta para os ângulos de *Pan* ou *Tilt*, ainda existe uma área capturada pela câmera que não seria representada, equivalente a metade do ângulo de visão da câmera. Logo, a imagem deve ter dimensões de 480x300.

A representação é feita em três canais, nos quais cada canal representa uma faixa de zoom de 33% do total. Esta escolha foi feita arbitrariamente com a intenção de se reduzir a faixa de valores para três magnitudes, de maneira que se possa representar a cena em curta, média e longa distância.

Por simplicidade e tirando proveito de estruturas típicas de representação de imagens coloridas, o zoom mínimo é representado pelo canal vermelho, o zoom médio, pelo canal verde e o zoom máximo pelo canal azul. A cada procura em uma região, os *pixels* da região de tamanho do ângulo de visão atual centralizados nas coordenadas relativas aos valores de *Pan* e *Tilt* da câmera, são marcados com o valor 255. Os níveis de zoom precisam ser separados pois uma face encontrada em um nível de zoom, seja ele menor ou maior, pode não ser detectada nos outros dois níveis devido a proximidade ou distância do tema à câmera.

A cada movimentação da câmera, é feita uma redução da intensidade dos *pixels* em 3 níveis, para os três canais considerando. Isto é realizado para representar a desatualização do conhecimento sobre a cena, ou o aumento da incerteza. Esta abordagem de representação do conhecimento da cena foi escolhida pois é de fácil representação, visualização, modificação e ainda processamento, dado que algoritmos de processamento de imagens como projeções e limiares podem ser utilizados diretamente para obter a finalidade escolhida.

Também existe a vantagem de se poder utilizar qualquer abstração de imagem existente que possua estes algoritmos além de outros relacionados como desenho de retângulos, por exemplo. As faces e outros ítens localizados em separado, os quais deseja-se comporta-

mento diferente podem ser representados através de uma segunda imagem, dedicada exclusivamente para tal.

A Figura 4.3 contém um exemplo da representação do conhecimento através destas imagens. Na imagem representando a incerteza de uma cena, a região preta é a da qual não se tem nenhum conhecimento. As cores isoladas (vermelho ou verde) indicam uma procura com apenas um nível de zoom. As cores combinadas (no exemplo, amarelo) indicam a procura em mais de um nível de zoom. Os pontos brancos, opcionais, representam a certeza da existência de uma face naquele ponto.

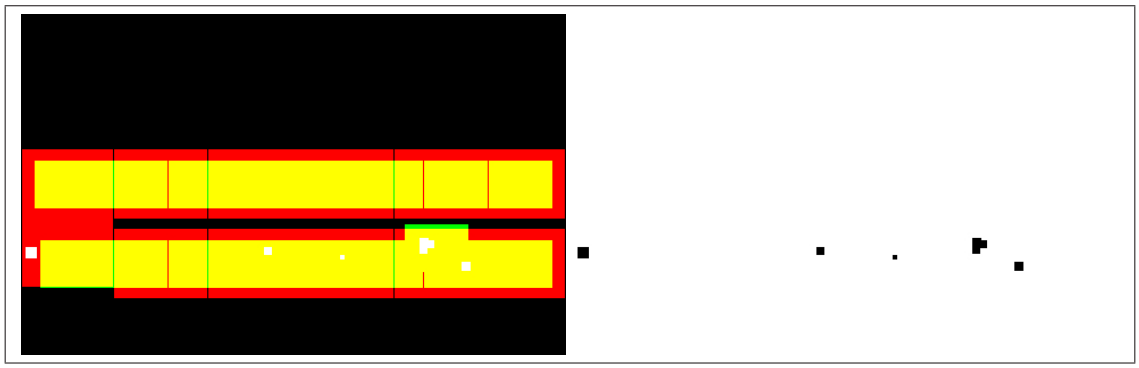


Figura 4.3: As imagens acima ilustram como é representado o conhecimento da certeza da cena (imagem superior) e da posição dos temas detectados (imagem inferior). As cores, na imagem superior, representam as buscas utilizando ângulos de visão menores que 33° (vermelho), entre 33° e 66° (verde), entre 0° e 66° (amarelo) e as faces localizadas (branco).

De posse destas imagens que representam a informação obtida na cena, é trivial encontrar-se alguns pontos de interesse. Através da imagem que representa as faces detectadas pode-se, utilizando projeção, saber onde estão as faces, aproximar a quantidade e qual o ângulo de visão que deve ser usado para enquadrar estas faces. Além disso, com base nesta informação, pode-se procurar nas mesmas alturas em locais diferentes por novas faces ainda não detectados, sustentado pelo princípio que, onde existe uma pessoa, as outras provavelmente seguirão um padrão de comportamento, como altura e distribuição. A representação do conhecimento da cena também pode ser utilizada para, em calculando-se áreas mais escuras da imagem, saber quais as regiões nas quais o conhecimento está mais desatualizado para assim buscar por temas nesta área.

A forma de representação utilizada nada mais é do que uma abstração visual da utilização

de matrizes para representação do mesmo conhecimento, podendo, sem perdas, ser utilizada esta outra forma de representação e armazenamento dos dados encontrados.

4.3.4 Estratégias de Busca por Alvos

Independentemente de qual cenário venha a ser utilizado, no problema abordado neste capítulo é considerado que, inicialmente, não há informação alguma sobre a possível localização do alvo na cena. Portanto toda e qualquer direção, em princípio, pode ter um alvo a qualquer distância da câmera. Dada esta condição, faz-se necessária a adoção de uma estratégia para procurar, com eficiência, pelo alvo em toda a extensão da cena.

Para tanto, são propostas quatro diferentes estratégias de busca para a localização de pessoas: estática, randômica, losangular e panorâmica, estratégias estas que devem satisfazer os cenários anteriormente descritos. Em todos os casos, o objetivo é encontrar alvos. Assim que um alvo é detectado, a câmera move-se para enquadrar este alvo. Mais detalhes de como é realizado o movimento para que o alvo seja enquadrado, é descrito adiante na Seção 4.4.1.

A primeira estratégia de busca, a estática, consiste em posicionar a câmera de forma que ela permaneça apontada para uma direção específica e, na medida em que localiza alvos, efetuar o enquadramento. Na medida em que se movimenta, a câmera pode perder o alvo, quer seja por falha no detector, quer seja pelo deslocamento do alvo, o qual pode sair do ângulo de visão da câmera. Quando isto acontece, a câmera volta para sua posição inicial. Esta estratégia é indicada para ambientes nos quais os alvos estão concentrados em uma região sempre contida dentro do campo de visão da câmera, como exemplo de uma sala de aula ou auditório estreitos e, não muito profundos. O Algoritmo 4.1 lista os passos desta estratégia de busca. A idéia é que o algoritmo repita o procedimento de fotografia indefinidamente ou até que alguma métrica seja atingida. No algoritmo descrito a métrica é o número de fotografias obtidas.

A estratégia de busca randômica consiste em efetuar pequenas movimentações na câmera, sendo a direção e a intensidade do vetor de movimentação determinados de forma aleatória. Esta estratégia é indicada para situações nas quais o alvo esteja disperso pelo ambiente e movimente-se sem direção ou velocidade bem definidas. A desvantagem é a possibilidade de uma grande perda de tempo em virtude da desordem da busca, dada a grande quantidade de possibilidades. Esta estratégia de busca é mostrada no Algoritmo 4.2.

Algoritmo 4.1 Estratégia de busca estática.

#Inicializar o contador de fotografias e mover a câmera para a posição inicial

câmera.setPanTiltZoom(0 , 0 , 41°26');

i=0;

#Executar até que um certo número alvo de fotos númeroDeFotos seja atingido

enquanto (*i* <númeroDeFotos), **faça**

#Marcar a hora atual

horaInício = horaAtual();

#Capturar a imagem da posição atual

Img = *câmera*.fotografar();

#Procurar por alvos na imagem da posição atual

ListaDeAlvos = procurarAlvos(*Img*);

#Para cada alvo encontrado, enquadrá-lo, fotografá-lo em alta definição e salvar a imagem

para todo $Alvo \in ListaDeAlvos$

câmera.setPanTiltZoom(enquadrar(*Alvos*));

$foto$ = *câmera*.fotografarAltaDefinicao();

$foto$.salvar(toString(*i*++).".jpg");

fim para todo

se (horaAtual() - *horaInício* > $limiteDeTempo$), **então**

câmera.setPanTiltZoom(0 , 0 , 41°26')

fim se;

fim enquanto;

Algoritmo 4.2 Estratégia de busca randômica.

```
#Iniciar a câmera com os valores (0,0) e ângulo de visão mais amplo
câmera.setPanTiltZoom( 0 , 0 , 41°26' );
i=0;
#Executar até que um certo número alvo de fotos númeroDeFotos seja atingido
enquanto ( i <númeroDeFotos ), faça
    #Fotografar a cena
    Img = câmera.fotografar();
    #Procurar por alvos na imagem da posição atual
    ListaDeAlvos = procurarAlvo(Img);
    #Para cada alvo encontrado, enquadrá-lo, fotografá-lo em alta definição e salvar a
    imagem
    para todo Alvo ∈ ListaDeAlvo, faça
        câmera.setPanTiltZoom( enquadrar( Alvo ) );
        foto = câmera.fotografarAltaDefinição();
        foto.salvar( toString( i++ )".jpg");
    fim para todo
    #Caso o tempo limite seja ultrapassado ou nenhuma pessoa for encontrada, mover a
    câmera aleatoriamente
    se ( horaAtual() - horaInício > limiteDeTempo ) ∨ ( ListaDeAlvos == ∅ ), entao
        NovoPan = câmera.panAtual() + aleatório( 0,0 , 20,0 );
        NovoTilt = câmera.tiltAtual() + aleatório( 0,0 , 10,0 );
        NovoZoom = câmera.zoomAtual() + aleatório( 0,0 , 5,0 );
        câmera.setPanTiltZoom( NovoPan , NovoTilt , NovoZoom );
    fim se;
fim enquanto;
```

Por sua vez, a estratégia de busca losangular descreve uma trajetória não-aleatória movendo a câmera de forma a ter os pontos médios do retângulo que representa o ângulo de visão completo da câmera, tais como pontos de partida e destino. O termo de busca losangular, portanto, vem do fato de a câmera percorrer uma trajetória semelhante a um losango. A diferença vem do fato de a câmera não ser simétrica, podendo girar mais na direção superior (90°) do que inferior (30°).

Na primeira execução, o retângulo é o determinado pelo ângulo de visão total da câmera. Já nas execuções subseqüentes este ângulo é reduzido em se definindo novas fronteiras limite para a câmera, fronteiras estas mais próximas ao centro. Esta estratégia de busca tem o objetivo de cobrir a região do ângulo de visão mais propícia a ter pessoas, ou seja, a região central.

Na Figura 4.4, ilustra-se o percurso que a câmera segue quando descreve a estratégia de busca losangular. Vale observar a redução do espaço de busca do primeiro passo (setas pretas) para o segundo passo (setas vermelhas) ilustrado aqui pelas marcas cinzas.

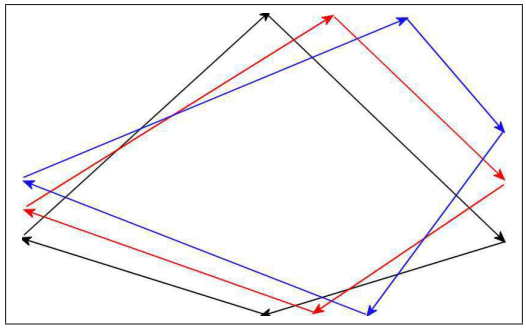


Figura 4.4: Três passos de busca losangular (primeiro preto, segundo vermelho). A área disponível é reduzida no segundo passo (área cinza representa locais que não serão mais visitados).

O Algoritmo 4.3 mostra detalhadamente a busca em forma de losango. Por fim, a estratégia de busca panorâmica faz um "retrato panorâmico" da cena, através da junção de imagens isoladas capturadas pela câmera que se assemelham à imagem que seria obtida caso fosse utilizada uma câmera com lente especial para permitir um grande ângulo de visão (neste caso, de mais de 200 graus) com a vantagem de sofrer menor distorção na imagem.

Para a obtenção da imagem panorâmica, o procedimento inicia-se ao se apontar a câmera para seu mínimo horizontal e vertical e variar sua rotação de forma que cada imagem cor-

Algoritmo 4.3 Estratégia de busca losangular.

#Iniciar a câmera com os valores (0,tilt_mínimo) e ângulo de visão mais amplo

câmera.setPanTiltZoom(0,*tilt_mínimo*,41°26'); *i*=0; *deslocamento*=0;

#Executar até que um certo número alvo de fotos númeroDeFotos seja atingido

Enquanto ($i <$*númeroDeFotos*), **faça**

para todo $z \in \{41°26', 28°17', 15°08', 1°97'\}$, **faça**

para todo $(x, y) \in \{$ (*pan_mínimo*, *deslocamento*), (*deslocamento*,*tilt_máximo*), (*pan_máximo*,*deslocamento*), (*deslocamento*, *tilt_mínimo*) $\}$, **faça**

$NúmeroDePassos = \frac{|camera.panAtual() - x|}{camera.AnguloDeVisao()}$

#Dividir o percurso entre origem e destino em passos iguais, fotografar a cena e armazenar os alvos encontrados na imagem

direçãoX = sinal(*x* - *PanAtual*); *direçãoY* = sinal(*y* - *TiltAtual*);

para todo $m \in \mathbb{N} : m <$*NúmeroDePassos*, **faça**

$NovoPan = PanAtual +$direçãoX $\times$*câmera*.ÂnguloDeVisão();

$NovoTilt = TiltAtual +$*direçãoY* $\times$*câmera*.ÂnguloDeVisão() $\div 1,3333$;

câmera.setPanTiltZoom(*NovoPan* , *NovoTilt* , *z*);

Img = *câmera*.fotografar();

ListaDeAlvos = *ListaDeAlvos* + procurarAlvo(*Img*);

fim para todo;

#Para cada alvo encontrado, enquadrá-lo, fotografá-lo e salvar a imagem

para todo *Alvo* $\in$ *ListaDeAlvos*

câmera.setPanTiltZoom(enquadrar(*Alvo*));

foto = *câmera*.fotografarAltaDefinição();

foto.salvar(toString(*i*++).".jpg");

fim para todo;

fim para todo;

deslocamento += *câmera*.ÂnguloDeVisão();

fim enquanto;

responda a uma região diferente. As variações horizontal e vertical devem ser equivalentes ao ângulo de visão da câmera de maneira que ao rotacionar-se a câmera nesta intensidade, a imagem deve ser imediatamente vizinha à imagem anterior.

Entretanto, devido à curvatura da lente da câmera e à irregularidade que pode existir no motor de rotação da câmera, a fotografia panorâmica realizada desta forma possui imperfeições, como pode ser visto adiante. Existem algoritmos de retificação da imagem que podem ser utilizados para minizar os erros decorrentes dos problemas apresentados [Sinha and Pollefeys, 2004]. Estes algoritmos, contudo, são custosos computacionalmente e usualmente precisam de uma grande quantidade de fotografias para sua calibração, fugindo ao escopo do problema aqui apresentado.

Pode-se fazer um paralelo com a busca realizada nesta dissertação com a movimentação de robôs em uma área. Em robótica, esta metodologia de busca é utilizada em situações onde não se tem nenhuma certeza sobre o ambiente. Como exemplo, tem-se os trabalhos descritos em [Stack and Smith, 2003] que referem-se a busca por minas aquáticas, tendo como principal diferença o fato de minas não serem colocadas depois de iniciada a busca e também do fato das minas aquáticas geralmente serem igualmente espaçadas. Esse estilo de busca é conhecido como varredura em linha (*line sweep*), porém também é conhecido como padrão do cortador de grama (*lawn mower*) ou padrão do disseminador de sementes (*seed sowing*).

Outros padrões de movimentação de robôs semelhantes ao proposto, podem ser encontrados nos artigos de Qiu et al e Yang & Luo [Qiu et al., 2006; Yang and Luo, 2004], sendo estes dois últimos ajudados por uma rede neural que é utilizada para calcular dinâmicamente o caminho de movimentação. Por questões de escopo e simplicidade, o cálculo do caminho neste trabalho é realizado de forma simples, calculando-se a menor distância entre os pontos.

Na Figura 4.5, mostra-se um ambiente de exemplo capturado pela câmera. O Algoritmo 4.4 lista os passos desta estratégia de busca.

A Câmera PTZ utilizada informa o ângulo de visão que está sendo utilizado, desta forma para qualquer nível de aproximação é possível se fazer uma fotografia panorâmica.

Na Figura 4.6, ilustra-se como é realizado o movimento no algoritmo anteriormente descrito.

A estratégia de busca panorâmica possui vantagens e desvantagens. A principal vantagem

Algoritmo 4.4 Estratégia de busca panorâmica.

#Criar imagem que receberá as fotografias das partes do ambiente

$ImgPanorâmica \longleftarrow \{\forall px \in ImgPanorâmica \mid px = 0\}$

Enquanto ($i <$*númeroDeFotos*), **faça**

#Procurar em níveis específicos de zoom

para todo $z \in \{41°26', 28°17', 15°08', 1°97'\}$, **faça**

#Iniciar a câmera apontando para seu canto inferior esquerdo

câmera.setPanTiltZoom(*pan_mínimo* , *tilt_mínimo* , z);

direcaoY = 1;

para todo $x \in \mathbb{R}$:*pan_mínimo* $\leq x \leq$*pan_máximo*, **faça**

#Descrever a trajetória panorâmica

para todo $y \in \mathbb{R}$: *pan_minimo* $\leq y \leq$ *pan_maximo*, **faça**

montaImagemPanorâmica(*câmera*.fotografar() , *ImgPanorâmica*);

$y = y +$*direçãoY* $\times$ *câmera*.ÂnguloDeVisão()$\div 1,3333$;

câmera.setPanTiltZoom(x , y , z);

fim para todo;

$x = x +$*câmera*.ÂnguloDeVisão();

#Fotografar durante o caminho de volta da câmera

direçãoY = direçãoY $\times -1$;

fim para todo;

#Buscar alvo e, em seguida, enquadrar e fotografar cada alvo

ListaDeAlvos = procurarAlvo(*ImgPanorâmica*);

para todo $Alvo \in ListaDeAlvos$, **faça**

câmera.setPanTiltZoom(enquadrar(*Alvos*));

$foto$ = *câmera*.fotografarAltaDefinição();

$foto$.salvar(toString(i++).".jpg");

fim para todo;

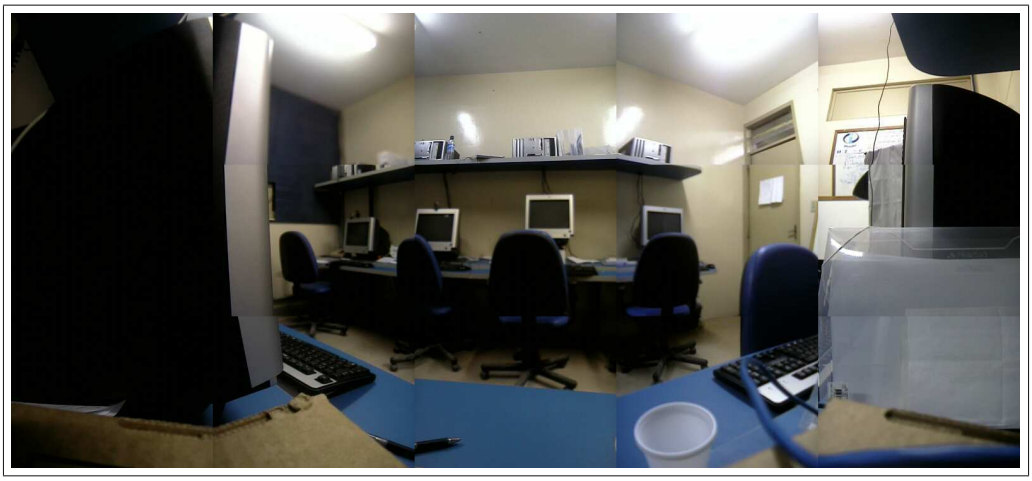


Figura 4.5: Foto panorâmica com variação horizontal de 43°25’ e variação vertical de 32°.

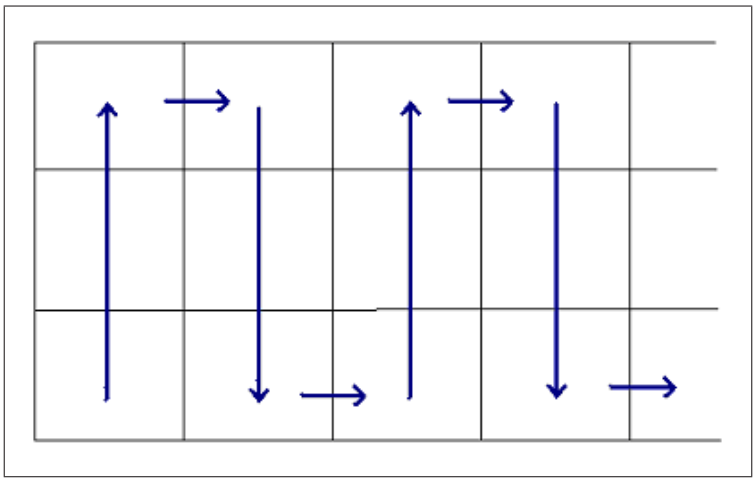


Figura 4.6: Movimentação feita pela câmera de forma a reduzir o número de movimentações.

é a possibilidade de cobrir toda a cena de forma ordenada. A principal desvantagem é a lentidão com que esta busca é feita, já que todos os pontos precisam ser cobertos. Esta desvantagem, contudo, só é relevante quando considera-se ambientes de alto dinanismo. Também, devido à uniformidade da pesquisa, algumas regiões ficam sem serem fotografadas pela câmera um longo período de tempo, o que não acontece em outras abordagens.

Após feita a fotografia panorâmica, pode-se localizar os alvos e, reduzir o espaço de busca por localizar regiões onde estão localizados os temas de interesse. Independente do método de procura utilizado para localizar em uma cena temas de interesse, o passo seguinte, o de enquadramento, é realizado de forma similar e é detalhado na Seção 4.4.1 que vem a seguir.

Além destas estratégias, outras podem ser utilizadas além de combinações dessas. Basta que existam informações que possam ser úteis na detecção de pessoas. Está sendo desenvolvida, para trabalhos futuros, uma outra estratégia que utiliza detector de movimento e filtro da cor da pele como evidência da presença de pessoas.

Comparação dos Métodos

Alguns métodos de busca, devido a suas características, são difíceis de serem comparados diretamente. Portanto, nestes casos, será feita uma análise do tempo máximo de busca por um alvo nestas condições. A idéia é mostrar as dificuldades e possíveis virtudes de cada método de busca a fim de que cada um possa ser utilizado da forma em que melhor se adapte.

Em primeiro lugar, é analisado o método estático. Este método é voltado para situações onde se sabe exatamente onde podem surgir temas, para só em caso de algum alvo ser encontrado, haver movimentação visando ao enquadramento e à fotografia. Portanto, é difícil comparar este método com outros métodos, devido à amplitude de situações em que o sistema pode ser utilizado, pois se o alvo estiver na direção da câmera ele sempre será encontrado enquanto se não estiver ele nunca será encontrado. Portanto este método não é indicado para cenários onde as pessoas estejam distribuídas pelo ambiente no qual a câmera é inserida.

O método randômico, por sua vez, tem sua dificuldade de comparação com os outros métodos descritos já que, em tese, toda execução deste método deveria gerar uma nova combinação de trajetórias, podendo existir, portanto, mais de uma trajetória, inviabilizando uma comparação direta.

Entretanto, pode-se fazer uma estimativa dos tempos mínimo e máximo necessários em uma determinada situação. Por exemplo, no caso de se ter uma única face no ambiente, o algoritmo randômico pode levar no mínimo um passo, caso a movimentação seja precisa e inifinitos passos, caso não seja realizado um registro dos pontos já visitados, ou poderá levar a quantidade de passos necessária para cobrir toda a área, caso seja feito um um registro.

Logo, o algoritmo de busca randômica, no pior caso, gastaria o mesmo tempo que a busca panorâmica. Entretanto, necessitaria de maior processamento para cálculo das regiões mais indicadas além de que a desordem da procura provocaria uma grande perda de tempo na movimentação do mecanismo da câmera. Por fim, há também a desvantagem da existência

de lacunas não procuradas o que não acontece no método panorâmico.

Para os dois métodos seguintes, pode ser realizado um teste em comum, de forma a compará-los diretamente. Na simulação de comparação os algoritmos são iniciados no mesmo ponto e as faces são posicionadas no mesmo ponto. É definida uma taxa de atualização da cena. Para efeito de comparação, foram utilizadas duas métricas. A primeira é a intensidade da atualização e a segunda, a dimensão de espaço desatualizado.

Para analisar a quantidade de área atualizada, a região de busca foi dividida em quatro quadrantes. A análise destes quadrantes pode indicar qual estratégia é mais promissora quando o principal fator é a quantidade de área pesquisada.

Os resultados mostram que o algoritmo de busca losangular possui a vantagem de se distribuir uniformemente pela imagem, não deixando um certo quadrante desatualizado por muito tempo. Pode-se concluir que se a cena for muito dinâmica, este deve ser o método mais indicado. Por outro lado, o algoritmo de busca panorâmico faz com que alvos sejam procurados na cena de maneira mais uniforme e completa, mais indicado para cenas com menor grau de dinamismo, onde o posicionamento das pessoas não variará ou variará pouco. Uma outra vantagem é que, independente do nível de desatualização utilizado, a cena estará em média 20% coberta no ângulo de visão de 41°26', calculo este efetuado utilizando as imagens contruídas com a informação da procura.

4.4 Composição e Fotografia do Tema

Após a localização do tema, independentemente de qual estratégia de busca tenha sido utilizada, o próximo passo é a fotografia do mesmo. Poder-se-ia apenas fotografar o alvo, entretanto o objetivo deste trabalho é a obtenção de fotografias dotadas de qualidade visual, logo faz-se necessária uma movimentação extra para um melhor posicionamento ou composição fotográfica.

Portanto, após localizado o alvo, foi definido que é essencial ao menos a centralização deste alvo na fotografia. Em se havendo condições, também é feita a composição dos elementos encontrados na imagem, o que pode gerar novos deslocamentos.

Esta seção vem tratar do processamento necessário para se obter uma fotografia de maior qualidade. Em primeiro lugar, a movimentação para um melhor posicionamento do alvo na

fotografia, em seguida a fotografia em si e, por fim, um pós-processamento da imagem para últimos retoque e/ou classificação da imagem por ordem de qualidade.

4.4.1 Estratégias de Movimentação

Após detectado um alvo, existe a necessidade de se fazer o enquadramento deste alvo para que ele ocupe uma determinada posição na imagem. Apesar de se conhecer a distância em pixels entre dois pontos quaisquer de uma imagem, essa informação apenas não é suficiente para que seja efetuado um correto enquadramento, pois não necessariamente existe correspondência entre o número de pixels e a rotação em graus da câmera. Isso se dá especialmente quando também se variam os níveis de zoom, uma vez que para cada novo nível há uma nova regra de rotação. Portanto, o cálculo tanto da rotação horizontal como da vertical deixa de ser trivial. Em não se tendo esta informação (situação mais comum), faz-se necessário o cálculo desta distância.

Alguma informação da qual se disponha seja qual for o nível de zoom é necessária, de forma a utilizá-la no cálculo da rotação correta. O ângulo de visão da câmera pode ser uma destas informações. Conhecendo-se o ângulo de visão da câmera e a distância do objeto de interesse ao centro da imagem, possibilita o preciso cálculo do ângulo de rotação necessário para que o alvo seja centralizado em um único passo. Entretanto, algumas câmeras podem não fornecer a informação sobre o ângulo de visão. É mostrado como, considerando-se estas duas possibilidades, pode-se fazer esta centralização do alvo na imagem.

O cálculo do ângulo de rotação, quando se conhece o ângulo de visão da câmera, é trivial pois pode ser efetuado através de trigonometria básica. Na Figura 4.7, ilustra-se o problema. Em (a) as linhas em preto representam o ângulo de visão da câmera. Em (b) é ilustrada a metade do ângulo de visão e em (c) o triângulo formado pela distância da face ao centro da imagem, a linha que liga a câmera ao objeto e a reta que liga ortogonalmente a câmera ao plano onde se encontra o tema encontrado. A seguir, é mostrado como é efetuado este cálculo.

Partindo da Figura 4.7, sabe-se que:

$$tg\beta = \frac{x}{z} \tag{4.1}$$

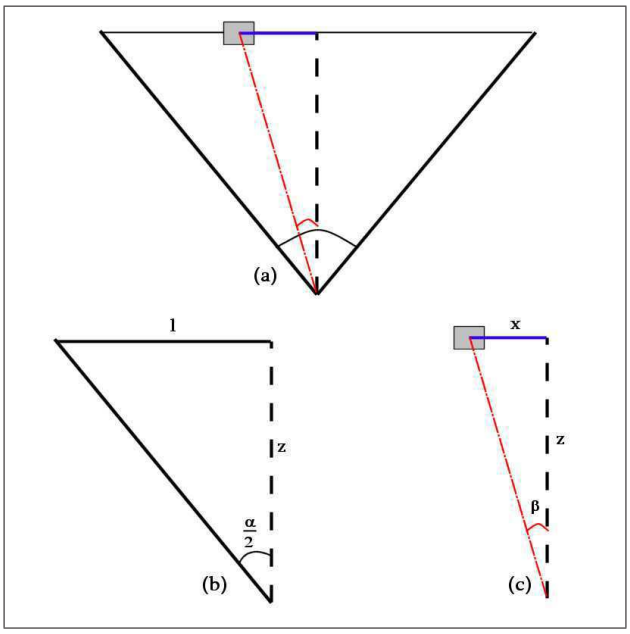


Figura 4.7: O triângulo (a) ilustra o campo de visão da câmera e um alvo detectado. Em (b) e (c) o triângulo (a) é fragmentado.

e

$$tg\frac{\alpha}{2} = \frac{l}{z} \tag{4.2}$$

onde x é a distância do alvo ao centro da imagem, z é a distância entre a câmera e o plano onde encontra-se o tema, l é a distância percorrida por metade do ângulo de visão, no caso metade da dimensão da imagem, α o ângulo de visão e β o ângulo necessário para o centro da câmera aponte diretamente ao alvo.

Considerando-se a distância z a mesma nas duas equações acima, tem-se que:

$$tg\beta = \frac{tg\frac{\alpha}{2} \times x}{l} \tag{4.3}$$

Como deseja-se obter o ângulo de rotação que centraliza o alvo, deduz-se:

$$\beta = arctg\left(\frac{x \times tg\frac{\alpha}{2}}{l}\right) \tag{4.4}$$

A mesma equação é utilizada para o cálculo da rotação vertical.

Em não se tendo informação alguma além da distância do centro do alvo à posição de interesse, é preciso estimar-se a movimentação necessária através de outras informações da

imagem.

Uma primeira proposta é inferir o ângulo de rotação da câmera para se chegar ao alvo através de um treinamento utilizando a estimativa do tamanho da face como informação de distância, dado que quanto mais longe da câmera, menor é a face. De posse destes dados, a construção de uma *Lookup Table* poderia ser suficiente para calcular situações similares. Apesar de bastante simples, essa estimativa não produz bons resultados pois o tamanho da face retornado pelo detector pode apresentar variações o que resulta em erros maiores em diferentes níveis de zoom. Como conseqüência dos problemas desta abordagem, tem-se que as faces são constantemente perdidas ou o zoom é excessivo ou insuficiente.

Uma outra alternativa é a utilização da visão estéreo, a qual permite, com a utilização de uma outra câmera, inferir a distância com base na triangulação de pontos similares nas duas imagens. Esta estratégia, entretanto, é de difícil implementação pois requer uma segunda câmera calibrada[1] e sincronizada[2] com a primeira. Portanto, devido às dificuldades envolvidas em se utilizar esta abordagem, ela sequer chegou a ser implementada.

Por fim, a abordagem utilizada nesta dissertação é inspirada no comportamento humano. Quando um ser humano resolve apontar uma câmera para um local, não o faz de forma precisa em um único passo, apesar da alta velocidade com que processa a informação visual (comparado com um computador) dar essa impressão [Hendee, 1997]. Portanto, o primeiro ajuste é realizado sem preocupação *a priori* com a precisão e na medida que se aproxima do objetivo, calcula os novos ajustes nos quadros subseqüentes. A velocidade do processamento é um indicador de quão precisos podem ser os ajustes, pois quanto mais ajustes puderem ser realizados, mais o ponto ideal há de se aproximar. Se o processo for muito lento, contudo, o ajuste não é tão preciso dada a necessidade de realizar as demais tarefas do sistema.

Para tanto, é necessário, em primeiro lugar, calcular a distância do centro da imagem ao centro do tema. Para melhor precisão e maior controle do movimento, a distância é calculada de forma independente para a componente horizontal e a componente vertical. As Equações 4.5 e 4.6 mostram como obter este valor.

[1] O processo de calibração consiste em eliminar os erros decorrentes da curvatura da câmera e de encontrar os pontos correspondentes entre as duas imagens

[2] A sincronização se dá quando duas ou mais câmeras obtém imagens simultâneamente

$$distancia\_face\_centro\_horizontal = coordenada\_x\_alvo - \frac{dimensao\_horizontal}{2} \quad (4.5)$$

$$distancia\_face\_centro\_vertical = coordenada\_y\_alvo - \frac{dimensao\_vertical}{2} \quad (4.6)$$

em que $distancia\_face\_centro\_horizontal$ e $distancia\_face\_centro\_vertical$ são as distâncias do centro horizontal e vertical da face ao centro da imagem respectivamente, a largura da imagem é representada por $dimensao\_horizontal$ e $dimensao\_vertical$ é a altura da imagem.

As Equações 4.7 e 4.8 equivalem à distância normalizada de cada componente ao centro da imagem, onde 1 equivale a maior distância possível (ou seja, quando o centro do alvo estiver por sobre a borda da imagem já 0 equivale a uma coincidência entre o centro da imagem e o centro do alvo (não havendo necessidade de movimentação). Desta forma, se pretende que a movimentação seja proporcional à distância do centro da face ao centro da imagem. Nas equações a seguir, $posicao\_h\_alvo$ e $posicao\_v\_alvo$ são, respectivamente, as distâncias do centro da face ao centro da imagem normalizadas e separadas horizontalmente e verticalmente.

$$posicao\_h\_alvo = \frac{distancia\_face\_centro\_horizontal}{dimensao\_horizontal} \quad (4.7)$$

$$posicao\_v\_alvo = \frac{distancia\_face\_centro\_vertical}{dimensao\_vertical} \quad (4.8)$$

Já as Equações 4.9 e 4.10 mostram como é feita a estimativa do ângulo de rotação necessário para se centralizar a face horizontalmente e verticalmente respectivamente. As rotações calculadas para pan e tilt são, respectivamente, $angulo\_horizontal$ e $angulo\_vertical$.

$$angulo\_horizontal = fator\_correcao\_pan \times posicao\_h\_alvo \quad (4.9)$$

$$angulo\_vertical = fator\_correcao\_tilt \times posicao\_v\_alvo \quad (4.10)$$

Os parâmetros $fator\_correcao\_pan$ e o $fator\_correcao\_tilt$ indicam a velocidade do movimento horizontal e vertical respectivamente. Com a intenção de simplficação, estes valores serão constantes e receberão os valores 10,00° e 5,00°, respectivamente.

Como forma de reduzir a quantidade de ajustes, é diminuída a precisão do posicionamento, permitindo que a face diste do centro da imagem até um certo valor de tolerância. Para esta aplicação foi definido heuristicamente a distância de 20 *pixels* em se utilizando imagens de 160x120. Este valor é arbitrário e, em se seguindo uma etapa de posicionamento, pode ser alto.

A nova posição da câmera é, portanto, representada pela Equação 4.11 a qual é denotada por *valor* onde α e β equivalem à variação do ângulo da câmera com relação a sua posição inicial (centralizada) o qual recebe valor 0:

$$(pan\_atual + \alpha, tilt\_atual + \beta) \tag{4.11}$$

No entanto, com a variação do zoom, esta rotação não mais é a mesma pois, enquanto uma rotação de 1,00° quase não produz diferenças visuais para a imagem original quando o zoom possui valor mínimo (ou seja, sem zoom) em um cenário de alto valor de zoom (a câmera de modelo VBC50i utilizada possui Zoom máximo de 26x) a mesma variação de 1,00° gera uma imagem bem diferente da original. Portanto, o cálculo da rotação precisa ser influenciado de forma inversamente proporcional pelo valor do zoom.

Com esta finalidade, utiliza-se o fator_zoom, o qual normaliza o valor do zoom entre 0 e 1 onde valores próximos de 0 são obtidos quando o zoom da câmera estiver próximo de seu máximo. Já na situação oposta, o fator_zoom tem um valor próximo de 1. Desta forma, se o zoom estiver próximo de seu valor máximo, a rotação tende a 0 (há uma adição para evitar que não haja movimento). Sendo *Zoom_Maximo* e *Zoom_Atual* os valores máximo e atual do Zoom respectivamente, a equação completa do fator_zoom é a Equação 4.12.

$$fator\_zoom = \frac{Zoom\_Maximo - Zoom\_Atual + 10}{Zoom\_Maximo} \tag{4.12}$$

Portanto, para o cálculo da rotação da câmera, precisa-se adaptar esta nova variável calculada na Equação 4.12 às equações anteriormente calculadas, 4.9 e 4.10. Com o fator_zoom, as equações finais de ângulo do movimento passam a ser as ilustradas nas Equações 4.13 e 4.14:

$$angulo\_horizontal = fator\_zoom \times fator\_correcao\_pan \times posicao\_h\_alvo \tag{4.13}$$

$$angulo\_vertical = fator\_zoom \times fator\_correcao\_tilt \times posicao\_v\_alvo \quad (4.14)$$

Pode ser necessário variar o fator_zoom, de forma a suavizar seu ajuste para algumas câmeras, em decorrência da diferença entre os níveis de zoom ou a própria não-linearização da aproximação, ou seja, situações onde um zoom de 4x não equivala exatamente ao dobro de 2x, por exemplo.

4.4.2 Composição do Tema Fotográfico

Após localização e enquadramento, o próximo passo é a composição da fotografia, mais precisamente do tema. No Capítulo 3, apresentou-se uma discussão detalhada sobre o que é a composição fotográfica e o porquê de sua utilização. A composição do tema com a câmera PTZ se dá de forma análoga. Como já foi dito, entretanto, existem regras que não são passíveis de correção dinâmica.

Outro agravante é a movimentação do alvo ou tema. Em dependendo do rigor adotado para a correção de uma dada regra, a câmera pode ter trabalho dobrado caso o movimento do alvo seja na mesma direção e intensidade do movimento previsto pela câmera. Desta forma, a correção, por ser lenta já que utiliza a detecção de faces como informação primordial sobre o posicionamento do alvo, pode gerar resultados desagradáveis.

Foram implementados no controle da câmera as seguintes regras: regra-do-zoom, regra-dos-terços e regra-da-integridade.

A regra-dos-terços funciona de forma semelhante ao algoritmo para enquadramento, tendo como principal diferença o fato de o tema não vir a ser centralizado e sim deslocado para um dos terços da imagem, logo o deslocamento é efetuado utilizando um dos pontos do terço como objetivo.

A regra-do-zoom e da Integridade, foram implementadas em conjunto. As fotografias serão obtidas, para cada face encontrada, em três distâncias diferentes: a distância máxima, em uma distância média e em close-up.

4.4.3 Fotografia do Alvo

Esta etapa é responsável em efetuar a fotografia e enviá-la ao sistema para pós-processamento seguido da classificação. No sistema proposto, a própria câmera utilizada para localizar e enquadrar, também é responsável por efetuar a fotografia. Poderia ser acoplada a este sistema, uma câmera de maior resolução (já que o modelo utilizado só permite uma resolução máxima de 640x480 pixels). Isto, contudo, causaria diferenças entre as duas imagens no tocante ao posicionamento do tema na imagem. Uma outra opção seria utilizar uma câmera PTZ com maior resolução máxima.

São adquiridas 10 fotografias em seqüência sem novas análises do posicionamento do alvo. Com isso, espera-se aumentar o número de fotografias, já que não há um passo de checagem a cada nova foto antes de ela ser obtida. Por outro lado, também se aumenta a quantidade de fotografias sem o posicionamento correto do alvo, portanto, carentes de filtragem. Esta abordagem, contudo, também é a utilizada por fotógrafos profissionais, muitas fotografias são obtidas para que depois seja feita uma filtragem das melhores fotografias.

As fotografias são armazenadas em um diretório à parte e renomeadas de forma a marcarem a seqüência. Durante a etapa de fotografia, pode ser dado o zoom na face localizada a fim de obter novos níveis de composição, em se dependendo da distância atual à face. Em existindo várias faces, repete-se o processo para cada face encontrada na imagem.

4.4.4 Pós-processamento da Imagem

Os algoritmos descritos no Capítulo 3 serão utilizados com o fim da filtragem de baixo nível das fotografias, utilizando o escore obtido por cada fotografia para que as fotografias sejam classificadas de forma ordenada. As fotografias que atingirem mais altos níveis de qualidade serão filtradas manualmente. Apesar de lento, o processo pode produzir fotografias de alta qualidade cuja pequena diferença de frações de segundo poderiam evitar a obtenção.

Além dos classificadores já descritos, pode ser utilizado um classificador de fotogenia para qualificar a fotografia final. A face enquadrada é classificada em fotogênica ou não-fotogênica. Esta classificação pode tanto ser *on-line*, i.e. realizada antes de a fotografia ser obtida, como *off-line*, i.e. na base de fotografias obtidas.

A vantagem da classificação on-line é a possibilidade de se decidir em qual momento se

obter a fotografia, diminuindo-se a quantidade de fotografias obtidas (já que apenas as classificadas como boas é que serão armazenadas) por outro lado permite que ocorram pequenos atrasos, em decorrência do processamento da imagem. Esses atrasos influenciam na qualidade final da imagem, pois pode ser armazenada uma imagem cuja cena foi ligeiramente alterada do momento da última fotografia classificada. Podem também existir atrasos para a obtenção de novas fotos - caso a classificação seja feita posteriormente à fotografia.

O módulo de classificação de fotogenia foi descrito no artigo de Batista et al. [Batista et al., 2006]. O objetivo é discernir quanto à fotogenia, ou seja, o quão agradável pode ser a fotografia de uma pessoa. Na Figura 4.8, ilustra-se o funcionamento do detector de fotogenia que, neste exemplo, circunscreve a face com um retângulo verde quando considera a face fotogênica ou com um retângulo vermelho caso contrário. O objetivo deste módulo não é indicar beleza física, mas a aparência em uma dada circunstância.

O classificador de fotogenia em questão foi treinado com imagens de faces previamente rotuladas através da avaliação manual de AU's (*Action Units*[3]) na qual uma dada combinação de AU's representa uma expressão facial. Para compor a base de pessoas fotogênicas, foram utilizadas imagens nas quais as expressões faciais foram rotuladas como alegria ou neutras. Os outros rótulos compuseram a base de faces não-fotogênicas.

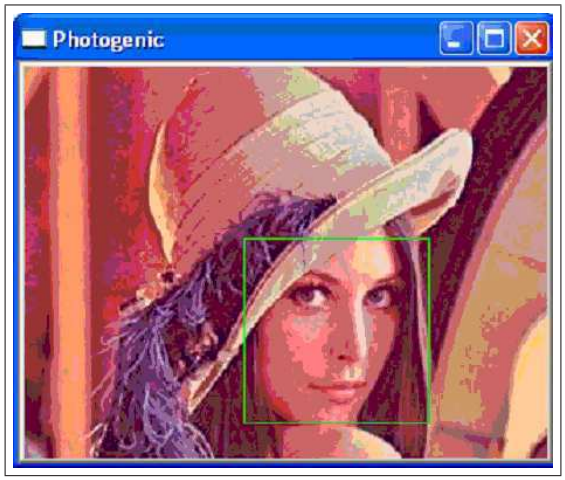


Figura 4.8: Exemplo de classificação quanto à fotogenia.

A interação entre o módulo de fotogenia e o sistema é feita através dos resultados da classificação do módulo de fotogenia nas faces da imagem. Portanto, se todas as faces ou a

[3]Os músculos da face são divididos em Action Units. Cada unidade pode ser excitada independentemente.

maioria delas apresentarem resposta positiva quanto à fotogenia, a imagem é guardada. Caso contrário, descartada ou colocada em um local diferente para reavaliação do usuário.

4.5 Experimentos

Pode-se dividir a etapa de experimentos em duas. Na primeira etapa, testou-se cada parte do sistema individualmente, conferindo se este atua da forma indicada. Na segunda etapa, testou-se o sistema com todas as suas partes reunidas em um ambiente real para validar o experimento como um todo.

O experimento do sistema completo, por si só, já seria suficiente para validar o experimento, até porque se alguma das partes não funcionasse corretamente, não seria possível obter o resultado correto. Ainda assim, é feita uma rápida descrição do que foi utilizado para validar os experimentos.

4.5.1 Testes das Partes do Sistema

Os experimentos serão descritos na ordem em que os correspondentes algoritmos foram apresentados neste capítulo. A primeira parte do sistema, o funcionamento da câmera, está fora do escopo do trabalho. Ainda assim, foi realizado um teste de obtenção de imagens e movimentação da câmera, conferindo se todas as movimentações indicadas são de fato realizadas pela câmera.

A busca por alvos em uma cena foi simulado através da criação de uma classe que atua semelhantemente à câmera. A diferença é que esta simulação sempre retorna a mesma imagem, com exceção da(s) posição(ões) escolhida(s) para conter(em) face(s), em que ela retorna uma imagem contendo uma face. Desta forma é possível analisar-se a trajetória e confirmar a corretude do algoritmo.

Também urge que seja analisado o enquadramento do alvo, pois as pequenas diferenças que podem existir devido à falta de calibração da câmera no momento da rotação poderiam resultar em um ajuste errôneo. Para testar a calibração do motor, várias fotos foram adquiridas com variação de $0{,}01^\circ$ e, em seguida, com variações maiores para verificar se a diferença na rotação era significativa.

Para testar o enquadramento, foram realizados testes em várias distâncias, onde haviam

movimentações horizontais e verticais, conferindo se para qualquer nível de zoom ou qualquer ângulo de visão o comportamento da câmera era satisfatório. Os testes foram realizados utilizando-se o valor do ângulo de visão e simulando-se o não conhecimento, para efeito de validação dos algoritmos de cálculo da movimentação necessária.

4.5.2 Experimentos Utilizando o Sistema Completo

Foram realizados alguns experimentos com todo o sistema integrado para verificar o comportamento do sistema quando este se depara com uma situação real, na qual as pessoas, não tendo o retorno do processamento, não interferem significantemente no sistema. Os experimentos a seguir foram realizados com a autorização dos coordenadores dos ambientes, entretanto como as pessoas fotografadas, mesmo sabendo que estavam sendo fotografadas, não autorizaram explicitamente a publicação de suas imagens, optou-se por não mostrá-las nesta dissertação.

Dentre os experimentos realizados, um ocorreu em uma sala de aula contendo aproximadamente 15 alunos, os quais foram informados do experimento. Neste experimento, realizado durante o dia, a iluminação era adequada para reconhecimento de faces. A câmera foi posicionada na mesma altura das pessoas e no ambiente não existiam elementos de oclusão. Foi utilizada para enquadramento a *Lookup Table*, com o sistema acoplado a um detector de fotogenia. Para faces próximas à câmera, o funcionamento foi adequado. Já para faces distantes da câmera, o enquadramento não foi preciso. Foram capturadas em torno de 20 imagens.

Em um segundo experimento, controlado, com pessoas olhando diretamente para a câmera em uma sala estreita, utilizando o enquadramento dinâmico e sem o detector de fotogenia acoplado, foram capturadas em torno de 40 imagens. Por ser um experimento realizado em um ambiente controlado, ou seja, as pessoas sabiam do experimento e na medida do possível cooperaram com o bom funcionamento deste, não se pôde chegar a resultados conclusivos, tendo sido mais útil como um experimento para avaliar partes do sistema como a detecção de faces e o enquadramento.

Um terceiro experimento, realizado em um largo salão, no qual as pessoas eventualmente transitavam, foi utilizado o enquadramento dinâmico e sem o detector de fotogenia acoplado. Foram capturadas menos de 10 imagens, devido à lentidão da detecção de faces. Pôde-

se concluir com este experimento que a velocidade de detecção de faces é muito aquém da velocidade das pessoas. Portanto, são necessárias estratégias para restringir o espaço de busca e de algoritmos de detecção mais velozes. Neste experimento, ao contrário dos anteriores, vários locais poderiam ter faces em diferentes ângulos com relação a câmera. Esta variedade de posições e ângulos que cada face pode assumir em diferentes momentos adiciona um componente de dificuldade, uma vez que o sistema não é capaz de delimitar, precisamente, um espaço de busca.

O quarto experimento, ocorreu em uma sala de aula larga, na qual com 10 alunos e um professor que, eventualmente, posicionava-se entre a câmera e os alunos, propiciando um ambiente mais real. As pessoas sabiam do experimento. Em aproximadamente 40 minutos de funcionamento, foram obtidas em torno de 15 fotografias que foram, posteriormente, aprovadas pela audiência. Esta experiência utilizou a busca randômica e o posicionamento do tema através da regra-dos-terços.

Foi detectado, contudo, necessidade de modificações no módulo de movimentação (diminuindo a correção do zoom enquanto é corrigido o posicionamento horizontal), no módulo de enquadramento e fotografia (de forma a permitir que a câmera insista mais em um ponto que estava perto de efetuar a fotografia). Apesar de que melhorias no sistema levaram a detecções mais velozes, o enquadramento de pessoas distantes da câmera continua sendo um problema uma vez que este enquadramento é vital ao sistema

O experimento final, foi a integração de todos os aspectos discutidos ao longo desta dissertação. Foi realizado em uma sala de aula contendo aproximadamente 15 alunos. Mais de 2000 fotografias foram obtidas. Este experimento é detalhado no Capítulo 5.

Com base nos experimentos realizados, alguns problemas puderam ser percebidos. Dentre os principais problemas, a detecção de faces tem sido mais lenta que a velocidade das pessoas. Conseqüentemente, para ambientes amplos onde pessoas caminhem, o detector de faces é insuficiente para efetuar corretamente o *tracking*.

4.6 Aplicações

O sistema proposto pode ser utilizado em uma série de situações. São destacadas três aplicações.

Uma primeira possibilidade é utilizá-la para fotografia de um evento sem que os participantes saibam do equipamento, obtendo fotografias do ambiente como um todo. Esta é a opção para qual o sistema descrito neste capítulo foi idealizado.

Uma segunda aplicação é a localização de elementos conhecidos em um ambiente. Treina-se qual a aparência do objeto a ser procurado e, em seguida, procura-se minuciosamente em toda a cena por ocorrências deste objeto. Para adaptar o sistema aqui proposto para esta abordagem, basta: (i) treinar o sistema para deteção de um outro item; (ii) modificar o cenário para estático, adquirindo-se assim fotografias da cena completa, sem a utilização de valores randômicos para a escolha do ponto a ser fotografado. (iii) alterar a quantidade de planos de visão de maneira a cobrir ainda mais minuciosamente a cena.

Por fim, pode-se utilizar a câmera como ponto de fotografia onde, sabendo que lá existe uma câmera fotográfica, pessoas posam para fotografias. Deve-se, portanto, deixar a câmera parada a espera de alvos. À medida que estes forem localizados, serão fotografados tal como proposto neste capítulo. Uma foto de grupo, nesta ocasião, implica que as pessoas precisam de tempo para posar. Uma alternativa é a de temporizar a partir do momento que a primeira face é encontrada. Uma segunda alternativa é utilizar um algoritmo para detecção de alguma informação visual para indicar a espera para a câmera. Como proposta, sugere-se o reconhecimento de gestos ou sinais além do algoritmo de Viola e Jones [Viola and Jones, 2001] para treinar o reconhecimento de elementos em uma imagem, podendo treinar, por exemplo, um símbolo que ao ser detectado indica para a câmera que ela deve permanecer sem fotografar em alta definição.

4.7 Considerações Finais

Neste capítulo, foi apresentado um protótipo otimizado para controle de uma Câmera *Pan-Tilt-Zoom*. Neste protótipo, a câmera é posicionada em um local estratégico para localização, enquadramento e captura de imagens dos alvos presentes na cena.

Para a movimentação, foram propostos três mecanismos de movimentação, onde cada qual ajusta seus parâmetros de forma a enquadrar alvos detectados no ambiente em que a câmera está localizada.

No tocante ao enquadramento, é proposta uma nova versão de enquadramento utilizando

o ajuste dinâmico, em que o ajuste é realizado com base nos valores detectados na última imagem capturada ao invés de um único ajuste efetuado com base em uma *Lookup Table*. O ajuste é mais lento, contudo, menos sensível à distância do alvo para a câmera.

Também foi proposta neste capítulo a utilização dos módulos de composição fotográfica com a intenção de aumentar a qualidade da imagem do ponto de vista subjetivo. Ainda nessa direção, também foi proposta a integração do sistema a um módulo de detecção de fotogenia, o qual pode indicar se as pessoas presentes em uma fotografia estão apresentando uma pose fotogênica ou não.

Capítulo 5

Experimento de Integração e Resultados

Neste capítulo, apresenta-se o experimento realizado com a intenção de dar maior suporte à validação da proposta como um todo. Outros experimentos foram apresentados em capítulos anteriores. Entretanto, o experimento aqui relatado integra características de cada parte do sistema descrito anteriormente.

O experimento aqui descrito foi realizado em uma sala de aula contendo 14 pessoas as quais não receberam quaisquer instruções específicas de posar para as fotos, ou seja, as fotografias foram obtidas de forma totalmente espontânea.

O experimento permitiu localizar as pessoas, guardar a informação de onde se encontravam e, utilizando esta informação, obter fotografias da cena, as quais deviam representar de forma satisfatória, de acordo com uma audiência, o evento.

5.1 Descrição do Experimento e Objetivo

O objetivo deste experimento foi testar o sistema de fotografia autônoma proposto nesta dissertação de maneira que, mesmo sem nenhuma informação prévia a respeito do ambiente, fosse possível localizar e fotograr as pessoas que estavam no local do experimento e, ao mesmo tempo, gravar um mapa do ambiente. Este mapa pode ser utilizado posteriormente na mesma cena, de forma que os temas previamente localizados possam ser fotografados sem que houvesse necessidade de uma nova etapa de localização.

5.1.1 Local do Experimento

Para que este experimento fosse realizado, foi preciso, antes de mais nada, definir o seu ambiente de execução. O local escolhido foi uma sala de aula contendo 14 pessoas, sendo elas 12 alunos, 1 professor e 1 operador do sistema. A sala continha diversos objetos, tais como cadeiras, computadores e mesas, contudo, a disposição e tamanho destes não chegaram a posar como obstáculos para o experimento.

A maior dificuldade com relação à geometria do lugar foi a pequena largura da sala, o que aumentou a proximidade entre as pessoas e fez com que aquelas localizadas mais à frente na sala fossem detectadas de forma desproporcional com relação às localizadas mais ao fundo, já que por muitas vezes estas últimas eram encobertas pelas primeiras. A ilustração da posição aproximada dos elementos na sala é mostrada na Figura 5.1. As pessoas estão representadas por retângulos pretos, enquanto a posição da câmera está representada pelo retângulo cinza.

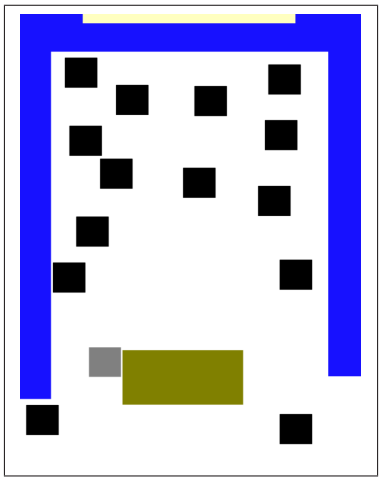


Figura 5.1: Vista superior do ambiente, representando a disposição de pessoas e objetos no espaço escolhido para promover o experimento. Os retângulos azuis representam bancadas de computadores, o retângulo marrom uma mesa, os quadrados pretos pessoas e o quadrado cinza a câmera.

5.1.2 Cenário

Considerou-se o cenário como semi-estático. Logo, as pessoas se movimentavam por um certo raio, mas não entravam nem saíam do ambiente. Dada esta restrição, o experimento precisou ser parado e reiniciado por algumas vezes, devido à entrada tardia de alguns alunos na sala-de-aula.

Para melhor utilização do tempo e à pedido dos alunos e professora que gentilmente se voluntariaram para o experimento, o grau de rotação vertical foi restringido, impedindo que a câmera rotacionasse e apontasse para abaixo da lilnha da cintura das pessoas. Esta decisão favoreceu o algoritmo, uma vez que era perdido tempo na procura de pessoas em localidades que certamente estas não estariam.

5.1.3 Funcionamento do Sistema

O experimento foi executado por 1 hora e 15 minutos, incluindo os períodos de inatividade devido ao reinício do experimento quando novas pessoas adentravam o ambiente e o tempo necessário para montar o equipamento. O equipamento é composto de 1 câmera *Pan-Tilt-Zoom* com seu tripé para nivelamento da altura da câmera em relação às pessoas na cena, um computador no qual foi executado o sistema e, por fim, um *hub* para comunicação entre a câmera e computador.

O computador no qual foi executado o sistema possuia um processador Athlon 1600+ com 1,5GB de memória RAM e espaço em disco suficiente para armazenar até 4 horas de fotografias. A taxa de processamento das fotografias capturadas era de 2 quadros por segundo, podendo chegar a 10 quadros por segundo quando apenas efetuava-se captura e salvamento em disco (sem processamento). A movimentação da câmera podia demorar até 3 segundos (considerando-se a movimentação entre os dois extremos horizontais).

O *software* foi desenvolvido utilizando a linguagem C++. Esta linguagem foi escolhida devido a eficiência do processamento, necessário para uma maior taxa de quadros por segundo. Em um momento futuro, após a otimização do código e refinamento do sistema, o programa deve ser disponibilizado de forma que outros desenvolvedores possam aprimorar o sistema.

As pessoas na cena estavam cientes do experimento, mas não posaram deliberadamente

para a câmera em nenhum momento, portanto não foi esperado que as pessoas estivessem em posições ou feições semelhantes a pessoas que posaram para uma fotografia. Neste experimento, foram obtidas imagens de um evento, no qual as pessoas estão interagindo com algo que não é a câmera e o papel do sistema foi fotografar esta interação. A característica do evento é diferente de uma seção fotográfica e, portanto, deve ser avaliado desta forma.

As pessoas não foram instruídas em nenhum momento a posicionaram-se de maneira favorável ou a contribuir de qualquer outra forma com o experimento. A contribuição, portanto, foi apenas permitir que o experimento fosse realizado naquele instante. Uma vez que nem todos os voluntários concordaram em ceder sua imagem para utilização em pesquisas, esta base de imagens não será disponibilizada publicamente.

Para este experimento, foi utilizada a estratégia de busca panorâmica, descrita na Seção 4.3.4.

5.2 Resultados Obtidos

5.2.1 Localização das Pessoas

Após 40 minutos efetivos e contínuos de experimento, puderam ser obtidas 2070 fotografias. Das 14 pessoas, em média 8 eram encontradas com duas interações do algoritmo. Contando todas as execuções do programa, 12 pessoas diferentes foram localizadas, contadas pelo número de faces diferentes detectadas nas imagens resultantes. Vale lembrar que, durante as execuções do programa, houve variação na quantidade de pessoas presentes na sala-de-aula. Também se salienta que existiam pessoas na cena de difícil localização devido à oclusão por alunos posicionados mais à frente no ambiente, o que já foi discutido e ilustrado na Seção 5.1.1.

Em se somando os mapas de detecção das três últimas execuções do algoritmo, foi obtida a imagem mostrada na Figura 5.2. Nesta imagem, pode-se notar como as pessoas, ilustradas pelos retângulos pretos, estavam distribuídas no ambiente no momento do experimento. As faces mais à extremidade representam o professor e o operador do sistema, também detectados pelo módulo localizador de pessoas. Entretanto, não é possível precisar a quantidade de pessoas encontradas. Isto porque as pessoas poderiam ser contadas mais de uma vez, uma

vez que a imagem armazena a informação ao longo do tempo. Isto, entretanto, não consiste em um problema uma vez que mesmo que uma quantidade maior de faces fosse indicada, ao efetuar o enquadramento da região, o detector de faces atualizaria a imagem de forma que esta pudesse representar o número correto de faces desta região.

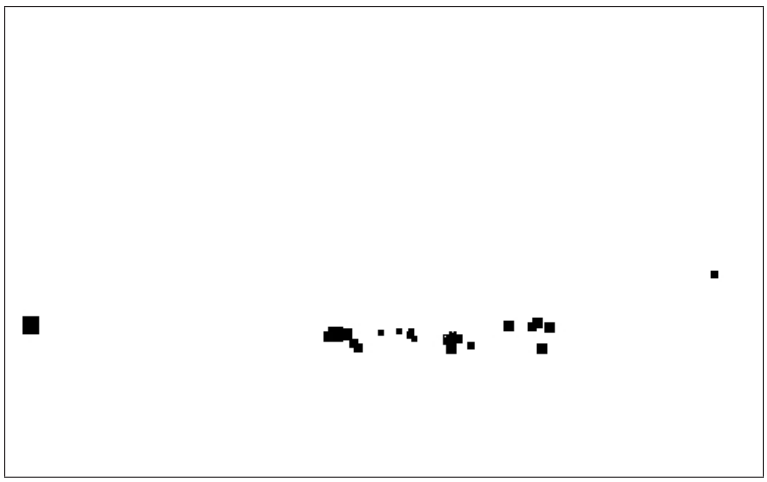


Figura 5.2: Vista horizontal da cena, no qual estão representadas, através de retângulos pretos, as faces detectadas.

5.2.2 Votação da Qualidade das Fotografias Obtidas

As imagens obtidas foram submetidas a uma votação para analisar a qualidade visual das mesmas. Uma audiência de leigos voluntariou-se para opinar sobre as melhores imagens obtidas pela câmera. Aos votantes eram apresentadas 10 fotografias adquiridas em seqüência de um mesmo tema fotografico. A escolha por 10 fotografias de um mesmo tema em seqüência serem mostradas juntas, teve a intenção de gerar uma comparação direta por fotos semelhantes. No grupo de fotografias, algumas poderiam ter pequenas modificações decorrentes da impossibilidade da câmera de informar o fim do movimento requisitado, resultando em algumas fotografias obtidas prematuramente. Este problema teve o aspecto positivo de avaliar o trabalho da composição.

As fotografias foram votadas por 11 pessoas diferentes, de forma circular, na qual cada um dos conjuntos de 10 fotografias era votado e, ao fim dos conjuntos, voltava-se ao primeiro. Dessa forma, não era preciso que os voluntários votassem em todas as fotografias e o

conjunto era bem distribuído entre os votantes. Em média, as fotografias receberam 3 votos cada.

O resultado da votação mostrou que, das 2070 fotografias obtidas inicialmente, 382 fotografias diferentes foram escolhidas como sendo boas, o que equivale a um percentual de 18% das fotografias obtidas. Em se considerando também as imagens escolhidas mais de uma vez, obtiveram-se 466 fotografias num percentual de 22%. Este percentual, contudo precisa levar em consideração os vários fatores que compuseram o cenário. Em primeiro lugar, esse resultado não quer dizer que os outros 82% das imagens não eram boas, apenas que não foram escolhidas como as melhores imagens. Em segundo lugar, devido à filtragem por imagens que contivessem faces, apenas 1550 fotografias foram apresentadas para votação, o que resultou num percentual aproximado de 25% (30% se consideradas as imagens repetidas).

O objetivo desta votação foi o de verificar três premissas: (1) a obtenção de várias fotos em seqüência, ao invés de apenas uma, é benéfico para que uma boa fotografia não seja perdida por questões de centésimos de segundos; (2) o sistema consegue obter fotografias que representem bem um evento em um ambiente parcialmente controlado. Por fim, também se desejou verificar se as regras de composição fotográfica aplicadas instantes antes da primeira fotografia, foram de fato obedecidas. Esta avaliação pode ser feita sem a análise de pessoas, sendo realizada apenas por algoritmos que irão mensurar a conformidade a uma determinada regra.

A primeira avaliação foi feita da seguinte forma: a partir da listagem de imagens escolhidas, pode-se afirmar que a primeira fotografia não foi escolhida em nenhuma vez neste experimento. Portanto, todas as 382 fotografias votadas são uma das 9 adicionais obtidas. Novamente, isso não quer dizer que as primeiras fotografias eram descartáveis, apenas que as 9 subseqüentes tinham mais qualidade e eram mais representativas. Uma explicação para este fato é que a primeira foto pode ter sido obtida enquanto a câmera ainda era ajustada de acordo com seus sensores, e.g., foco automático ou iluminação, o que pode ter prejudicado a qualidade destas fotografias.

Mais uma vez, a câmera não sinaliza ao sistema quando está pronta para obter fotografias, portanto o erro de foco em virtude da movimentação da mesma não pode ser evitado. Também foi contabilizado que das 382 fotografias, as 5 últimas foram preferidas em 64% das

ocasiões, contra 36%. Isto se explica pelo fato das 5 últimas terem sido obtidas com a composição final, enquanto as 5 primeiras ainda em fase de ajustes. Logo, pode-se considerar que a obtenção de fotografias extra é desejável.

Já a segunda avaliação foi realizada com respeito à quantidade de fotografias. Como foi mostrado, das 2070 fotografias, 1550 possuíam faces, dessas quais 382 foram escolhidas, o que dá 25% do total. Apesar do número aparentar ser baixo, é preciso considerar que o experimento foi executado por 1 hora (a duração da aula descontada de aproximadamente 30 minutos de instalação de equipamento e reparos no sistema).

Também é preciso levar em consideração que caso fosse desejado publicar as fotos de um determinado evento, apenas uma pequena parte desse total seria, de fato, publicada. O que dá uma grande margem de escolha para uma seleção mais rigorosa. Portanto, levando estes fatores em consideração, pode-se afirmar que as fotos obtidas pela câmera podem ser utilizadas para que seja feita uma representação do evento. A Figura 5.3 mostra duas imagens consideradas boas pela audiência votante. Outras imagens tanto consideradas boas como ruins podem ser vistas no Apêndice C.

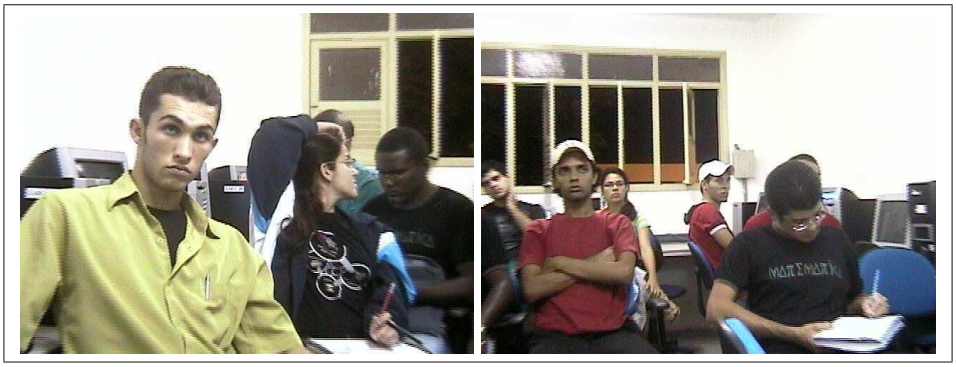


Figura 5.3: Imagens consideradas boas.

Por fim, a terceira avaliação visou a determinar se as regras de composição foram corretamente aplicadas. A Tabela 5.1 mostra os valores médios obtidos pela execução dos algoritmos de extração de características descritos na Seção 3.5 nas 1550 imagens que continham faces. O Apêndice D apresenta os valores obtidos para uma amostra de imagens.

A partir destes resultados (nos quais maiores valores representam maior conformidade à regra), pode-se avaliar o quão bem o algoritmo de composição executou sua tarefa. As regras de composição utilizadas foram a regra-dos-terços, a regra-do-zoom, e a regra-da-

Tabela 5.1: Valores médios e desvios padrão obtidos por cada regra de composição.

Integridade. Horizontal	**Integridade Vertical**	**Espaçamento sup.**
0,82 , DP: 0,30	0,50 , DP: 0,28	0,65 , DP: 0,22
Terços Vertical	**Terços Horizontal**	**Terços pontos**
0,69 , DP: 0,2	0,69, DP: 0,13	0,74 , DP: 0,14

integridade.

Pode-se perceber que a Integridade Horizontal foi respeitada, como também a regra-do-zoom, analisado a partir das integridades e do cálculo do espaçamento superior. Os valores que destoaram, no caso a Integridade Vertical e a regra-dos-terços (quando analisadas as retas horizontais isoladamente), são complementares e podem ser explicados pela presença de muitas fotografias de grupos.

As fotografias de grupo atendiam a média da altura das faces o que, em decorrência da diferença de estatura das pessoas e da distância até a câmera, levou o sistema a optar por deixar espaço em excesso e, conseqüentemente, distanciar-se da reta dos terços horizontal superior.

No caso da reta dos terços vertical, este problema foi minimizado, pois as duas retas verticais (as que dividem a imagem horizontalmente em 3 partes) eram contabilizadas. Pode-se considerar, portanto, que o sistema obedece com uma precisão aceitável às regras de composição fotográfica apresentadas.

5.3 Considerações Finais

Neste capítulo, foram descritos os experimentos realizados em um ambiente parcialmente controlado, com o objetivo de avaliar o comportamento do sistema desenvolvido nesta dissertação, o qual consiste na integração de módulos de composição automática e de busca e fotografia de pessoas a partir de uma câmera *Pan-Tilt-Zoom*.

Em virtude dos resultados obtidos, é possível afirmar que o sistema realizou com êxito as tarefas propostas. Em primeiro lugar, este foi capaz de localizar as pessoas distribuídas por um ambiente através da estratégia de busca panorâmica, descrita no Capítulo 4. Em segundo lugar, de fazer um mapeamento da cena de forma a que o sistema armazenasse

conhecimento acerca de uma dada situação possuindo, assim, uma informação de partida para futuras execuções do sistema.

O sistema também se mostrou capaz de enquadrar o alvo com qualidade, utilizando as regras de composição fotográfica como guias de precisão de posicionamento, com uma pequena quantidade de movimentos, o que beneficia a agilidade do processo como um todo. Por fim, as fotografias foram avaliadas por uma audiência a qual considerou que em 25% dos casos, a fotografia representava de forma satisfatória o evento fotografado.

Uma comparação com outros trabalhos é difícil de ser feita apenas analisando-se os valores numéricos. Isso se dá pois os experimentos envolvem fotografias de terceiros. Não é possível reproduzir o comportamento das pessoas, mas apenas a localização. Também não é possível isentar os votantes e, como já foi dito, muitas vezes a publicação da base de fotografias não é possível por questões dos direitos de imagem dos voluntários, o que permitiria uma votação de bases de dados distintas por um mesmo conjunto de votantes.

Mesmo com esta dificuldade, o sistema foi comparado a outros trabalhos em outros aspectos. Em se comparando com o trabalho de Byers et al. [Byers et al., 2003], podem-se observar alguns ganhos. No trabalho citado, 2.000 fotografias são avaliadas entre muito boas, boas, neutras, ruins e muito ruins por um conjunto de votantes voluntários. O resultado mostra que 9% das fotos foram consideradas muito boas enquanto 20% foram consideradas boas. Em somando os dois valores, obtém-se um resultado melhor que os 25% aqui apresentados.

Entretanto, há de se considerar que na situação proposta por Byers et al. as pessoas posavam intencionalmente para a câmera e esperavam por seus ajustes. Esta diferença faz com que, por exemplo, a taxa de aceitação dos votantes do experimento acima seja maior. Isso acontece pois como os votantes posaram para a foto, a possibilidade da foto ser mais agradável para o votante é maior. Uma pequena diferença pode, portanto, modificar o resultado. Logo, não é possível analisar-se apenas numericamente o resultado, há de se levar em consideração o ambiente em que os problemas se inserem.

Capítulo 6

Conclusão

Este capítulo apresenta um sumário dos principais pontos discutidos nesta dissertação, bem como as contribuições da pesquisa desenvolvida e propostas de trabalhos futuros acerca dos problemas encontrados durante o desenrolar da pesquisa, os quais necessitam de estudos mais aprofundados.

6.1 Sumário da Dissertação

No Capítulo 2 foi discutido o estado-da-arte do problema da composição fotográfica. Pôde ser observada a lacuna ainda existente de trabalhos na área de Composição Automática de fotografias, especialmente em softwares visando uma plataforma de fotografia autônoma, objetivo deste trabalho.

No Capítulo 3, foi realizado um estudo sobre possíveis maneiras de se efetuar correções em imagens já capturadas, de forma que fosse possível aumentar sua qualidade no tocante ao aspecto de composição fotográfica. Partindo de uma revisão bibliográfica sobre composição automática de fotografias, foram descritos algoritmos para efetuar as correções nas imagens. Os algoritmos propostos e desenvolvidos nesta dissertação são controlados por um módulo que escolhe dinamicamente qual regra de composição deve ser utilizada. Este capítulo também descreve a extração de características com o fim de se obter informações relevantes sobre uma fotografia assim como as regras comumente utilizadas por fotógrafos humanos. No Capítulo 3, também há uma discussão sobre a influência da utilização de uma dada regra no resultado final segundo a opinião do fotógrafo e de uma audiência imparcial.

O Capítulo 4 complementa o Capítulo 3, na medida em que propõe um sistema que tem por objetivo a aplicação de regras de composição antes da fotografia ser obtida, numa tarefa preventiva, visando uma correção on-line. Esse sistema foi demonstrado através do uso de uma câmera *Pan-Tilt-Zoom*. Como requisito essencial a esta demonstração, houve a necessidade de localizar o alvo em uma cena da qual não se possuia qualquer informação *a priori*. Desta forma, foi descrito um algoritmo para movimentar a câmera em busca de pessoas e, a partir da experiência aprendida com o passar do tempo, aumentar o número de procuras em posições de maior probabilidade de se encontrarem as mesmas pessoas, ao mesmo tempo que ainda permite a busca por pessoas ainda não localizadas.

No Capítulo 5, foram apresentados os experimentos realizados na integração das técnicas discutidas nos Capítulos 3 e 4. O objetivo destes experimentos foi dar ainda mais suporte ao funcionameno das regras aqui descritas. O experimento foi executado em uma sala de aula contendo 14 alunos, em que 2070 fotografias foram obtidas resultando em uma aprovação de 25% por um grupo de observadores (votantes).

6.2 Contribuições

Nesta dissertação, são propostas abordagens que objetivam o aumento da qualidade de uma fotografia. Outros trabalhos abordaram o mesmo problema, porém este trabalho teve o objetivo de ser mais completo, tendo de cada abordagem o que de melhor ela apresenta para a construção de um sistema que, com qualidade, reúna conhecimento de todas as áreas. Logo, foram abordados nesta dissertação metodologias inéditas para o problema de composição fotográfica e automação de fotografia.

Em primeiro lugar, é apresentado um sistema autônomo para efetuar, mesmo após a fotografia ter sido obtida, ajustes considerados benéficos à luz da composição fotográfica. Esta abordagem utiliza a detecção de faces como ponto de partida e calcula o espaço ocupado pelas pessoas através de medidas antropométricas, heurísticas utilizadas pela Anatomia e Artes Plásticas, para respeitar as proporções normais dos seres humanos.

Em seguida, é apresentado um sistema que utiliza uma câmera *Pan-Tilt-Zoom* (câmera dotada de motores para rotação e ajuste do ângulo de visão) para localizar, enquadrar e fotografar alvos em uma cena. São mostradas estratégias que podem ser utilizadas para

agilizar ou, pelo menos organizar, a busca por alvos em uma cena. São também apresentadas propostas para que, após localizados os alvos, sejam enquadrados, mesmo sem ter informação do ângulo de visão da câmera. Também é um diferencial deste trabalho o fato de apenas uma câmera ser utilizada, sem sensores externos ou câmeras adicionais. Também não é preciso nenhuma informação ou condição prévia para o funcionamento do sistema. Este é um problema pouco tratado na literatura e cujas soluções normalmente utilizam suporte de sensores ou câmeras adicionais, por exemplo ver os trabalhos de Clady et al., Funahashi et al. e Senior e Hampapur [Clady et al., 2001; Funahasahi et al., 2004; Senior and Hampapur, 2005]. Nesta dissertação, mostra-se como a correção pode ser realizada antes mesmo da fotografia ser obitida através de um método *on-line* de correção.

Por fim, um outro grande diferencial é a junção destas duas etapas(localização e composição), pouco exploradas na literatura. Logo, em um único sistema, pode-se localizar, enquadrar de acordo com a composição, fotografar, analisar a qualidade da fotografia, em sendo necessário, corrigir a composição da fotografia já obtida e poder classificar o resultado final de acordo com a sua qualidade.

Parte da abordagem aqui apresentada, mais especificamente a parte da dissertação que trata da composição automática de fotografias, foi publicado na conferência VIIP - *Visualization, Imaging and Image Processing* [Cavalcanti et al., 2006].

6.3 Trabalhos Futuros

Esta seção apresenta algumas sugestões de trabalhos futuros com respeito à obtenção de um melhor desempenho e a novas aplicações do sistema proposto.

Em primeiro lugar, deve-se procurar novas abordagems para os problemas mais relevantes encontrados nesta dissertação. Um dos principais problemas que existe é a incapacidade do sistema detectar o alvo em alguns casos. Isto gera uma cascata de erros, já que a informação primordial - posição e dimensões da face - estando errada, todo o processo que se utiliza dessa informação, irá errar. Portanto um primeiro trabalho futuro é a investigação da possibilidade de utilização de outros detectores, seja de pessoas, seja de pele ou de movimento. Uma proposta é a utilização de pontos de Atenção Visual aliado à detecção de faces.

Um outro projeto de trabalho futuro é a ampliação do grau de liberdade da câmera aco-

plando outros módulos de movimento. Desta forma, podem-se adquirir fotografias de qualidade ainda maior. Em compensação, há um grande desafio que é a compreensão da cena como um todo, de forma que a câmera possa mover-se de forma segura, porém objetiva.

Após localizado o tema, existe também uma proposta de melhoria para os passos seguintes. Uma delas é a capacidade de interpretar elementos da cena que possam servir como sinais. Estes sinais - gestos, placas, sons, cores, etc - podem ser utilizados como uma forma de permitir interatividade entre o sistema e o alvo. Assim, pode-se conseguir um controle bem maior do sistema, sem a necessidade de se fazer complexos algoritmos que entendam a cena como um todo.

Na área da composição fotográfica, existe uma demanda por mais regras de composição que possam ser integradas, cuja aplicação seria indiscutivelmente benéfica à fotografia. Novos módulos de composição além daqueles apresentados no Apêndice A podem ser um ponto de partida nesta direção. Além das regras de composição, propriamente ditas, faz-se desejável o estudo e compreensão de outras etapas do processo da fotografia, de maneira que todas as etapas estejam confluindo para a obtenção de uma excelente fotografia..

Por fim, é importante um refinamento na etapa de extração de características de forma que o treinamento de um Sistema Inteligente para aprender informações sobre a qualidade de fotografias seja possível. Este refinamento consiste em um melhor entendimento do processo como um todo, especialmente no tocante à obtenção da informação a partir do votante. Por este se comportar de uma forma nem sempre ideal, é preciso que seja direcionado o foco do problema de forma a não deixar margens ao erro.

Além destes, outros trabalhos futuros poderiam derivar do que foi apresentado nesta dissertação. Estas possibilidades foram destacadas por apresentarem ser as mais promissoras.

Referências Bibliográficas

[Arnheim, 1980] Arnheim, R. (1980). *Arte & Percepção Visual: Uma psicologia da visão criativa*. Pioneira Thomson Learning, São Paulo.

[Banerjee and Evans, 2004] Banerjee, S. and Evans, B. L. (2004). Unsupervised automation of photographic composition rules in digital still cameras. *Proc. IS&T SPIE Conf. on Sensors, Color, Cameras, and Systems for Digital Photography*, 5301:364–373.

[Batista et al., 2006] Batista, L. B., Gomes, H. M., and Carvalho, J. M. (2006). Photogenic facial expression discrimination. Em *International Conference on Computer Vision Theory and Applications (VISAPP'06)*, páginas 166–171.

[Berg et al., 2005] Berg, A. C., Berg, T. L., and Malik, J. (2005). Shape matching and object recognition using low distortion correspondences. Em *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 01, páginas 26–33.

[Boyd, 2000] Boyd, J. (2000). Enhacing photographic composition − wellington photoghraphic society. URL: http://asp.photo.free.fr/enhancing_composition_wellington.htm. Último acesso: [03/06/07].

[Busselle, 1999] Busselle, M. (1999). *Better Picture Guide to Photographing People*. Rotovision, New York.

[Byers et al., 2003] Byers, Z., Dixon, M., Goodier, K., Grimm, C., and Smart, W. (2003). An autonomous robot photographer. *Proc. of IROS - International Conf. on Robots and Systems*, 3:2636–2641.

[Byers et al., 2004] Byers, Z., Dixon, M., Smart, W. D., and Grimm, C. M. (2004). Say cheese!: Experiences with a robot photographer. *AI Magazine*, 25(3):37–46.

[Cavalcanti et al., 2006] Cavalcanti, C. S. V. C., Gomes, H., Meireles, R., and Guerra, W. (2006). Towards automating photographic composition of people. Em *Proceedings of IASTED-VIIP 2006 - Visualization, Imaging and Image Processing*. IASTED.

[Cavalcanti and Gomes, 2005] Cavalcanti, C. S. V. C. and Gomes, H. M. (2005). People detection in still images based on a skin filter and body part evidence. Em *CDRom Proceedings of XVIII Brazilian Symposium on Comp. Graphics and image Proc.*, Natal. 2 pages.

[CImg, 2005] CImg (2005). The CImg Library - C++ Template Image Processing Library. URL: http://cimg.sourceforge.net/. Último acesso: [03/06/07].

[Clady et al., 2001] Clady, X., Collange, F., Jurie, F., and Martinet, P. (2001). Object tracking with a pan-tilt-zoom camera: application to car driving assistance. Em *IEEE International Conference on Robotics and Automation, 2001*, volume II, páginas 1653–1658, LASMEA, CNRS, Aubiere, France.

[Clark, 2001] Clark, R. N. (2001). Image detail (how much detail can you capture and scan?). URL: http://www.clarkvision.com/imagedetail/scandetail.html. Último acesso [03/06/2007].

[Dace, 2006] Dace, M. (2006). Human proportion. URL: http://www.dace.co.uk/proportion.htm. Último acesso: [03/06/07].

[Datta et al., 2006] Datta, R., Joshi, D., Li, J., and Wang, J. Z. (2006). Studying aesthetics in photographic images using a computational approach. *Lecture Notes in Computer Science*, 3953(3):288–301.

[Desolneux et al., 2004] Desolneux, A., Moisan, L., and Morel, J.-M. (2004). *Gestalt Theory and Computer Vision*. Kluwer Academic Publishers.

[Fossa and Erickson, 2005] Fossa, J. A. and Erickson, G. W. (2005). The divided line and the golden mean. *Revista Brasileira de História da Matemática*, 5(9):59–77.

[Freeman, 2004] Freeman, M. (2004). *Photographing People*. The Ilex Press Limited.

[Funahasahi et al., 2004] Funahasahi, T., Tominaga, M., Fujiwara, T., and Koshimizu, H. (2004). Hierarchical face tracking by using ptz camera. Em *Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, páginas 427–432.

[Gomes Filho, 2000] Gomes Filho, J. (2000). *Gestalt do Objeto - Sistema de Leitura Visual da Forma.* Ed. Escrituras, São Paulo.

[Grill and Scanlon, 1990] Grill, T. and Scanlon, T. (1990). *Photographic Composition*. New York: Amphoto Books. Watson-Guptill Publications, Broadway.

[Hager and Belhumeur, 1998] Hager, G. and Belhumeur, P. (1998). Efficient region tracking with parametric models of geometry and illumination. *IEEE Transactions on Pattern Analysis And Machine Intelligence*, 10(20):1025–1039.

[Haykin, 1999] Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*. Prentice Hall, Upper Saddle River, NJ.

[Hedgecoe, 2003] Hedgecoe, J. (2003). *The new manual of photography*. D.K. - Dorling Kindersley Limited, London.

[Hedgecoe, 2005] Hedgecoe, J. (2005). *The Book of Photography*. D.K. - Dorling Kindersley Limited, Great Britain.

[Hendee, 1997] Hendee, W. R. (1997). *The Perception of Visual Information*. Springer-Verlag, New York.

[Hu et al., 2000] Hu, C., Yu, Q., Li, Y., and Ma, S. (2000). Extraction of parametric human model for posture recognition using genetic algorithm. *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 1:518–523.

[Hurter, 2004] Hurter, B. (2004). *The Portrait Photographer's Guide to Posing*. Amherst Media, New York.

[IJG, 2005] IJG (2005). Independent JPEG group. URL: http://www.ijg.org/. Último acesso: [03/06/07].

[Itti et al., 2005] Itti, L., Rees, G., and Tsotsos, J. K. (2005). *Neurobiology of Attention*. Elsevier Inc.

[Kant, 1993] Kant, I. (1993). *Crítica da Faculdade de Juízo - Trad. Valerio Rohden e António Marques.* Forense Universitária, Rio de Janeiro.

[Koffka, 1955] Koffka, K. (1955). *Principles of Gestalt Psychology*. Routledge and Kegan Paul Ltd., London.

[Latzel et al., 2005] Latzel, M., Darcourt, E., and Tsotsos, J. (2005). People tracking using robust motion detection and estimation. Em *Proceedings of the 2nd Canadian Conference on Computer and Robot Vision*, páginas 270–275.

[Lawrence, 2004] Lawrence, G. (2004). Composition. URL: http://www.geofflawrence.com. Último acesso: [03/06/07].

[Li et al., 1999] Li, J., Wang, J. Z., Gray, R. M., and Wiederhold, G. (1999). Multiresolution object-of-interest detection for images with low depth of field. Em *Proceedings of the 10th International Conference on Image Analysis and Processing*, volume 1, páginas 32–37.

[Linhares, 2004] Linhares, J. (2004). *Fotografando Mulheres*. Brasport, Rio de Janeiro, Brasil.

[Luo et al., 2004] Luo, J., Singhal, A., Etz, S. P., and Gray, R. T. (2004). A computational approach to determination of main subject region in photographic images. *Image and Vision Computing*, 01(22):227–241.

[Lyman and Varian, 2003] Lyman, P. and Varian, H. R. (2003). How much information 2003. URL: http://www.sims.berkeley.edu/how-much-info-2003. Último acesso: [03/06/07].

[Mitchell, 1999] Mitchell, T. (1999). Machine learning and data mining. *Communications of the ACM*, 42(11):30–36.

[Nikos, 2005] Nikos (2005). Group snapshots - travel photography blog. URL: http://www.travelphotoblog.com/archives/000735.shtml. Último acesso: [03/06/07].

[Okazaki, 1998] Okazaki, R. Y. (1998). Creating a storyboard for video production. URL: http://www2.hawaii.edu/ ricky/etec/basicshot.html. Último acesso: [03/06/07].

[OpenCV, 2005] OpenCV (2005). OpenCV - Open Source Computer Vision Library. URL: http://www.intel.com/technology/computing/opencv/index.htm. Último acesso: [03/06/07].

[Ozer and Wolf, 2002] Ozer, I. B. and Wolf, W. (2002). Real-time posture and activity recognition. Em *IEEE Proceedings of the Workshop on Motion and Video Computing (MOTION'02)*.

[Parker, 1997] Parker, J. R. (1997). *Algorithms for Image Processing and Computer Vision*. John Wiley & Sons.

[Pereira and Gomes, 2006] Pereira, E. T. and Gomes, H. M. (2006). Guiding a bottom-up visual attention mechanism to locate specific image regions using a distributed genetic optimization. Em *CIARP - Iberoamerican Congress on Pattern Recognition*, páginas 257–266.

[Pereira et al., 2006] Pereira, E. T., Gomes, H. M., and Florentino, V. F. C. (2006). Bottom-up visual attention guided by genetic algorithm optimization. Em *Eigth IASTED International Conference on Signal and Image Processing*, páginas 228–233.

[photography.com, 2006] photography.com (2006). Film vs. digital statistics. URL: http://www.photography.com/topics/film-vs-digital-statistics/. Último acesso: [03/06/07].

[Photoinfo.com, 2000] Photoinfo.com (2000). Composition basics − how to get good pictures. URL: http://photoinf.com/General/ITRC_ UMT. Último acesso: [03/06/07].

[Qiu et al., 2006] Qiu, X., Song, J., Zhang, X., and Liu, S. (2006). A complete coverage path planning method for mobile robot in uncertain environments. Em *Proceedings of the 6th World Congress on Intelligent Control and Automation*, páginas 8892–8896, Dalian, China.

[Ramalho and Palacin, 2004] Ramalho, J. and Palacin, V. (2004). *Escola de Fotografia*. Futura, São Paulo.

[Rowley et al., 1998a] Rowley, H. A., Baluja, S., and Kanade, T. (1998a). Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1):23–28.

[Rowley et al., 1998b] Rowley, H. A., Baluja, S., and Kanade, T. (1998b). Rotation invariant neural network-based face detection. *Computer Vision and Pattern Recognition*, 20(1):38–44.

[Santella et al., 2006] Santella, A., Agrawala, M., DeCarlo, D., Salesin, D., and Cohen, M. (2006). Gaze based interaction for semi automatic photo cropping. Em *Proceedings of the SIGCHI conference on Human Factors in computing systems*, páginas 771–780.

[Senior and Hampapur, 2005] Senior, A. and Hampapur, A. (2005). Acquiring multi-scale images by pan-tilt-zoom control and automatic multi-camera calibration. Em *Workshop on Applications of Computer Vision Jan. 2005.*, páginas 433–438.

[Sinha and Pollefeys, 2004] Sinha, S. N. and Pollefeys, M. (2004). Towards calibrating a pan-tilt-zoom camera network. Em *OMNIVIS 2004 - Workshop on Omnidirection Vision and Camera Networks held in conjunction with ECCV 2004*, páginas 1–13.

[Sprague and Luo, 2002] Sprague, N. and Luo, J. (2002). Clothed people detection in still images. Em *Proceedings of the 16 th International Conference on Pattern Recognition (ICPR'02)*, volume 3, páginas 585–589.

[Stack and Smith, 2003] Stack, J. and Smith, C. (2003). Combining random and data-driven coverage planning for underwater mine detection. Em *Oceans 2003 Proceedings*, volume 5, páginas 2463 – 2468.

[Takahashi and Sugakawa, 2004] Takahashi, K. and Sugakawa, S. (2004). Remarks on human posture classification using self-organizing map. Em *IEEE International Conference on Systems, Man and Cybernetics*, volume 03, páginas 2623– 2628.

[Tomono, 2003] Tomono, M. (2003). Path planning with target finding by single camera under map uncertainty. Em *Computational Intelligence in Robotics and Automation, 2003. Proceedings. 2003 IEEE International Symposium on*, volume 1, páginas 465 – 470.

[Tremblay, 2003] Tremblay, P. (2003). Basic rules of photography. URL: http://trem.ca/learning.html. Último acesso: [03/06/07].

[Vapnik, 1999] Vapnik, V. N. (1999). *The Nature of Statistical Learning Theory*. Springer, 2nd edition.

[Viola and Jones, 2001] Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. *Computer Vision and Pattern Recognition*, 1:511–518.

[Yamada et al., 1998] Yamada, M., Ebihara, K., and Ohya, J. (1998). A new robust real-time method for extracting human silhouettes. Em *Proceedings of the 3rd. International Conference on Face & Gesture Recognition*, páginas 528–533.

[Yang and Luo, 2004] Yang, S. X. and Luo, C. (2004). A neural network approach to complete coverage path planning. *IEEE Transactions on Systems, Man, and Cybernetics*, 34(1):718–725.

[Zhang et al., 2005] Zhang, M., Zhang, L., Sun, Y., Feng, L., and Ma, W. (2005). Auto cropping for digital photographs. Em *IEEE International Conference on Multimedia and Expo*, páginas 4–8.

Apêndice A

Noções de Composição Fotográfica

Existem casos em que a fotografia após revelada não agrada aos olhos do fotógrafo e paira a dúvida sobre o que faltou para que o sentimento que o fotógrafo viu a olho nu pudesse ser passado por completo. Isto se dá pois na conversão de uma cena para uma fotografia perde-se muita informação. Em primeiro lugar, o ângulo de visão de um ser humano é maior que o da câmera. É perdida também informação com relação às dimensões de representação dos objetos, que caem de 3 dimensões na cena real para 2 dimensões na fotografia. Estes, além de diversos outros fatores tais como o movimento, as cores, o cheiro, etc. são perdidos quando da transposição da imagem para o papel fotográfico.

Professores e estudiosos de arte costumam ser rigorosos quanto ao uso de palavras na atividade de descrição de artes. A informação visual é praticamente impossível de ser descrita verbalmente. São meras descrições limitadas de todos os sentidos que atuam simultaneamente, pois as palavras não são a via ideal para o contato com a realidade, servem apenas para nomear aquilo que se vê, ouve e sente [Arnheim, 1980]. De forma análoga, a fotografia, como ela é, estática, não é capaz de passar muito do ambiente que foi fotografado. É preciso mais do artista (do fotógrafo) para que o ambiente, o sentimento, etc. possam estar presentes em um pedaço de papel que meramente representa uma pequena parte do que os olhos captavam em um dado momento.

Como são incompletas por si só, já que possuem uma dimensão a menos e tem ângulo de visão reduzido, as fotografias precisam objetivar a informação que trazem. As regras de composição trazem alguma informação sobre o que de uma cena deve ser enfatizada para que o tema da fotografia seja evidenciado. Elas tentam fazer com que a perda da informação

seja minimizada e a fotografia, mesmo com tantos fatores a menos que a cena, ainda possa passar uma parte do sentimento do fotógrafo ao visualizador.

Um termo muito utilizado na fotografia como um todo e que vai ser muito citado nas explicações que se seguem é o termo "Tema" ou "Alvo". A escolha do tema pode ser vista como uma "pré-regra" para qualquer abordagem de composição utilizada.

O tema está para uma foto como o sujeito (gramaticalmente falando) está para uma frase. O tema é o que se deseja, de fato, fotografar [Ramalho and Palacin, 2004; Grill and Scanlon, 1990]. O tema pode ser uma pessoa, um lugar, uma característica de uma pessoa, um animal, uma ação, etc. Enfim, independentemente do observador, é fundamental que fique claro o tema - o verdadeiro centro de interesse de uma foto.

Todas as fotografias tem um tema. Mas nem sempre o tema da fotografia é de fato o tema que o fotógrafo quis enfatizar. Essa diferença é o que pode, visualmente, deixar uma foto amadora desagradável. O erro, nesses casos, é que o fotógrafo não conseguiu deixar claro "O que" ele estava querendo fotografar [Hedgecoe, 2003]. E este erro pode e deve ser evitado sempre [Hurter, 2004].

As regras de composição de fotografia precisam ter claro o tema. Para que uma regra faça sentido, é necessário que o "tema" sofra a composição. Ele é o ponto principal de uma fotografia e as regras de composição servem para, prioritariamente, dar mais ênfase ao tema, deixá-lo cada vez mais saliente.

Neste trabalho, o tema tratado será sempre pessoas. Toda e qualquer regra aqui descrita visará a boas práticas de composição de forma que pessoas sejam enfatizadas como o tema de uma fotografia. Apesar de terem o mesmo objetivo, as regras de composição específicas para fotos com pessoas diferem em muito das demais (paisagens, animais, etc). Uma regra aplicável para uma paisagem pode ser aplicada para uma pessoa, porém, a forma como será feita essa aplicação deve ser diferente. As regras descritas, contudo, além de poderem ser aplicadas em pessoas, podem também ser aplicadas caso o tema da foto seja outro. Para um sistema computacional, por exemplo, o que muda é o detector de alvos, que, no caso desta disseratção, está configurado como um detector de pessoas, podendo ser um detector de flores, animais, ou até mesmo um ponto de atenção visual. Esta escolha, contudo, é justificável dado que a grande maioria das fotografias no mundo tem pessoas como tema principal [Freeman, 2004].

Portanto, como base para todas as técnicas propostas, tem-se a detecção de pessoas como um forte requerimento. Dado que todas as pessoas têm face, a solução para nosso problema pode, a priori, basear-se neste requisito. Sabendo-se a posição da face pode-se, por exemplo, inferir a posição do resto do corpo da pessoa. Um detector de faces é, portanto, de fundamental importância para esta finalidade.

A seguir, serão descritas as regras de composição desenvolvidas nesta dissertação para a melhoria automática de uma fotografia.

A.1 A Regra-dos-Terços

Esta regra - uma das mais antigas e conhecidas - sugere que o tema da fotografia esteja em pontos de referência da fotografia. Estes pontos são os pontos de encontro de duas retas verticais e horizontais eqüidistantes entre si (e nas retas em si). A regra é conhecida por Regra-dos-Terços [Joseph e Hamptom, 2003] por estas retas (horizontais ou verticais) distanciarem-se das bordas da fotografia em 1/3 do tamanho total disponível. A imagem é dividida, portanto, em 3 regiões de mesma área tanto horizontalmente como verticalmente. A Figura A.1 ilustra esta regra.

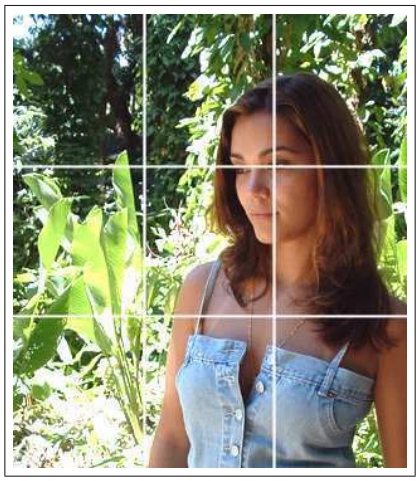


Figura A.1: Ilustração do uso da Regra-dos-Terços. Exemplificando o posicionamento ideal do tema − o olho da modelo.

A Regra-dos-Terços teve origem há milênios e era utilizada (antes do advento da fotografia) em pinturas dos mais famosos nomes. Outra regra de posicionamento utiliza a “medida

áurea − Golden Mean" [Fossa and Erickson, 2005]. Tem seus pontos de referência muito próximos aos pontos da Regra-dos-Terços o que, de certa forma, ajuda a validá-la.

Os fotógrafos amadores, inicialmente, sofrem um impacto ao conhecer esta regra. Em sua grande maioria, os amadores buscam centralizar o tema. A justificativa desta descentralização do tema é que na medida que o tema é centralizado, o cérebro é forçado a fixar sua atenção no tema ignorando o resto do contexto da fotografia. Isto pode ser desejável, por exemplo, quando da fotografia de uma única pessoa, quando se deseja máxima atenção àquele único ponto. Temas centralizados, contudo, produzem um visual muito estático [Hurter, 2004]. Ainda assim é possível uma fotografia que obedeça à regra e na qual a pessoa continue como centro da atenção.

Em casos onde a imagem já está completamente preenchida, escolhe-se o ponto principal para colocá-lo em um ponto de terço, por exemplo, no caso de um close-up no quaj se deve colocar um dos olhos em um dos pontos de terço [Hedgecoe, 2005; Hurter, 2004; Freeman, 2004; Ramalho and Palacin, 2004].

Enquanto isso, uma fotografia que obedece à regra-dos-terços, faz com que o observador, em um primeiro momento, olhe para o centro de atenção − a pessoa − e, em seguida, busque outros elementos na fotografia. Isto dá à fotografia dinamicidade tendo como resposta, um observador menos "cansado" pois seus olhos são levados a "viajar" pela superfície da fotografia em busca de novas informações.

A.2 Regra-do-Arranjo e Regra-dos-Ímpares

As regras-do-arranjo da cena e regra-dos-ímpares tratam da disposição dos temas pelo espaço da foto, mesmo que com sentidos diferentes.

Na regra-do-arranjo [Boyd, 2000], deseja-se que a disposição dos temas seja feita de forma uniforme porém dinâmica. Essa disposição, contudo, é feita a partir de duas nuances: na primeira define-se o posicionamento das pessoas de forma que elas ocupem o espaço da fotografia, a segunda, entretanto, preocupa-se de fato com a distribuição dos temas pela fotografia.

O arranjo das pessoas quando de um grupo de pessoas com uma certa heterogenia na altura deve ser realizado de forma a evitar que as pessoas posicionem-se de forma a causar

uma confusão visual, ou seja, os olhos não sabem para onde olhar. A forma estática do posicionamento, também é um importante fator a ser evitado: fotos nas quais todos estão alinhados tal como um "time de futebol posando para uma foto" normalmente não causam uma sensação visual agradável, ainda que o caminho que os olhos percorrerão seja bem definido. Existem regras para que a disposição das pessoas seja a mais agradável e atraente o possível aos olhos do observador. Na Figura A.2, um exemplo de uma regra de arranjo é mostrado. Neste exemplo, as pessoas são distribuídas com o objetivo de formar triângulos. Algumas figuras geométricas (tais como os triângulos, as pirâmides, diamantes, linhas curvas e os triângulos invertidos dentre outras formas) causam maior interesse ao olho humano que simplesmente linhas retas [Nikos, 2005].

Figura A.2: Regra da disposição triangular aplicada a pessoas em uma fotografia.

Vale ressaltar, que a disposição das pessoas para uma fotografia também passa pelo bom senso de não obrigar mudanças que deixem as pessoas fotografadas em posição desconfortável em detrimento de se obedecer a uma regra.

As fotos com pessoas alinhadas, contudo, não são automaticamente descartáveis. Assim como mostrado na Figura A.3, uma disposição utilizando apenas linhas também pode ser agradável se puder formar um circuito levando o olho a "passear" pela imagem até um determinado ponto. Diagonais, por exemplo, possuem um dinamismo muito útil a uma fotografia.

A Regra-dos-Ímpares sugere que as fotos são mais agradáveis quando o número de alvos é impar - alvos de número par geram visuais estáticos e sem atratividade. Isto não quer dizer que fotos com duas pessoas são desagradáveis, apenas se as pessoas estejam tão separadas

Figura A.3: (1) A disposição das pessoas leva a um passeio pela foto retornando ao meio. (2) O alinhamento cria uma diagonal no sentido superior direito - direcionando o fluxo.

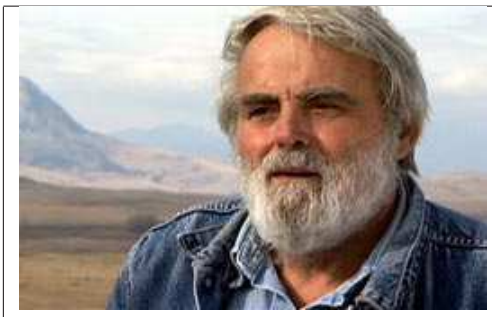
a ponto de se formarem dois grupos. Ao juntarem-se duas pessoas, forma-se um par o que obedece a regra. Ao juntarem-se mais de duas pessoas, forma-se 1 grupo - obedecendo também à regra. Esta regra é justificada pela atração que a desordem provocada pelos ímpares provoca ao cérebro.

A.3 Regra-do-*Zoom* / *Headroom*

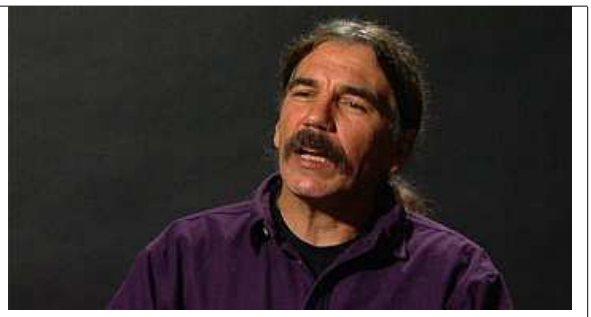
Uma foto de pessoas precisa que os temas estejam visíveis a ponto de ser possível reconhecê-los. Ao mesmo tempo, uma foto que praticamente toda a extensão é ocupada apenas pela face do alvo também pode não ser a solução ideal, em virtude do interesse pela paisagem.

Em suma, o tema deve estar em uma posição agradável ao contexto geral da imagem. Nem tão longe nem tão perto. Se assim for procedido, a quantidade de fotografias cuja qualidade é boa aumenta. Apesar disso, não é simples saber o quão perto/longe se deseja que os temas estejam.

Quando a fotografia é de pessoas (especialmente um close-up de uma única pessoa) o zoom pode ser controlado na medida do *Headroom* [Photoinfo.com, 2000] - o espaço que a cabeça do alvo deve ter para que tenha condições de se "deslocar" pela fotografia. Fotos nas quais as cabeças estão muito próximas às bordas dão a sensação que o fotografado irá "bater a cabeça" na borda ou, se a cabeça estiver muito baixa, dá a impressão que a pessoa foi "degolada". A Figura A.4 ilustra um *headroom* correto.



(a)



(b)

Figura A.4: (a) *Headroom* incorreto estando a cabeça "batendo nos limites" já em (b) a cabeça está posicionada corretamente.

O *Headroom* também não pode ser excessivo a ponto de se dar a impressão que o *background* é mais importante que o tema [Busselle, 1999].

A.4 Regra-da-Integridade

Uma regra importante para uma foto ter uma boa aceitação é que os temas da foto não sejam "cortados" pelos limites da foto. Óbviamente, fotos de corpo inteiro nem sempre são necessárias podendo haver cortes sem prejuízo da qualidade [Hurter, 2004; Busselle, 1999]. Entretanto, é fundamental que, em especial, os membros superiores não sejam ocultados. Mãos, braços e cabeça - ou seja, partes do corpo próximas às juntas - produzem um efeito desagradável ao serem excluídos da fotografia. Portanto uma regra interessante é a de sempre verificar se os temas estão por completo no espaço da fotografia.

A regra-do-zoom/*headroom* (item anteriormente abordado) ilustra um pouco desta questão mais direcionado ao posicionamento da cabeça. Neste item fala-se do tema em si, independentemente de qual parte do corpo está sendo omitida - o desejável é que não haja nenhuma parte do corpo do alvo ocultada. Evidentemente, a foto não precisa ser de corpo inteiro, basta que, a partir de um dado ponto, preferencialmente pouco acima da linha da cintura, não sejam cortadas outras partes do corpo que, porventura, estejam separadas do corpo (tronco). Isto não quer dizer, contudo, que a foto só pode ser tirada desta forma. O ideal é que o alvo maximize a área ocupada na fotografia. Isto pode fazer, por exemplo, que seja preciso cortar algo do corpo [Lawrence, 2004]. O que importa é que esta parte do corpo cortada não dê a impressão que esta parte do corpo foi amputada.

A Figura A.5 ilustra o quão desagradável pode parecer uma fotografia a qual os membros e as cabeças das pessoas foram cortados.

Figura A.5: Nesta foto, a cabeça e o corpo de várias pessoas do tema foram recortadas por descuido do fotógrafo.

Observe ainda que esta mesma figura poderia tornar-se bem mais agradável se algum algoritmo de atuação prévia impedisse o fotógrafo de obter esta fotografia, alertando-o sobre as conseqüências negativas deste incorreto enquadramento.

Apêndice B

Imagens Representando a Composição Fotográfica

Este Apêndice mostra imagens processadas pelo algoritmo de composição automática de fotografias e votadas por voluntários que decidiram pela qualidade do processamento. Os resultados da votação são mostrados a seguir. Na primeira coluna, duas imagens sendo a da esquerda a original e a da direita a modificada. Na segunda coluna, a quantidade de votos para cada opção (na ordem apresentada anteriormente) e, na terceira coluna, a conclusão da votação onde 'Iguais Adequada' indica que não havia necessidade de modificação na imagem original e 'Iguais Inadequada' que o votante achava que deveria ter sido feita alguma mudança.

Percebe-se que não existe unanimidade mesmo para imagens cujas mudanças foram consideradas benéficas por uma larga maioria, indicando subjetividade na avaliação. Também percebe-se dificuldade na avaliação da necessidade de mudanças, uma vez que as opções 2 e 3 eram votadas incoerentemente. As Figuras B.1 e B.2 mostram imagens consideradas boas. Já as Figuras B.3 e B.4 mostram imagens consideradas ruins pelos votantes após as modificações efetuadas pelo algoritmo.

 	4 - 0 - 0 - 1 Melhorou
 	5 - 0 - 0 - 0 Melhorou
 	5 - 0 - 0 - 0 Melhorou
 	5 - 0 - 0 - 0 Melhorou
 	4 - 0 - 0 - 1 Melhorou
 	5 - 0 - 0 - 0 Melhorou
 	4 - 0 - 0 - 1 Melhorou
 	5 - 0 - 0 - 0 Melhorou

Figura B.1: Imagens consideradas melhores

 	5 - 0 - 0 - 0 Melhorou
 	3 - 0 - 2 - 0 Melhorou
 	3 - 1 - 1 - 0 Melhorou
 	3 - 0 - 0 - 2 Melhorou
 	5 - 0 - 0 - 0 Melhorou
 	0 - 3 - 2 - 0 Iguais Adequado
 	0 - 3 - 2 - 0 Iguais Adequado
 	0 - 4 - 1 - 0 Iguais Adequado
 	0 - 4 - 1 - 0 Iguais Adequado
 	0 - 3 - 2 - 0 Iguais Adequado

Figura B.2: Imagens consideradas melhores ou iguais

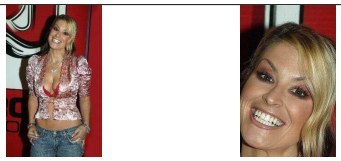
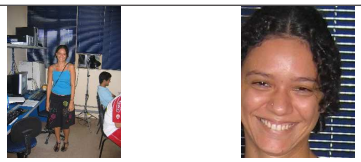
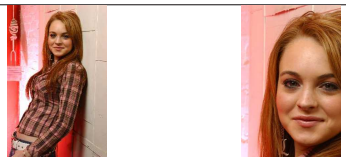
	2 - 0 - 0 - 3 Piorou
	2 - 0 - 0 - 3 Piorou
	2 - 0 - 0 - 3 Piorou
	2 - 0 - 0 - 3 Piorou
	2 - 0 - 0 - 3 Piorou
	1 - 0 - 0 - 4 Piorou
	0 - 2 - 3 - 0 Iguais Inadequado
	0 - 0 - 5 - 0 Iguais Inadequado
	0 - 2 - 3 - 0 Iguais Inadequado
	1 - 1 - 3 - 0 Iguais Inadequado

Figura B.3: Imagens consideradas piores ou ruins por ausência de mudanças

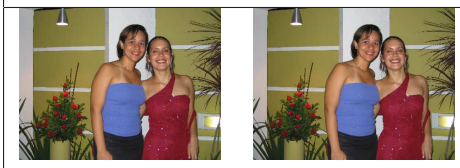
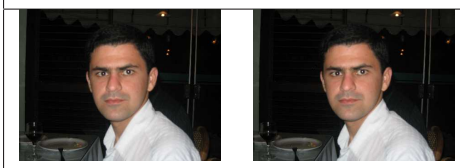
	0 - 2 - 3 - 0 Iguais Inadequado
	0 - 1 - 4 - 0 Iguais Inadequado
	0 - 2 - 3 - 0 Iguais Inadequado
	0 - 2 - 3 - 0 Iguais Inadequado
	1 - 1 - 3 - 0 Iguais Inadequado
	0 - 1 - 3 - 1 Iguais Inadequado
	0 - 2 - 3 - 0 Iguais Inadequado
	0 - 2 - 3 - 0 Iguais Inadequado
	0 - 2 - 3 - 0 Iguais Inadequado
	0 - 2 - 3 - 0 Iguais Inadequado

Figura B.4: Imagens consideradas ruins por ausência de mudanças

Apêndice C

Imagens Resultantes do Experimento com a Câmera *Pan-Tilt-Zoom*

Este apêndice traz algumas imagens capturadas pela câmera PTZ no experimento descrito no Capítulo 5. Na Figura C.1 são apresentadas imagens obtidas durante os passos de movimentação para ajuste da câmera, e as Figuras (C.2)-(C.5) apresentam algumas das imagens que foram consideradas boas pelos votantes no mesmo experimento. As Figuras (C.6)-(C.9) apresentam imagens que não foram escolhidas como sendo boas pelos votantes.

Pode-se perceber, ao analisar os resultados, que existem mais fotografias de duas pessoas em particular. Isso se dá devido a geometria do local, que impediu que outras pessoas fossem visualizadas uma vez que estas duas estavam, na grande maioria do tempo do experimento, na frente das outras.

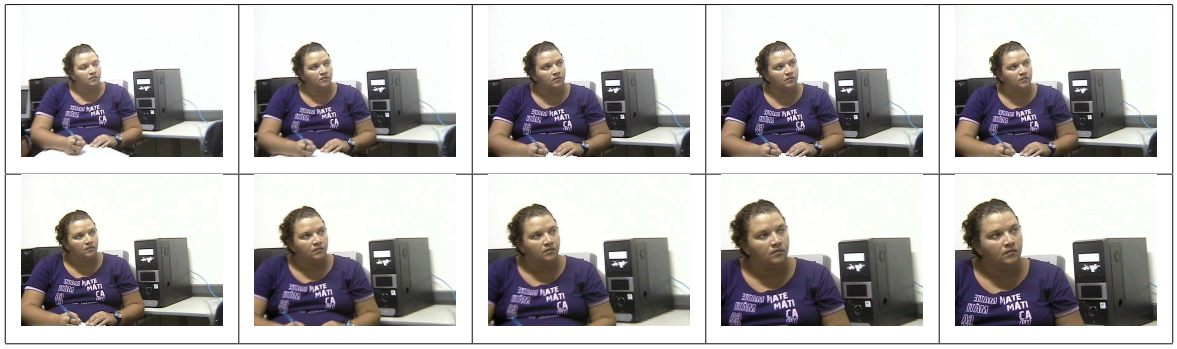


Figura C.1: Exemplo do ajuste da câmera. Ajusta-se primeiro o posicionamento e em seguida o zoom.

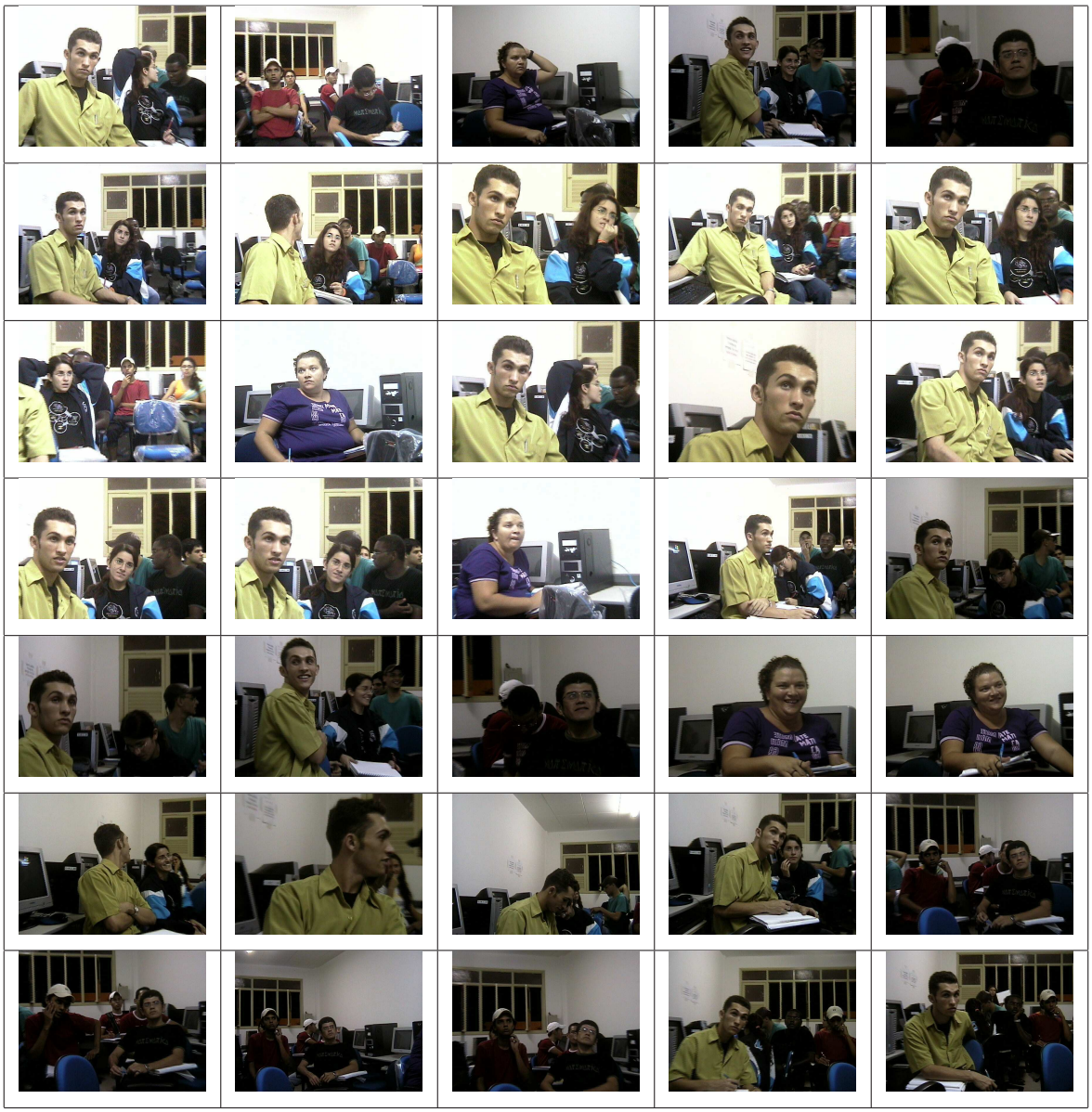


Figura C.2: Alguns exemplos de fotografias consideradas boas pelos votantes 1/4.

Figura C.3: Alguns exemplos de fotografias consideradas boas pelos votantes 2/4.

Figura C.4: Alguns exemplos de fotografias consideradas boas pelos votantes 3/4.

Figura C.5: Alguns exemplos de fotografias consideradas boas pelos votantes 4/4.

Figura C.6: Alguns exemplos de fotografias não escolhidas como boas pelos votantes 1/4.

Figura C.7: Alguns exemplos de fotografias não escolhidas como boas pelos votantes 2/4.

Figura C.8: Alguns exemplos de fotografias não escolhidas como boas pelos votantes 3/4.

Figura C.9: Alguns exemplos de fotografias não escolhidas como boas pelos votantes 4/4.

Apêndice D

Imagens Representando a Extração de Características

Este apêndice apresenta o experimento realizado para extração de características. As imagens cujas características foram extraídas aparecem lado a lado com os valores das características obtidas nas Figuras (D.1)-(D.7). Os valores estão normalizados entre 0 e 1 sendo 0 a nota dada para uma imagem mal avaliada enquanto 1 indica uma correta composição no quesito. Os rótulos foram abreviados para facilitar a visualização: Int. Horizontal e Int. Vertical são, respectivamente, integridade horizontal e integridade vertical, Esp. Superior indica o espaçamento superior, Terços Horizontal, Vertical e Pontos, simplifica o termo regra-dos-terços para as retas horizontais, verticais e para o ponto dos terços, respectivamente.

Percebe-se que para as imagens cujas faces foram detectadas corretamente, os valores calculados são coerentes e podem dar um bom indicativo da qualidade da fotografia, uma vez que seja definida como métrica de qualidade a conformidade às regras de Composição Fotográfica. Entretanto, pode haver um impacto negativo nas imagens em que nem todas as faces tenham sido detectadas, uma vez que o tamanho e posicionamento das faces consistem no principal dado utilizado para o cáculo destes valores.

	Int. Horizontal 1,00	Int. Vertical 0,54	Esp. Superior 0,50
	Terços Vertical 0,94	Terços Horizontal 0,73	Terços Pontos 0,55
	Int. Horizontal 1,00	Int. Vertical 0,78	Esp. Superior 0,61
	Terços Vertical 0,92	Terços Horizontal 0,84	Terços Pontos 0,76
	Int. Horizontal 1,00	Int. Vertical 0,13	Esp. Superior 0,73
	Terços Vertical 0,96	Terços Horizontal 0,94	Terços Pontos 0,92
	Int. Horizontal 1,00	Int. Vertical 0,82	Esp. Superior 0,59
	Terços Vertical 0,94	Terços Horizontal 0,84	Terços Pontos 0,74
	Int. Horizontal 1,00	Int. Vertical 0,97	Esp. Superior 0,53
	Terços Vertical 0,93	Terços Horizontal 0,75	Terços Pontos 0,59
	Int. Horizontal 1,00	Int. Vertical 0,63	Esp. Superior 0,57
	Terços Vertical 0,95	Terços Horizontal 0,80	Terços Pontos 0,68
	Int. Horizontal 1,00	Int. Vertical 0,58	Esp. Superior 0,56
	Terços Vertical 0,99	Terços Horizontal 0,79	Terços Pontos 0,66
	Int. Horizontal 1,00	Int. Vertical 0,20	Esp. Superior 1,00
	Terços Vertical 0,95	Terços Horizontal 0,84	Terços Pontos 0,74

Figura D.1: Amostra da extração de características 1/7.

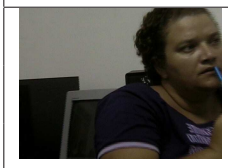
	Int. Horizontal 1,00	Int. Vertical 0,78	Esp. Superior 0,45
	Terços Vertical 0,96	Terços Horizontal 0,78	Terços Pontos 0,64
	Int. Horizontal 1,00	Int. Vertical 0,78	Esp. Superior 0,61
	Terços Vertical 0,91	Terços Horizontal 0,86	Terços Pontos 0,79
	Int. Horizontal 0,82	Int. Vertical 0,56	Esp. Superior 1,00
	Terços Vertical 0,99	Terços Horizontal 0,88	Terços Pontos 0,79
	Int. Horizontal 1,00	Int. Vertical 0,00	Esp. Superior 1,00
	Terços Vertical 0,95	Terços Horizontal 0,81	Terços Pontos 0,69
	Int. Horizontal 0,73	Int. Vertical 0,91	Esp. Superior 1,00
	Terços Vertical 0,94	Terços Horizontal 0,71	Terços Pontos 0,53
	Int. Horizontal 1,00	Int. Vertical 0,63	Esp. Superior 1,00
	Terços Vertical 0,93	Terços Horizontal 0,81	Terços Pontos 0,69
	Int. Horizontal 1,00	Int. Vertical 0,57	Esp. Superior 0,65
	Terços Vertical 0,91	Terços Horizontal 0,91	Terços Pontos 0,90
	Int. Horizontal 1,00	Int. Vertical 0,38	Esp. Superior 0,64
	Terços Vertical 0,95	Terços Horizontal 0,91	Terços Pontos 0,86

Figura D.2: Amostra da extração de características 2/7.

	Int. Horizontal 1,00	Int. Vertical 0,58	Esp. Superior 0,63
	Terços Vertical 0,97	Terços Horizontal 0,85	Terços Pontos 0,75
	Int. Horizontal 1,00	Int. Vertical 0,76	Esp. Superior 0,49
	Terços Vertical 0,95	Terços Horizontal 0,71	Terços Pontos 0,53
	Int. Horizontal 1,00	Int. Vertical 0,99	Esp. Superior 0,53
	Terços Vertical 1,00	Terços Horizontal 0,74	Terços Pontos 0,57
	Int. Horizontal 1,00	Int. Vertical 1,00	Esp. Superior 0,52
	Terços Vertical 0,96	Terços Horizontal 0,74	Terços Pontos 0,56
	Int. Horizontal 1,00	Int. Vertical 0,68	Esp. Superior 0,47
	Terços Vertical 1,00	Terços Horizontal 0,74	Terços Pontos 0,57
	Int. Horizontal 1,00	Int. Vertical 0,95	Esp. Superior 0,47
	Terços Vertical 0,92	Terços Horizontal 0,73	Terços Pontos 0,57
	Int. Horizontal 1,00	Int. Vertical 0,80	Esp. Superior 0,49
	Terços Vertical 0,96	Terços Horizontal 0,72	Terços Pontos 0,54
	Int. Horizontal 1,00	Int. Vertical 0,74	Esp. Superior 0,47
	Terços Vertical 0,94	Terços Horizontal 0,77	Terços Pontos 0,64

Figura D.3: Amostra da extração de características 3/7.

	Int. Horizontal 1,00	Int. Vertical 0,71	Esp. Superior 0,60
	Terços Vertical 0,97	Terços Horizontal 0,86	Terços Pontos 0,78
	Int. Horizontal 1,00	Int. Vertical 0,77	Esp. Superior 0,38
	Terços Vertical 0,96	Terços Horizontal 0,90	Terços Pontos 0,84
	Int. Horizontal 1,00	Int. Vertical 0,71	Esp. Superior 0,60
	Terços Vertical 0,96	Terços Horizontal 0,86	Terços Pontos 0,78
	Int. Horizontal 1,00	Int. Vertical 0,82	Esp. Superior 0,59
	Terços Vertical 0,96	Terços Horizontal 0,84	Terços Pontos 0,74
	Int. Horizontal 1,00	Int. Vertical 0,82	Esp. Superior 0,56
	Terços Vertical 0,99	Terços Horizontal 0,80	Terços Pontos 0,66
	Int. Horizontal 1,00	Int. Vertical 0,87	Esp. Superior 0,59
	Terços Vertical 0,93	Terços Horizontal 0,82	Terços Pontos 0,72
	Int. Horizontal 1,00	Int. Vertical 0,78	Esp. Superior 0,61
	Terços Vertical 0,95	Terços Horizontal 0,87	Terços Pontos 0,79
	Int. Horizontal 1,00	Int. Vertical 0,17	Esp. Superior 1,00
	Terços Vertical 0,94	Terços Horizontal 0,80	Terços Pontos 0,68

Figura D.4: Amostra da extração de características 4/7.

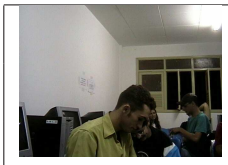
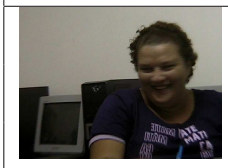
	Int. Horizontal 1,00	Int. Vertical 0,96	Esp. Superior 0,27
	Terços Vertical 0,99	Terços Horizontal 0,90	Terços Pontos 0,83
	Int. Horizontal 1,00	Int. Vertical 1,00	Esp. Superior 0,26
	Terços Vertical 0,97	Terços Horizontal 0,89	Terços Pontos 0,83
	Int. Horizontal 1,00	Int. Vertical 0,85	Esp. Superior 1,00
	Terços Vertical 0,94	Terços Horizontal 0,89	Terços Pontos 0,84
	Int. Horizontal 0,97	Int. Vertical 0,43	Esp. Superior 1,00
	Terços Vertical 0,97	Terços Horizontal 0,76	Terços Pontos 0,61
	Int. Horizontal 0,64	Int. Vertical 0,66	Esp. Superior 1,00
	Terços Vertical 0,99	Terços Horizontal 0,86	Terços Pontos 0,78
	Int. Horizontal 0,74	Int. Vertical 0,66	Esp. Superior 1,00
	Terços Vertical 0,95	Terços Horizontal 0,82	Terços Pontos 0,70
	Int. Horizontal 0,92	Int. Vertical 0,18	Esp. Superior 1,00
	Terços Vertical 0,94	Terços Horizontal 0,87	Terços Pontos 0,80
	Int. Horizontal 1,00	Int. Vertical 0,36	Esp. Superior 1,00
	Terços Vertical 0,92	Terços Horizontal 0,89	Terços Pontos 0,84

Figura D.5: Amostra da extração de características 5/7.

	Int. Horizontal 1,00	Int. Vertical 0,57	Esp. Superior 0,65
	Terços Vertical 0,96	Terços Horizontal 0,93	Terços Pontos 0,90
	Int. Horizontal 1,00	Int. Vertical 0,57	Esp. Superior 0,65
	Terços Vertical 0,99	Terços Horizontal 0,94	Terços Pontos 0,90
	Int. Horizontal 1,00	Int. Vertical 0,37	Esp. Superior 0,62
	Terços Vertical 0,92	Terços Horizontal 0,88	Terços Pontos 0,83
	Int. Horizontal 0,97	Int. Vertical 0,07	Esp. Superior 1,00
	Terços Vertical 0,97	Terços Horizontal 0,98	Terços Pontos 0,99
	Int. Horizontal 1,00	Int. Vertical 0,83	Esp. Superior 0,66
	Terços Vertical 0,97	Terços Horizontal 0,95	Terços Pontos 0,93
	Int. Horizontal 1,00	Int. Vertical 0,93	Esp. Superior 0,51
	Terços Vertical 0,99	Terços Horizontal 0,71	Terços Pontos 0,53
	Int. Horizontal 1,00	Int. Vertical 0,76	Esp. Superior 1,00
	Terços Vertical 0,93	Terços Horizontal 0,89	Terços Pontos 0,88

Figura D.6: Amostra da extração de características 6/7.

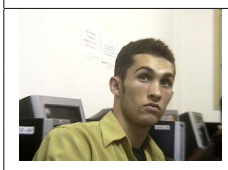
	Int. Horizontal 0,92	Int. Vertical 0,48	Esp. Superior 1,00
	Terços Vertical 0,94	Terços Horizontal 0,74	Terços Pontos 0,57
	Int. Horizontal 1,00	Int. Vertical 0,88	Esp. Superior 0,63
	Terços Vertical 0,97	Terços Horizontal 0,89	Terços Pontos 0,82
	Int. Horizontal 0,85	Int. Vertical 0,85	Esp. Superior 1,00
	Terços Vertical 0,92	Terços Horizontal 0,84	Terços Pontos 0,77
	Int. Horizontal 0,95	Int. Vertical 0,38	Esp. Superior 1,00
	Terços Vertical 0,96	Terços Horizontal 0,78	Terços Pontos 0,64
	Int. Horizontal 1,00	Int. Vertical 0,85	Esp. Superior 1,00
	Terços Vertical 0,97	Terços Horizontal 0,90	Terços Pontos 0,84
	Int. Horizontal 0,79	Int. Vertical 0,57	Esp. Superior 1,00
	Terços Vertical 0,98	Terços Horizontal 0,80	Terços Pontos 0,67
	Int. Horizontal 1,00	Int. Vertical 0,74	Esp. Superior 0,48
	Terços Vertical 0,93	Terços Horizontal 0,72	Terços Pontos 0,54

Figura D.7: Amostra da extração de características 7/7.