

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM
CIÊNCIA DA COMPUTAÇÃO

*“Métodos para um Sistema de Animação Facial Baseado em
Aquisição Estereoscópica”*

Rodrigo Assaf

São Carlos
2006

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária da UFSCar**

A844ms

Assaf, Rodrigo.

Métodos para um sistema de animação facial baseado em aquisição estereoscópica / Rodrigo Assaf. -- São Carlos : UFSCar, 2006.

104 p.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2006.

1. Computação gráfica. 2. Processamento de imagens. 3. Animação por computador. 4. Aquisição 3D. I. Título.

CDD: 006.6 (20^a)

Universidade Federal de São Carlos

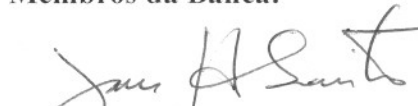
Centro de Ciências Exatas e de Tecnologia
Programa de Pós-Graduação em Ciência da Computação

“Métodos para um Sistema de Animação Facial Baseado em Aquisição Estereoscópica”

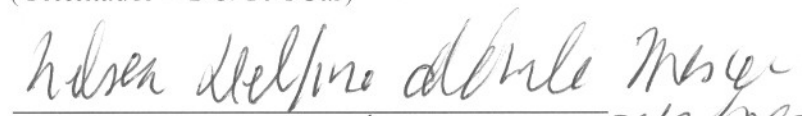
RODRIGO ASSAF

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Membros da Banca:



Prof. Dr. José Hiroki Saito
(Orientador – DC/UFSCar)



Prof. Dr. Nelson Delfino d'Ávila Mascarenhas
(DC/UFSCar)



Prof. Dr. Hani Camille Yehia
(Depto. Engenharia Eletrônica/UFMG)

São Carlos
Maio/2006

“Animação não é a ilusão da vida; ela é a vida.”

(Chuck Jones | 1912 – 2002)

Agradecimentos

Agradeço primeiramente a Deus por suas graças cedidas e por ter me iluminado todos estes anos de minha vida.

Agradeço, com todo meu amor, aos meus pais. Ao meu pai, Alexandre Assaf Neto, que muito me ajudou, apoiou, aconselhou e acompanhou toda minha formação e vida. À minha mãe, Anita Assaf, que acreditou e perseverou ao meu lado sempre com muito amor.

À minha namorada Marcela, pela compreensão, carinho e amor até nos momentos mais difíceis, e pelas suas sugestões no meu trabalho.

Ao meu irmão, Alexandre, pelo apoio, força e companheirismo durante minha trajetória.

Ao meu orientador, Prof. Dr. José Hiroki Saito, pela paciência, sabedoria e comprometimento com a pesquisa durante estes anos de mestrado.

A toda minha família, pelo amor e carinho.

Aos meus amigos de mestrado, da turma BCC 2000 e de São Carlos pela força e ajuda. Em especial ao Mario à ajuda na montagem do *Hardware*, ao Thadeu pelos estudos em estereoscopia, ao pessoal do GAPIS e ao grupo de estudos do Prof. Dr. José Hiroki Saito.

Agradeço também aos profissionais do Departamento de Computação e da Universidade Federal de São Carlos por toda minha formação e dedicação.

E, finalmente, a todos meus amigos de Ribeirão Preto e aos profissionais da área de computação gráfica que muito me ensinaram.

Sumário

Lista de Figuras	V
Lista de Tabelas	VIII
Resumo	1
Abstract	2
Capítulo 1: Introdução	3
1.1. Identificação do Problema	5
1.2. Objetivos	5
1.2.1. Objetivo Geral	5
1.2.2. Objetivos Específicos	5
1.3. Resumo dos Capítulos.....	6
Capítulo 2: Animação	7
2.1. O Desenvolvimento da Animação	7
2.2. Princípios Básicos da Animação	10
Capítulo 3: Animação Facial	15
3.1. O Desenvolvimento da Área de Animação Facial.....	15
3.2. A Anatomia da Face	17
3.3. Análise das Expressões Faciais	20
3.4. As Expressões Universais	21
3.5. A Linguagem <i>Mimic</i>	21
3.6. O Sistema de Codificação de Ação Facial	22
3.7. Paradigmas da Animação Facial	22
3.7.1. Animação por Controle de Parâmetros	22
3.7.2. Animação por Interpolação de Quadros Chaves	23
3.7.3. Animação Baseada em Performance Real	23
3.7.4. Animação Baseada em Pseudo-Músculo	23

3.7.5. Abstrações de Ações Musculares	23
3.7.6. Deformações Livres de Formas.....	25
3.7.7. Animação Facial Baseada em Músculo	26
3.8. Sincronização Labial (<i>Lip Sync</i>)	27
3.9. Áreas de Aplicação da Animação Facial	28
3.9.1. Indústria de Jogos	28
3.9.2. Medicina	28
3.9.3. Teleconferência	29
3.10. Exemplos de Sistemas de Animação Facial Existentes	29
3.10.1. <i>CrazyTalk (Reallusion)</i>	29
3.10.2. <i>Facial Studio (Digimation)</i>	29
3.10.3. <i>LifeStudio Head</i>	30
3.10.4. <i>Motion Capture</i> Óptico da Universidade de Ohio	30
3.10.5. Modelo Muscular de Parke e Waters	31
3.10.6. Modelo Muscular do “ <i>Max-Planck-Institut Informatik</i> ”	31
3.10.7. <i>Magpie (Third Wish Software and Animation)</i>	32
3.10.8. <i>Game Engine: Jupiter 3D</i>	32
3.10.9. <i>Facade (Stanford Facial Animation System)</i>	32
3.10.10. <i>The Expression Toolkit</i>	33

Capítulo 4: Estereoscopia 35

4.1. Introdução à Estereoscopia	35
4.2. O Olho Humano	35
4.3. Projeção Perspectiva	37
4.4. Princípios Básicos da Estereoscopia	41
4.4.1. Alinhamento de Câmera e Sistema de Coordenadas	41
4.4.2. Sistema de Estereoscopia	42
4.4.3. Calibração da Câmera	44
4.5. Buscando Pontos Correspondentes (<i>Matching Process</i>)	45
4.5.1. Método LSM	45
4.5.1.1 Algoritmo LSM adaptativo	46
4.5.2. <i>Matching</i> por Contornos com operadores morfológicos	51

4.5.3. <i>Matching</i> por Regiões	52
4.6. Outros Métodos de se Obter Informações Tridimensionais	53
4.6.1. Introdução aos Sensores Ativos	53
4.6.2. <i>Scanner</i> a Laser	53
4.6.3. Luz Estruturada	54
4.6.4. Sistemas de <i>Motion Capture</i>	55
4.6.4.1. Sistemas Acústicos	56
4.6.4.2. Sistemas Mecânicos	56
4.6.4.3. Sistemas Ópticos	57
4.6.4.4. Sistemas Magnéticos	58

Capítulo 5: Definição de Métodos para um Sistema de Animação Facial baseado em Estereoscopia

5.1. Sistema de Animação Proposto	59
5.2. Materiais	60
5.3. Desenvolvimento do Sistema	63
5.3.1. <i>Hardware</i>	63
5.3.2. <i>Software</i>	68
5.3.2.1. Performance do Movimento Facial Humano	68
5.3.2.2. Seqüência de Quadros (<i>Frames</i>) Capturados	68
5.3.2.3. Estereoscopia	69
5.3.2.4. Modelo Facial 3D	70
5.3.2.5. Associação (<i>Fitting</i>)	71
5.3.2.6. Interpolação	72
5.3.2.7. Animação Facial	72

Capítulo 6: Resultados e Discussão

6.1. Expressões	73
6.1.1. Surpreso	73
6.1.1.1 Estereoscopia dos pontos correspondentes de cada par de quadros equivalentes (Processo de <i>Matching</i>)	76

6.1.2. Sorriso	78
6.1.2.1. Estereoscopia dos Pontos Correspondentes de cada Par de Quadros Equivalentes (Processo de <i>Matching</i>)	80
6.1.3. Emburrado	81
6.1.3.1. Estereoscopia dos pontos correspondentes de cada par de quadros equivalentes (Processo de <i>Matching</i>)	82
6.2. Sincronismo Labial (<i>Lip Sync</i>)	83
6.2.1. Modelo Aplicado em Face Humana 3D	84
6.2.2. Modelo Aplicado em uma Face <i>Cartoon</i>	86
6.3. Discussão	87
Capítulo 7: Conclusões, Contribuições e Trabalhos Futuros	90
7.1. Contribuições	90
7.2. Limitações	91
7.3. Trabalhos Futuros	93
Glossário	94
Apêndice	96
Referências Bibliográficas	99

Lista de Figuras

Figura 2.1: A lanterna mágica	8
Figura 2.2: Taumatroscópio	8
Figura 2.3: Fenaquistoscópio	9
Figura 2.4: Estroboscópio	9
Figura 2.5: Zootroscópio	9
Figura 2.6: Squash e Stretch.	11
Figura 2.7: As bolas vermelhas demonstram uma antecipação de movimento	12
Figura 2.8: Enquadramento.....	12
Figura 2.9: Slow in e Slow out.	13
Figura 2.10: Exemplo de animação utilizando arco (a) e não utilizando arco (b).	13
Figura 2.11: Exemplo de exagero no rosto da personagem.	14
Figura 3.1: Globeleza Digital	17
Figura 3.2: Cortes e nomenclatura usada.	18
Figura 3.3: Ossos da face.	19
Figura 3.4: Músculos da face.	20
Figura 3.5: (A) tristeza, (B) raiva, (C) satisfação, (D) medo, (E) aversão, (F) surpresa.	21
Figura 3.6: Parâmetros de modelo 3D (Aki Ross) do filme Final Fantasy	23
Figura 3.7: Animação por interpolação de quadros-chave	24
Figura 3.8: Características da Interpolação: (a) Linear e (b) Curva.	24
Figura 3.9: Exemplo de expressões utilizando a técnica de FFD.	26
Figura 3.10: Modelo 3D na camada de músculos: Shrek	27
Figura 3.11: Face animada pelo Software <i>CrazyTalk</i>	29
Figura 3.12: Face modelada e animada pelo <i>Facial Studio</i>	30
Figura 3.13: Face construída pelo <i>software</i> da <i>Lifemode Interactive</i>	30
Figura 3.14: Face com fontes do <i>Motion Capture</i>	31
Figura 3.15: Modelo muscular utilizado por Parke.	32
Figura 3.16: Modelo muscular posicionado na face 3D usado para a animação.....	32
Figura 3.17: Resultado da animação facial pelo <i>Facade</i> : Na ordem da esquerda para a direita, expressão de bravo, aversão, medo, alegria, triste e surpreso	33

Figura 3.18: Resultados de expressões pelo <i>Toolkit</i> : Da esquerda para a direita, expressão neutra, bravo, medo e com a sobrancelha levantada.....	34
Figura 4.1 (a): O olho humano.	36
Figura 4.1 (b): A figura mostra o esquema de transmissão das imagens dos olhos para o cérebro. [58]	37
Figura 4.2: O sistema de coordenadas da câmera (x, y, z) está alinhado com o sistema de coordenadas do mundo (X, Y Z)	38
Figura 4.3: Sistemas desalinhados.....	41
Figura 4.4: Modelo de processamento de estereoscopia	42
Figura 4.5: A partir dos seed points (pontos brancos), é aplicado o voronoi tessellation e a busca por pontos correspondentes (pontos pretos) é iniciada dentro das limitações da geometria.....	45
Figura 4.6: Matching por Contornos com operadores morfológicos.....	52
Figura 4.7: Matching por regiões.....	53
Figura 4.8: Típica geometria de luz estruturada.....	55
Figura 5.1: Métodos para o sistema de animação facial	59
Figura 5.2: Duas câmeras de segurança monocromáticas.	61
Figura 5.3: Placa gráfica para captura de vídeo FX5600.	62
Figura 5.4: Adaptador que realizada a conversão de interfaces DVI-I para RCA.	62
Figura 5.5: Esquema da placa PixelView para captura de vídeo.	63
Figura 5.6: Esquema de câmeras sincronizadas.	64
Figura 5.7: Câmera sem o compartimento de proteção (a) e com seu hardware (b).....	64
Figura 5.8: Tempo dos sinais de FSI e FSO.....	65
Figura 5.9: Relação entre os sinais de FSO e a saída de vídeo.	65
Figura 5.10: Identificação do FSO e FSI e a ligação entre eles.	66
Figura 5.11: Câmeras sincronizadas.	66
Figura 5.12 Disposição das duas câmeras com uma distância de 40mm entre seus centros ópticos.	67
Figura 5.13: Esquema do sistema de animação facial montado.	67
Figura 5.14: Pontos que serão capturados pela estereoscopia na face humana.	68
Figura 5.15: Seleção dos quadros 1 e 6 apenas para o cálculo estereoscópico.....	69

Figura 5.16: Esquema da estereoscopia em alto nível.	70
Figura 5.17: Fotos de referência e o resultado modelado.	71
Figura 5.18: (a) Pontos identificados na face real; (b) Mesmos pontos associados nos vértices da face 3D. Total de 66 pontos de controle.	71
Figura 5.19: Animação Facial concluída baseada no vídeo capturado.	72
Figura 6.1: Imagens faciais de expressão de surpreso pela câmera esquerda	73
Figura 6.2: Imagens faciais de expressão de surpreso pela câmera direita	74
Figura 6.3: Quadros da câmera esquerda que foram selecionados para a interpolação. (a) quadro inicial e (b) quadro final	75
Figura 6.4: Quadros da câmera direita que foram selecionados para a interpolação. (a) quadro inicial e (b) quadro final.....	75
Figura 6.5: Encontro de pontos semelhantes em imagens correspondentes da esquerda e da direita no par de quadros.....	76
Figura 6.6: Associação no modelo e interpolação entre as poses chaves.....	77
Figura 6.7: Imagens faciais de expressão de sorriso obtidas pela câmera esquerda.....	78
Figura 6.8: Imagens faciais de expressão de sorriso obtidas pela câmera direita.....	79
Figura 6.9: Quadros da câmera esquerda que foram selecionados para a interpolação. (a) quadro inicial, (b) quadro intermediário e (c) quadro final.....	79
Figura 6.10: Quadros da câmera direita que foram selecionados para a interpolação. (a) quadro inicial, (b) quadro intermediário e (c) quadro final.....	80
Figura 6.11: Encontro de pontos semelhantes, em imagens correspondentes da esquerda e da direita, nos 3 pares de quadros.....	80
Figura 6.12: Associação no modelo e interpolação entre as poses-chave.....	81
Figura 6.13: Quadros da câmera esquerda que foram selecionados para a interpolação, no caso da expressão facial de emburrado. (a) quadro inicial e (b) quadro final.....	81
Figura 6.14: Quadros da câmera direita que foram selecionados para a interpolação, no caso da expressão facial de emburrado. (a) quadro inicial e (b) quadro final.....	82
Figura 6.15: Encontro de pontos semelhantes em imagens correspondentes da esquerda e da direita no par de quadros, no caso de expressão de emburrado.....	82
Figura 6.16: Associação no modelo e a interpolação entre as poses-chave para o caso de expressão de emburrado.....	83

Figura 6.17: Gráfico da fala usado para referência no sincronismo labial.....	83
Figura 6.18: Quadros da câmera direita que foram selecionados para a interpolação, objetivando o estudo do sincronismo labial.....	84
Figura 6.19: Quadros da câmera esquerda que foram selecionados para a interpolação, objetivando o estudo do sincronismo labial.....	85
Figura 6.20: Encontro de pontos semelhantes em imagens correspondentes da esquerda e da direita no par de quadro.....	85
Figura 6.21: Associação no modelo e a interpolação entre as poses-chave.....	86
Figura 6.22: Associação no modelo e interpolação entre as poses-chave em uma face cartoon.....	87

Lista de Tabelas

Tabela 1: Características de sistemas mecânicos.....	57
Tabela 2: Características de sistemas ópticos.....	58
Tabela 3: Características de sistemas magnéticos.....	59
Tabela 4: Comparação entre o sistema óptico de <i>motion capture</i> da Universidade de Ohio com o sistema proposto.....	89

Resumo

A animação facial tem sido amplamente utilizada em diversas áreas de aplicação e entretenimento, apresentando altos níveis de detalhamento e avanços tecnológicos. Contudo, sua implementação ainda exige um trabalho árduo da equipe de produção podendo desprender muito tempo nesta etapa.

Com o objetivo de adequar um sistema estereoscópico a um sistema de animação facial de baixo custo financeiro, este trabalho trata de assuntos da área de animação e de seus principais fundamentos para serem aplicados em uma face.

Estudos da anatomia da face, como sistema muscular e ósseo, são fundamentais para um melhor desempenho da animação facial, pois é notável que a face é, naturalmente, um sistema complexo. Diversos paradigmas de sistemas faciais são abordados mostrando os atuais estágios desta área da computação gráfica.

Os princípios fundamentais de um sistema estereoscópico, que utiliza duas imagens como referência, são abordados, enfocando seu uso no desenvolvimento de métodos para uma animação facial.

Esta pesquisa baseou-se na demarcação de pontos de controle faciais em uma face humana real, que foi modelada em 3D utilizando a técnica de *box modeling*, para a realização dos métodos de estereoscopia, associação e interpolação obtendo uma expressão facial e de sincronismo labial.

A utilização da estereoscopia mostrou-se, neste trabalho, como sendo uma técnica eficiente para a captação das coordenadas 3D provenientes da face real, resultando em uma animação facial realista. Encontrou-se o re-uso, a portabilidade, a criação de *target morphings* e a criação de *keyframes* como as principais contribuições da estereoscopia, permitindo sua ampla utilização.

Abstract

Facial animation has been widely used in many areas of applications and entertainment, presenting high levels of details and technology. However, its implementation still demands lots of hard work from the producer team.

Aiming at the adaptation of a stereoscopic system to a financial low cost facial animation system, this research covers animation subjects and their main principles, in order to use them on an animated face.

Studies on face anatomy, such as bones and muscles, are considered important to enhance the facial animation results. Several paradigms are covered in order to demonstrate the actual stage of this computer graphic area.

The main aspects of a stereoscopic system, which uses two images as reference, presented in this work, are focused on a method development for a facial animation system.

This research is based on facial control points on a human face, which is 3D modeled using the box modeling technique. The 3D virtual face is used to apply an ending result of a facial expression or lip sync obtained from the methods of stereoscopy, association and interpolation.

The use of the stereoscopy demonstrated to be an efficient technique to capture 3D coordinates from real faces and use them to develop a realistic facial animation. The re-use, portability, target morphing, and key frames are the main contributions of the use of the stereoscopy in a facial animation system.

Capítulo 1: Introdução

A animação facial, bem como outras técnicas de animação em geral, tem sido amplamente usada em diversas áreas de aplicação, tais como: cinema, jogos, desenhos animados, medicina, ambientes virtuais, Internet, entre outras.

Na área de cinema, o desenvolvimento da animação facial tem mostrado um alto nível de detalhamento e avanço tecnológico perante as outras áreas, em razão à alta resolução requerida nos filmes digitais e ao tempo de produção do produto, ou seja, os filmes não são processados em tempo real. Com isso, as animações faciais mais complexas podem passar horas ou talvez meses no estágio de produção final (estágio que consiste em gerar a animação digital final com todos os elementos da animação juntos na cena, incluindo iluminação, reflexos, entre outros efeitos) para depois serem apresentadas ao público.

Em contrapartida, existe uma problemática no processamento em tempo real na arte de animar faces e expressões em jogos eletrônicos, Internet, ambientes virtuais, entre outros meios. Nos jogos eletrônicos tridimensionais, por exemplo, a animação dos personagens ou dos objetos da cena ainda é realizada em baixa quantidade de polígonos (abaixo de 10.000 polígonos, considerando a média de polígonos utilizados em personagens de jogos) deixando, portanto, as animações faciais simplificadas e vista do elevado número de músculos faciais envolvidos e da qualidade gráfica do jogo em tempo real.

Os avanços da capacidade de processamento dos computadores pessoais (PC) e da taxa de definição dos monitores e televisores, bem como o desenvolvimento de *softwares* voltados à computação gráfica, trouxeram inúmeros progressos na área da animação facial. Tais mudanças vêm proporcionando uma melhor qualidade e realismo em suas aplicações. Vale destacar ainda, que esses avanços tem sido amplamente utilizados nas novas gerações de *consoles* de entretenimento digital (MICROSOFT TM, 2006) contribuindo para o avanço das animações faciais em tempo real. Apesar dos progressos alcançados na área de animação facial, atingir um alto grau de realismo, em suas diversas aplicações, ainda é uma tarefa árdua, em função da complexidade, particularidade e diversidade da face humana. Deste modo, o reconhecimento de expressões humanas por outras pessoas, se torna possível mesmo com pequenas e sutis mudanças na expressão facial.

Desta forma, para que sejam promovidos avanços na área de animação facial, torna-se necessário, adquirir um amplo conhecimento técnico sobre o grau de desenvolvimento tecnológico da área e sobre os métodos de aquisição de dados faciais existentes atualmente.

Para tanto, deve-se obter o domínio, não somente da implementação computacional, mas também das técnicas de animação em si.

Tradicionalmente, a animação facial por computador segue dois procedimentos básicos: (1) construção de uma malha 3D da face, e (2) animação da malha 3D em alguma forma de controle na qual simule ações faciais.

No procedimento (1), é desejável a construção de uma malha com o máximo de detalhes, como peles dobráveis, rugas, entre outros, quando for construída em alta quantidade de polígonos (acima de 10.000 polígonos, considerando o nível de detalhe perceptível ao olho humano). Contudo, se a aplicação exigir menos quantidade de polígonos (abaixo de 10.000), o desafio será de fazer uma malha 3D da face convincente. Para isso, outros recursos são utilizados, como textura e iluminação.

A modelagem da malha facial pode ser realizada manualmente através de técnicas de modelagem conhecidas como *Box modeling* e *Patch*, entre outras. Ambas as técnicas possuem características distintas, porém geram os mesmos resultados. A técnica de *Box modeling* consiste em começar a modelagem de uma face, a partir de uma forma primitiva básica como um cubo ou uma esfera chegando em sua forma final através de extrusões de arestas e translações dos vértices das faces poligonais da mesma. A técnica *Patch*, visa usar formas geométricas 2D, usando uma imagem referência, formando assim, a forma 3D final através de translação de vértices.

A modelagem da malha facial também poderá ser realizada, usando uma face real como informações iniciais, automaticamente através de métodos estereoscópicos (usando duas ou mais fotos) ou através de *scanners* à laser e de luz estruturada. A grande vantagem destes métodos automáticos é a facilidade quanto à modelagem. Porém, para se obter uma boa precisão é necessário equipamentos de altos custos e algoritmos robustos para resolver problemas como a oclusão, eventuais ruídos, entre outros.

No procedimento (2), o animador deve decidir quais vértices da malha deverão ser movimentados para que se atinja uma animação facial adequada e convincente. Várias técnicas foram propostas como: modelos parametrizados (PARKE, 1989), modelos com pontos de controle (KURIHARA & ARAI, 1991), modelos musculares com *Kinematic* (WALTERS, 1987), modelos de textura (YAU & DUFFY, 1988), modelos *spline* (NAHAS, 1990), modelos de rastreamento (WILLIAMS, 1990) e modelos musculares dinâmicos (PLATT & BARDLER, 1981).

1.1. Identificação do Problema

Os principais problemas que motivaram o desenvolvimento deste projeto de mestrado foram:

- A artificialidade presente na arte de animar faces, a qual é feita de forma não natural, fazendo com que o animador atente-se ao excesso de músculos faciais e de seus comportamentos sem uma referência real.
- A dificuldade em se alcançar o grau de realismo e veracidade desejado, principalmente na realização de sincronismo labial (*lip sync*), a qual é considerada uma tarefa árdua para o animador.
- A necessidade de diminuir o tempo gasto na etapa de animação de faces em produções e jogos eletrônicos 3D.
- O elevado custo financeiro dos sistemas ópticos de animação facial existentes no mercado.

1.2. Objetivos

Este projeto de mestrado tem como objetivos:

1.2.1 Objetivo Geral

O objetivo geral do projeto consiste em: Animar uma face 3D, previamente obtida por métodos de modelagem, através de uma performance humana real captada por meio de duas câmeras utilizando o procedimento da estereoscopia.

1.2.1 Objetivos Específicos

- Determinar métodos para atingir a animação facial a partir de um modelo real humano usando o método da estereoscopia.
- Atingir um grau de resultado e realismo visivelmente de acordo com a performance humana real.
- Animar expressões e *Lip-Sync* (sincronização labial) utilizando a técnica de interpolação.
- Aplicar a performance humana facial em faces de personagens 3D.
- Desenvolver métodos para um sistema de baixo custo financeiro.

1.2. Resumo dos Capítulos

A presente dissertação de mestrado está disposta nos seguintes capítulos, além deste capítulo 1, de introdução, apresentação dos problemas e dos objetivos:

- **Capítulo 2:** Serão abordados o desenvolvimento da animação e seus princípios básicos.
- **Capítulo 3:** Serão discutidos assuntos condizentes a animação facial, seus paradigmas e suas aplicações.
- **Capítulo 4:** Serão discutidos assuntos como a estereoscopia e outros métodos de obtenção de informações tridimensionais.
- **Capítulo 5:** Abrangerá os aspectos metodológicos da pesquisa.
- **Capítulo 6:** Serão apresentados os resultados alcançados e a discussão destes.
- **Capítulo 7:** Apresentação da conclusão do projeto abordando as limitações, contribuições e trabalhos futuros.

Capítulo 2: Animação

Este capítulo é dedicado ao desenvolvimento da animação desde as primeiras invenções e descobertas até os principais princípios utilizados atualmente. A compreensão do desenvolvimento histórico da animação é importante para o entendimento da evolução tecnológica da mesma.

2.1. O Desenvolvimento da Animação

A palavra “animação” ou animar, deriva-se do latim *animare*, que significa *dar a vida*. Mas só foi a partir do século XX, que este termo passou a ser utilizado na descrição de imagens em movimento, em decorrência do desenvolvimento tecnológico surgido nesse século (SOLOMON, 1987).

A essência da animação, que faz parte das artes visuais, é o movimento, todavia, para realizar suas obras de arte, os desenhistas e os pintores já dispunham dos recursos necessários para sua produção visual na própria natureza, não sendo necessários recursos técnicos e científicos para sua viabilização. Com o surgimento da animação, os artistas depararam-se com um novo problema: a criação da ilusão do movimento. Para tal, era necessário um alto grau de desenvolvimento científico e técnico que permitisse uma rápida sucessão de imagens, fazendo da animação uma arte. Para o alcance deste grau de desenvolvimento, foi necessário uma série de formulações tecnológicas que possibilitaram a prática da animação, tal como é conhecida nos dias atuais.

Uma das primeiras invenções nesse sentido, data de 1645, onde Athanasius Kircher, publicou um texto intitulado *Ars Magna Lucis et Umbrae (A poderosa arte de luz e sombra)*, descrevendo uma lanterna mágica (figura 2.1), que se tratava de uma caixa com uma fonte de luz e um espelho com formato curvo em seu interior, a qual projetava *slides* pintados em uma lâmina de vidro. Apesar da novidade em decorrência deste experimento, o mesmo despertou o interesse dos cientistas em utilizar tais dispositivos na área de entretenimento. Foi então que, em meados do século XVIII, a Lanterna Mágica passou a ser utilizada para este fim, por meio de exposições itinerantes, na projeção de *slides* de ilustrações. Vale ressaltar que, a movimentação das lâminas de vidros projetadas em uma superfície reta, criava a ilusão do movimento (LUCENA JUNIOR, 2002).

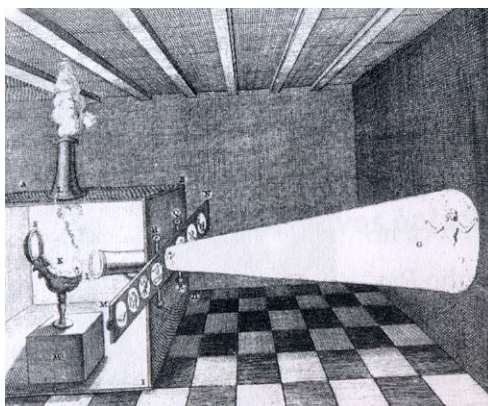


Figura 2.1: A lanterna mágica (Fonte: LUCENA JUNIOR, 2002).

Outra descoberta importante se refere ao artigo publicado por Peter Mark Roget, em 1824: “*The Persistence of Vision with Regard to Moving Objects*”, no qual o autor estabelecia a persistência de uma imagem por uma fração de tempo pelo olho humano, ao mesmo tempo em que outra imagem está sendo percebida (LUCENA JUNIOR, 2002). Desta forma, o olho humano combinava uma seqüência de imagens num único movimento, quando as mesmas eram exibidas rapidamente, com regularidade e iluminação adequadas. A partir dessa descoberta, surgiram inúmeras invenções que constituíram os objetos nos quais a animação é utilizada. Uma dessas invenções foi o taumatoscópio (figura 2.2), criado em 1825 e que se tratava de um disco preso em seus lados opostos por cordões e que possuía uma imagem na frente e outra em seu verso (LUTZ, 1998). Quando girado em uma velocidade rápida, o disco dá a impressão de que as imagens assumem uma única aparência, ou seja, uma continuidade que se obtêm da mistura óptica.

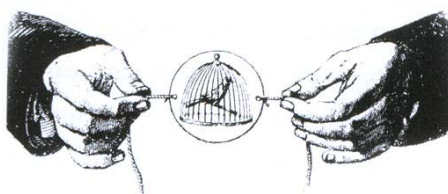


Figura 2.2: Taumatoscópio (Fonte: LUCENA JUNIOR, 2002).

Outras criações de destaque são: o fenaquistoscópio (figura 2.3), que consistia de dois discos, um com seqüências de imagens pintadas em torno do eixo e outro com frestas na mesma disposição; o estroboscópio (figura 2.4), que possuía apenas um disco e um orifício para o acompanhamento da seqüência de imagens; o zootrocópio (figura 2.5), cujas imagens

eram projetadas na parte interior do objeto circular para dar a ilusão de movimento quando rotacionado. Outra invenção foi o *kineograph*, mais conhecido como *flipbook*, que consiste de páginas com desenhos (ou fotografias) em seqüência, montadas como um livro. Quando as páginas são viradas rapidamente, a ilusão do movimento é criada, sendo considerado um meio prático e eficiente. Atualmente, esse recurso é amplamente utilizado ao se produzirem filmes baseados em animação com desenhos.

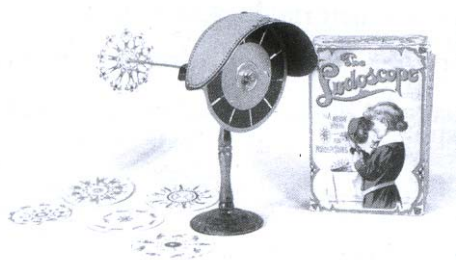


Figura 2.3: Fenaquistoscópio



Figura 2.4: Estroboscópio

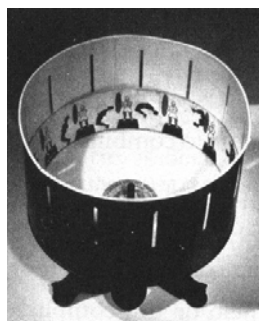


Figura 2.5: Zootrocópio

(Fonte: LUCENA JUNIOR, 2002).

Outra descoberta importante do século XIX foi o primeiro projetor de cinema, criado pelo austríaco Franz Von Uchatius, por meio da combinação da lanterna mágica com dois discos giratórios (um no qual se montavam dúzias de *slides* pintados em vidro e outro com frestas que atuavam como obturadores) (MORRISON, 1994). Além disso, pode-se destacar também, a criação do praxinoscópio, que se tratava de outro aparelho de projeção de imagens e que resultou no surgimento do famoso *teatro praxinoscópio*. Mas, somente a partir do surgimento da fotografia, em 1820, que foi possível iniciar as pesquisas acerca do uso de seqüências fotográficas na análise do movimento humano e animal (LUCENA JUNIOR, 2002).

Muitos anos se passaram até que os irmãos Lumière projetassem os primeiros filmes, apresentando em 1895 a primeira exibição de fotografias animadas, com o seu cinematógrafo. O ambiente propício à fantasia e ao imaginário, numa época em que viviam escritores como Júlio Verne, atiçava a imaginação popular. Assim, percebeu-se que a arte do cinema estava em trapacear a realidade.

Apesar dessas primeiras tentativas de animação de figuras, foi a partir do século XX que houve um substantivo avanço no desenvolvimento da tecnologia para a animação (LUCENA JUNIOR, 2002).

Foi em 1906, que o artista plástico James Stuart Blackton realizou o primeiro desenho animado, intitulado *Humourous Phases of Funny Faces*, filme curto, com a apresentação apenas de animação *frame a frame* em seqüências desenhadas por pouco instantes. Todavia, para que esse intento fosse realizado, Blackton teve um trabalho árduo e realizou outras técnicas experimentais (SOLOMON, 1987).

Outro desenho animado (*Gertie the Dinosaur*) foi lançado em 1914, por McCay, sendo sua produção considerada como grande marco da história da animação, da qual originaram vários dos princípios de animação (SOLOMON, 1987).

A partir dessas primeiras produções, outros filmes e artistas foram surgindo, trazendo consigo a evolução do desenvolvimento das técnicas de animação, bem como da tecnologia de projetores.

Desta forma, o desenvolvimento da animação e de suas técnicas só foi possível por meio dos resultados alcançados por pessoas que conseguiram unir técnica, imaginação e talento artístico.

Com o advento da computação gráfica, dos rápidos processadores e placas gráficas, a animação por computador tornou-se uma ferramenta poderosa para sua execução. Porém, apesar do surgimento de uma nova forma de animação proveniente da computação gráfica 3D, os conceitos e princípios desenvolvidos anteriormente de animação continuam a ser largamente utilizados neste novo meio de produção.

2.2 Princípios básicos da Animação

Segundo Frank Thomas e Ollie Johnston, dos estúdios Disney (TOMAS & JOHNSON, 1981), existem doze princípios da animação a serem considerados pelo animador. Estes são universalmente aceitos em qualquer produção de animação e influem diretamente no modo

como uma animação é feita e conseqüentemente no seu resultado. Foram vários anos de “*arte experimental*” até se chegar a esses doze princípios.

Dez desses princípios são aplicáveis, não somente em animação de personagens, mas também em animação facial. Os outros dois princípios (aparência e desenho sólido) estão intimamente relacionados à animação de desenho 2D.

Segue abaixo a descrição dos princípios básicos da animação:

- 1. Comprime e Estica:** É considerado o conceito mais importante e consiste em esticar ou comprimir as formas geométricas que constituem a parte deformada preservando o volume. Este princípio é usado para dar mais vida e expressão a um personagem, uma vez que, qualquer figura viva muda consideravelmente sua forma quando se move durante uma ação, o que não ocorre com objetos fixos, como um prato ou uma mesa. Assim, como é demonstrado na figura 2.6, a animação de uma bola caindo ganha mais expressão quando sua forma muda com a utilização do comprime e estica (*squash e stretch*).

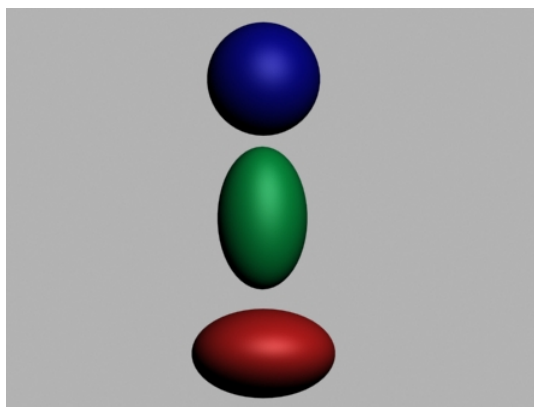


Figura 2.6: *Squash e Stretch.*

- 2. Antecipação:** Parte do princípio de que, para que haja o entendimento por parte das pessoas acerca de um acontecimento durante a animação, deve haver uma seqüência de ações que levem claramente de uma atividade à outra, afim de que o público possa estar preparado para o próximo movimento. Para tanto é necessário introduzir suaves mudanças de expressão ou uma enorme ação física. Este princípio está ilustrado na figura 2.7, aonde a bola vai realizar um movimento vertical em direção ao teto, porém antes ela antecipa o movimento, comprimindo-se, buscando uma melhor impulsão.

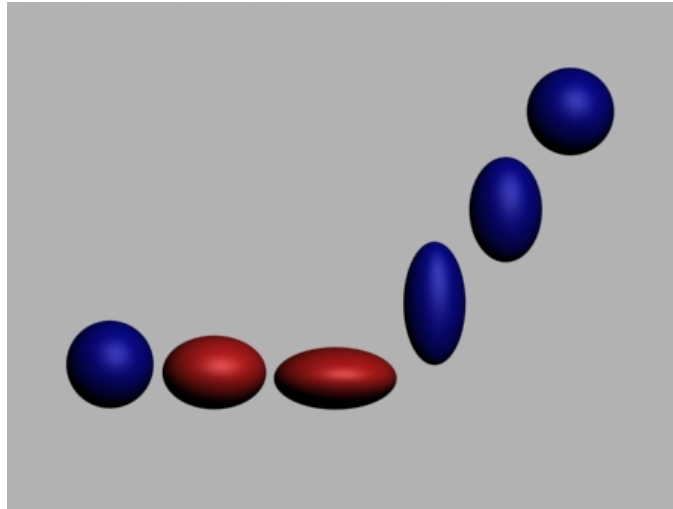


Figura 2.7: As bolas vermelhas demonstram uma antecipação de movimento.

- 3. Enquadramento:** Baseia-se na apresentação da ação de forma que fique claro visualmente para o espectador. Uma ação tem bom enquadramento quando a expressão é bem vista, ou seja, o movimento é claro e visível. A aplicação deste princípio pode ser visualizada na figura 2.8.



Figura 2.8: Enquadramento (Fonte: BLAIR, 1994).

- 4. Animação Direta (*Straight ahead*) ou Pose a Pose:** Existem dois métodos para animar uma cena, o “direto” (*straight ahead*) e o “pose a pose”. O método “direto” é usado em cenas de movimentos rápidos, no qual cada quadro é considerado chave. Para tanto, é realizada uma pose após a outra sem o estudo de *keyframe* (quadro-chave), o que deixa a cena mais espontânea e menos mecânica. Já o método “pose a pose”, é usado em situações em que há definição prévia da ação do personagem. Assim, neste caso, os *keyframes* e a quantidade de intervalos entre os mesmos são planejados cuidadosamente para alcançar o *timing* (tempo) desejado.

5. **Follow Through:** Consiste na aplicação das leis mecânicas da física em todos os elementos animáveis da cena, para que a mesma se torne mais convincente e suave.
6. **Slow In e Slow Out:** Relaciona-se ao uso de uma “chave de intervalação” na qual os desenhos se movem rapidamente de um extremo ao outro, como pode ser visto na figura 2.9. O uso de quadros desenhados perto dos extremos resulta em uma ação na qual o personagem vai menos rapidamente de uma pose a outra, dando a sensação de suavidade.

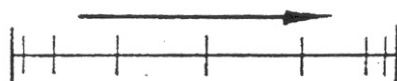
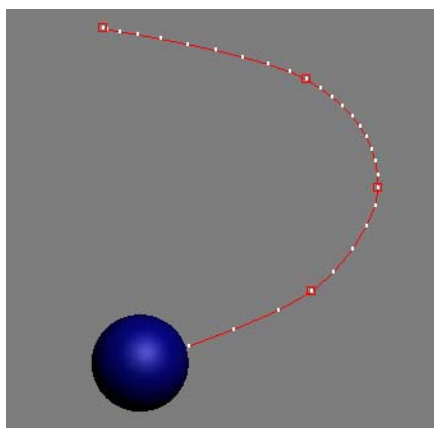
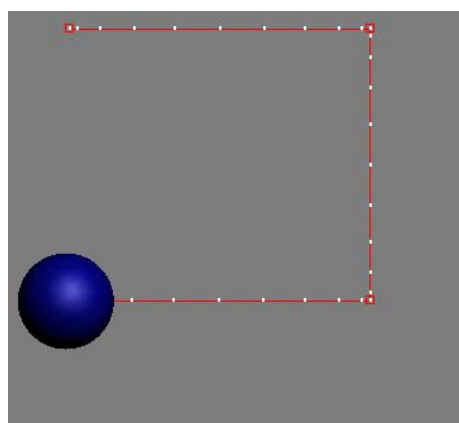


Figura 2.9: *Slow in e Slow out* (Fonte: HGN PRODUÇÕES Ltda, 1997).

7. **Arcos:** Relaciona-se ao uso de movimentos circulares ou em arcos para animar um personagem, possibilitando maior realidade do movimento, uma vez que os seres vivos normalmente executam movimentos segundo uma trajetória circular ou “arco” e raramente executam movimentos retilíneos horizontalmente ou verticalmente. Este princípio encontra-se ilustrado na figura 2.10.



(a)



(b)

Figura 2.10: Exemplo de animação utilizando arco (a) e não utilizando arco (b).

8. **Ações secundárias:** Parte do princípio de que as ações secundárias, como o movimento dos cabelos, movimentos de roupa, ou outro tipo de movimento em objetos, podem fortalecer uma cena dando mais realismo à mesma.
9. **Exagero:** Como mostra a figura 2.11, o exagero pode ser essencial para se alcançar uma boa comunicação do que a imagem significa. Assim, o uso do exagero torna-se fundamental, desde os personagens realistas aos mais caricatos. Vale ressaltar que é difícil precisar o grau de exagero necessário, uma vez que este depende da emoção a ser traduzida pelo personagem (PARKE & WATERS, 1996).

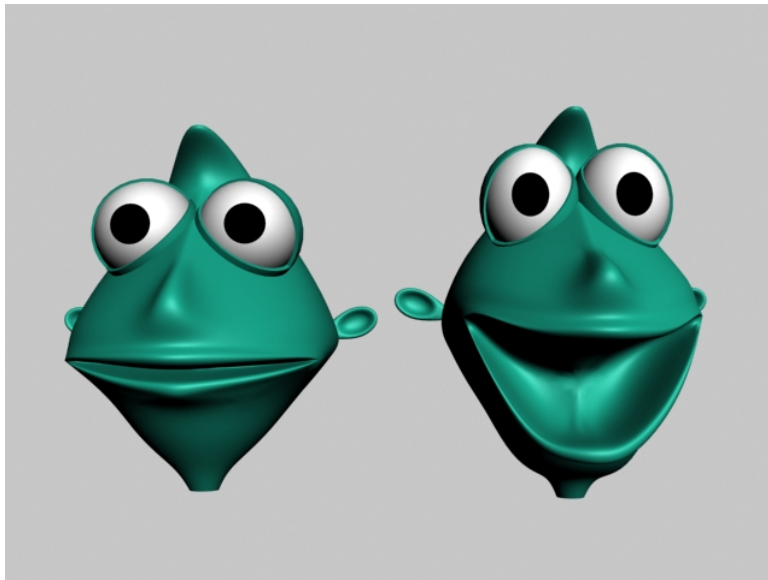


Figura 2.11: Exemplo de exagero no rosto da personagem.

10. **Design atraente:** O *design* do personagem torna-se imprescindível quando se trata da atração que ele deve exercer sobre as pessoas. Deste modo, os personagens devem possuir charme, simplicidade, comunicação e magnetismo (HGN PRODUÇÕES Ltda, 1997).

Capítulo 3: Animação Facial

Este capítulo tratará do desenvolvimento histórico da animação facial e introduzirá conceitos importantes sobre as expressões faciais e seus sistemas de codificações. Serão contemplados os principais paradigmas da animação facial, as aplicações existentes nesta área, assim como o conceito de *Lip Sync*, que é a sincronia labial com um arquivo de som de uma fala.

3.1. O Desenvolvimento da Área de Animação Facial

Historicamente, as primeiras imagens faciais geradas por computadores foram feitas por Parke durante seu curso de computação gráfica na Universidade de Utah no início dos anos 70. Parke começou com uma parte poligonal bem primitiva para representar a cabeça, o que resultou em uma animação dos vértices dos olhos e da boca abrindo e fechando. Henri Gouraud, aluno da mesma universidade, desenvolveu em sua dissertação um algoritmo denominado *smooth polygon shading algorithm*, aplicando-o em um modelo facial digitalizado de sua esposa. Este autor usou este algoritmo inovador para produzir algumas animações mais realísticas, uma vez que o algoritmo suavizava as superfícies poligonais. Realizou a coleta de dados poligonais de expressões faciais reais a partir da técnica de fotogrametria (uso de fotos) e simplesmente interpolou as expressões para criar a animação. Por volta de 1974, motivado pelo desejo de produzir rapidamente animações faciais, Parke completou seu primeiro modelo facial parametrizado (PARKE & WATERS, 1996).

Em 1971, Chernoff publicou seu trabalho usando desenhos de faces 2D gerados por computadores para representar um espaço de k -dimensões (CHERNOFF, 1971). De 1974 a 1978, o desenvolvimento de animação facial 3D foi muito pequeno, entretanto, o desenvolvimento da animação 2D usando sistemas de computador continuou na *New York Institute Of Technology*, *Cornell University* e mais tarde na *Hanna-Barbera* (empresa de animação). Estes sistemas deram suporte à animação de *cartoons* 2D e à animação facial destes (PARKE & WATERS, 1996).

Na década de 80 outros avanços foram alcançados. Platt (1980), publicou sua tese de mestrado na *Pennsylvania University* sobre um modelo de expressão facial baseado no controle muscular físico. Brennan (1982), no MIT, divulgou seu trabalho sobre técnicas de caricaturas faciais 2D produzidas por computador. Também no MIT, baseado neste trabalho,

Burson desenvolveu técnicas computacionais para envelhecer imagens faciais, especialmente em imagens de crianças.

A partir da metade dos anos de 1980, o desenvolvimento em animação facial teve uma considerável evolução. O filme de curta-metragem (pequena duração) animado chamado *Tony de Peltrie*, produzido por Bergeron & Lachapelle, em 1985, foi um marco para a animação facial. Este foi o primeiro filme de curta-metragem animado gerado por computador, onde expressões faciais 3D e de fala foram partes fundamentais no enredo da história (PARKE & WATERS, 1996).

Além disso, Waters (1987) divulgou seu novo modelo muscular enfocando a animação de expressões faciais. Este enfoque permitiu que uma variedade de expressões faciais fosse criada apenas controlando a musculatura da face. Um ano depois, Magnenat-Thalmann e colaboradores (1988) também desenvolveram um sistema parecido.

Em 1987, Lewis e em 1988, Hill divulgaram técnicas para uma fala sincronizada automatizada e para animações faciais (PARKE & WATERS, 1996).

Outro filme de curta-metragem animado inovador foi o *Tin Toy*, produzido em 1988 pela Pixar® (empresa de animação) e que recebeu um prêmio da academia de Hollywood Oscar®. Foi uma amostra da grande capacidade de animação facial gerada por computador. Em particular, um modelo de músculos foi usado para articular a geometria facial do bebê em uma variedade de expressões.

O desenvolvimento de *scanners* ópticos no início da década de 90, como o *Cyberware optical laser scanner* (Cyberware Laboratory, 1990) concedeu novas maneiras de aquisição de dados faciais.

Williams (1990) divulgou o uso de texturas de imagens faciais com o intuito de realizar animações de expressões 3D.

Lee, Terzopoulos e Waters (1993) divulgaram técnicas de mapear indivíduos em uma representação canônica da face.

Em outras áreas, como a de planejamento cirúrgico, também foram desenvolvidos trabalhos, quando Deng (1988) usou um modelo de elementos finitos da pele para simular incisões e fechamento por meio de pontos.

Nos últimos anos, grandes empresas de animação vêm trabalhando arduamente em projetos de longa e curta metragem usando, e ainda desenvolvendo, muitas técnicas de animação facial em 3D.

Com o advento das novas gerações de *consoles*, equipados com um poderoso *hardware*, as empresas de jogos estão investindo cada vez mais na animação facial de seus personagens.

O crescente volume de atividades em animação de personagens em geral, acarretou o aumento do interesse pela animação facial. Porém, o desenvolvimento da animação facial dos personagens não acompanhou o avanço da animação de outras partes do corpo. Isto se dá por alguns motivos como: a) o número elevado de músculos faciais, o qual dificulta a veracidade da animação; b) a especificidade das expressões faciais individuais, pois usamos a face para reconhecer as pessoas, além de conseguirmos detectar pequenas mudanças de expressões faciais, as quais variam de indivíduo para indivíduo; c) o processo histórico, pois a base da animação 3D veio da animação 2D, sendo que, neste caso, os movimentos da face não precisam de muito realismo para conseguir atingir um resultado eficaz, ao contrário do que ocorre na animação facial 3D; d) o grande número de peles dobráveis aumentando assim os polígonos do modelo facial; e) o tamanho do arquivo que armazena as informações e f) o tempo de *renderização* (processo de produção final da imagem contendo luz, textura e outros materiais aplicados na cena), podendo tornar a animação inviável e com um alto custo financeiro. Como evidenciado na figura 3.1, apesar do alto grau de realismo dos movimentos da personagem 3D, a animação facial da mesma, não conseguiu atingir o mesmo grau de realismo.



Figura 3.1: Globeleza Digital (Fonte: www.seagullsfly.com.br).

3.2. A Anatomia da Face

Um dos objetivos de se estudar a anatomia da face é se aproximar ao máximo do realismo e das características que esta possui quando se implementa uma animação facial.

Serão apresentados neste trabalho, apenas os principais músculos faciais e a nomenclatura dos ossos faciais.

A figura 3.2 denomina os planos de corte da face. O plano que corta a face em dois lados simétricos é chamado de mediana, conseqüentemente, o plano que corta a face não simetricamente é chamado de plano coronal (PARKE & WATERS, 1996).

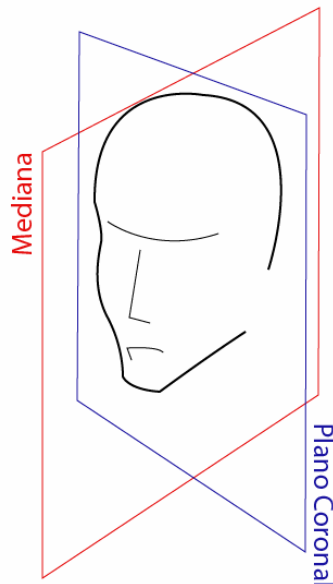


Figura 3.2: Cortes e nomenclatura usada.

A figura 3.3 mostra os principais ossos faciais, os quais podem ser divididos em duas classes principais: o crânio, que protege o cérebro e o esqueleto da face, cuja mandíbula (D) é o único osso com juntas livres para movimento. Os principais ossos da figura 3.3 são: (A) sutura internasal, (B) osso frontal, (C) osso nasal (D) mandíbula, (E) osso zigomático e (F) margem supra orbital. Os ossos são importantes para a modelagem 3D da face pois são eles que definem a posição dos músculos e da pele.

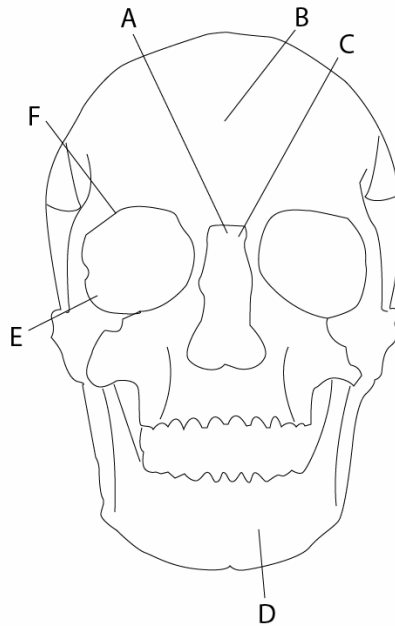


Figura 3.3: Ossos da face (Fonte: PARKE & WATERS, 1996).

A figura 3.4 define os principais músculos da expressão facial. Alguns destes músculos também exercem importantes funções como mover as bochechas e os lábios durante a mastigação e a fala ou até mesmo o fechamento e abertura dos olhos.

Os principais músculos da figura 3.4 são: (A) ondulador, (B) ondulador do supercílio, (C) levator superior nasal, (D) zigomático menor, (E) zigomático maior, (F) levator superior, (G) levator anguli oris, (H) mastigador, (I) bucinador, (J) mentalis, (K) orbicularis oris, (L) platysma, (M) depressor labil inferioris, (N) depressor anguli oris, (O) risorius, (P) zigomático maior, (Q) zigomático menor, (R) levator labil maior superior, (S) obicularis oculi maior, (T) epicranius maior.

Os músculos faciais trabalham em conjunto e nunca independentes. Um grupo de músculos trabalhando em conjunto e bem organizado leva a uma expressão facial bem definida.

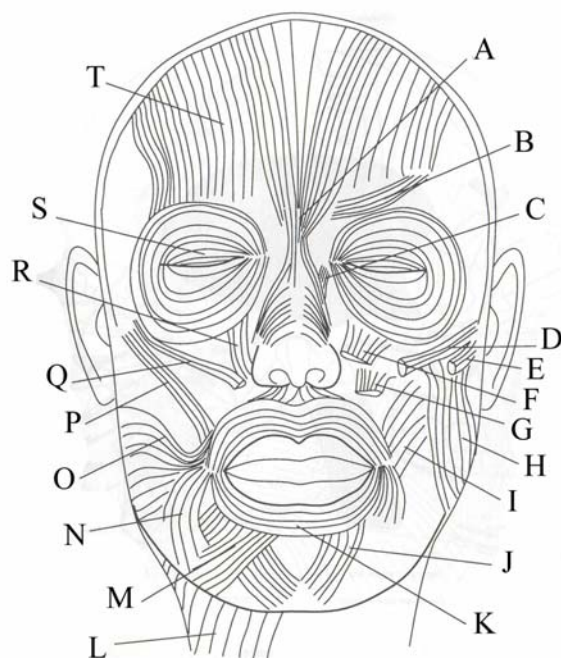


Figura 3.4: Músculos da face. (Fonte: PARKE & WATERS, 1996).

3.3. Análise das Expressões Faciais

As expressões humanas têm sido largamente investigadas pela comunidade científica. Em particular, o tópico sobre a universalidade de expressões faciais entre diversas culturas já consumiu considerável atenção. A forma exata como os músculos faciais interagem para expressarem estados emocionais é um tema que já foi objeto de estudo para muitos cientistas. Uma das primeiras publicações desta investigação foi de John Bulwer no final dos anos de 1640. Outros cientistas que estudaram o assunto foram Charles Bell, Duchenne de Boulogne e Charles Darwin.

Darwin foi o primeiro a demonstrar a universalidade das expressões e sua continuidade em homens e animais na obra *The Expression of the Emotions in Man and Animals*, sendo que essa obra teve um valioso valor para o estudo de expressões faciais, assim como para outras áreas, sendo que até hoje, cientistas vêm classificando e refinando as teorias postuladas por Darwin (PARKE & WATERS, 1996).

Contudo, a investigação mais marcante sobre as expressões faciais foi realizada por Duchenne (1962 *apud* LUCENA JUNIOR, 2002) que investigou as articulações faciais por meio da estimulação dos músculos faciais com eletrodos que geravam uma corrente elétrica nos pontos motores chaves da superfície da face. Assim, ele conseguiu manipular e gravar as

atividades musculares classificando-as. Este pesquisador classificou músculos em expressivos (completamente independentes), não-expressivos e pouco expressivos ou expressivos de maneira complementar. Além disso, categorizou as expressões advindas dos movimentos desses três tipos de músculos.

3.4. As Expressões Universais

Pesquisas em expressões faciais concluíram que existem seis categorias universais de expressões faciais que são reconhecidas entre várias culturas diferentes (EKMAN, 1989). Estas categorias são tristeza, raiva, satisfação, medo, aversão e surpresa (vide figura 3.5). Todavia, é importante ressaltar que cada uma destas categorias poderá ter diferentes intensidades e algumas variações individuais.

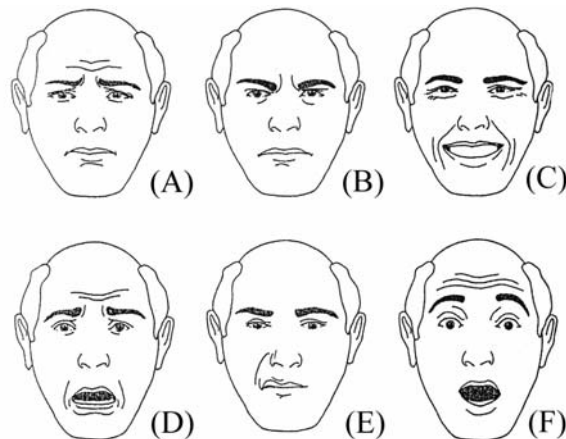


Figura 3.5: (A) tristeza, (B) raiva, (C) satisfação, (D) medo, (E) aversão, (F) surpresa. (Fonte: PARKE & WATERS, 1996).

3.5. A Linguagem *Mimic*

A linguagem *Mimic* (HJORTSJO, 1970), foi uma das primeiras a tentar investigar e sistematizar as atividades musculares que criam as diversas expressões faciais. A estrutura proposta por Hjortsjo, aponta que as expressões faciais são o resultado direto de aspectos estruturais estáticos da face junto com aspectos dinâmicos da mesma. Os aspectos estáticos são influenciados pelas condições mentais agindo sobre o sistema endócrino, sendo determinados pelos ossos faciais e pela formação da pele macia da face. Os aspectos dinâmicos são determinados pelas condições mentais e o estado emocional do indivíduo.

3.6. O Sistema de Codificação de Ação Facial - (*Facial Action Coding System- FACS*)

Desenvolvido por Paul Ekman e Wallace Friesen (EKMAN & FRIESEN, 1978), o sistema de codificação de ação facial (FACS) foi particularmente relevante à animação facial, pois quebrou as ações faciais em pequenas unidades chamadas *unidades de ação* (AUs). Cada AU representa uma ação muscular individual ou uma ação de um pequeno grupo de músculos em uma expressão facial reconhecível. No total, foram classificados 66 AUs, que em combinação, podem gerar expressões faciais bem definidas. Este sistema preocupa-se apenas com a descrição dos movimentos faciais e não com o que os mesmos significam. Atualmente, este esquema é o mais usado como base para o controle de expressão.

3.7. Paradigmas da Animação Facial

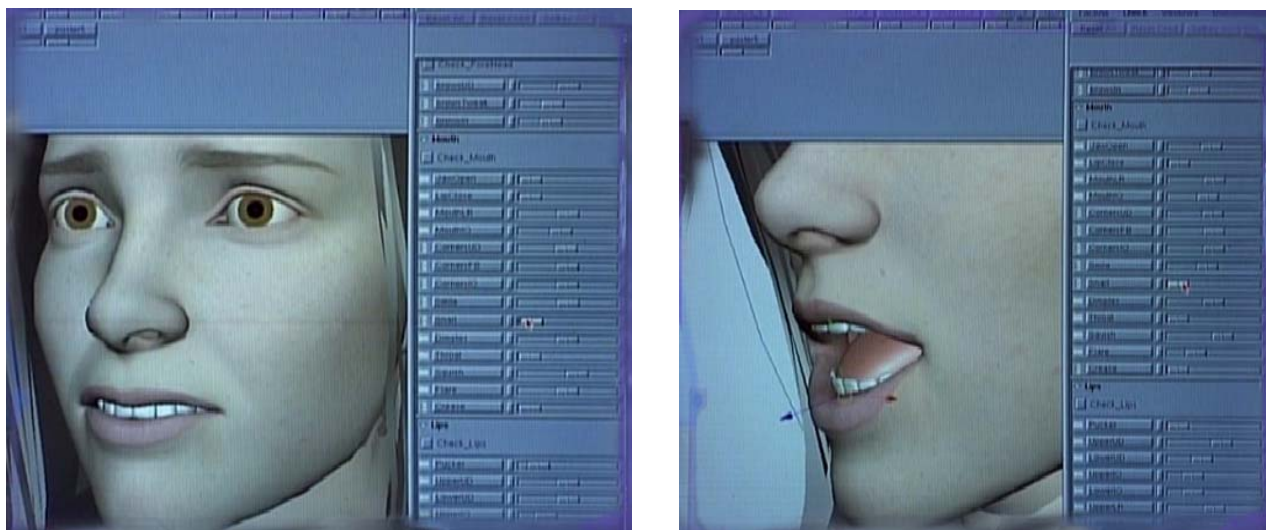
Neste tópico serão mostrados vários meios de se animar uma face. Cada paradigma de animação facial possui uma gama de características que serão abordadas.

3.7.1. Animação por Controle de Parâmetros (*Control Parameterizations*)

Neste caso, a animação torna-se um processo de especificar e controlar os parâmetros, atribuindo os valores em função do tempo.

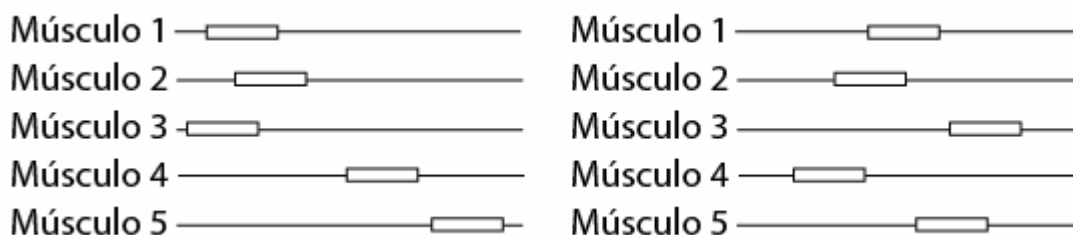
Este paradigma, pode ser visto como duas atividades independentes: (1) o desenvolvimento de controles de parâmetros, os quais são associados a uma interface que possa ser usada por um usuário, e (2) o desenvolvimento de técnicas para implementar a animação facial baseado nestes parâmetros.

Na figura 3.6, pode-se notar a diversidade de parâmetros que um modelo tridimensional possui, assim como o desenvolvimento da interface. O parâmetro de controle da abertura da boca, como é visto na figura 3.6 (a), começa com um valor padrão, que quando alterado pelo animador por meio da interface de *slider*, como é visto na figura 3.6 (b), a face 3D realizará o movimento proposto. No item (c) da figura 3.6 é possível ver claramente os nomes dos parâmetros e suas barras de controle antes (a) e depois (b).



(a)

(b)



(c)

Figura 3.6: Parâmetros de modelo 3D (*Aki Ross*) do filme Final Fantasy (Fonte: Making Of DVD Final Fantasy 2001)

3.7.2. Animação por Interpolação de Quadros-Chave (*Key Expression Interpolation*)

O conceito deste paradigma é o de coletar, de alguma maneira, os dados da geometria da face em dois quadros (ou expressões) diferentes. Deste modo, um simples parâmetro de controle, como o coeficiente de interpolação, é usado como função para mudar de uma pose (expressão) para outra.

A mudança da expressão da face é gerada devido ao movimento de cada ponto de controle da superfície (vértices) em pequenas distâncias durante sucessivos quadros (*frames*).

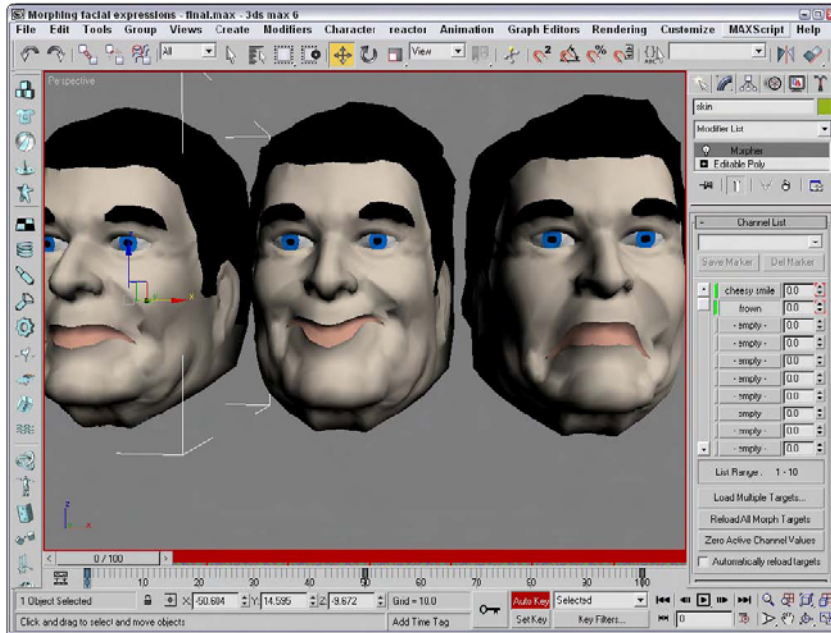


Figura 3.7: Animação por interpolação de quadros chaves (Fonte: MURDOCK, 2004).

A interpolação pode ser caracterizada pela forma linear, ou seja, a região entre os quadros-chave será ocupada por novos quadros criados igualmente espaçados, gerando um movimento constante, ou pela forma curva, que produz o efeito de aceleração e desaceleração devido à criação de novos quadros, irregularmente espaçados.

Essas características podem ser notadas na figura 3.8 , onde no item (a) indica uma interpolação linear descrevendo um movimento retilíneo uniforme e no item (b), uma interpolação curva descrevendo um início de movimento acelerado e um fim desacelerado.

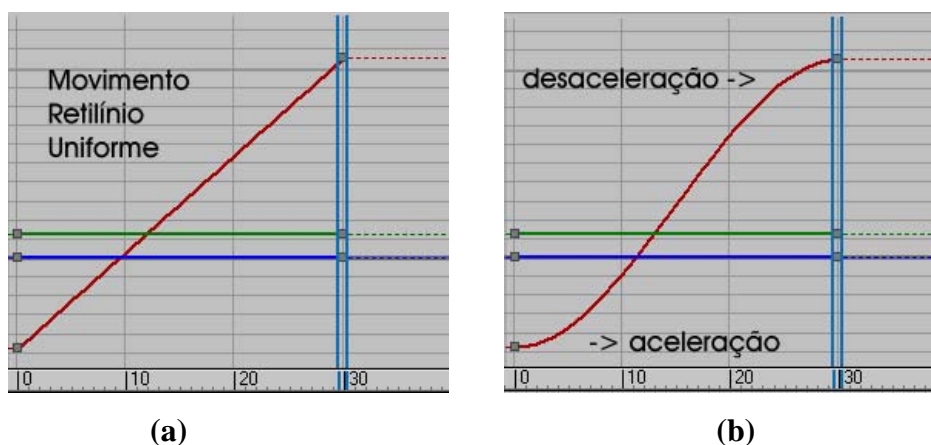


Figura 3.8: Características da Interpolação: (a) Linear e (b) Curva.

3.7.3. Animação Baseada em Performance Real (*Performance-Based Animation*)

Este paradigma envolve o uso de informações derivadas de ações de um modelo humano real para serem usadas em um modelo virtual 3D.

A animação facial pode ser realizada usando mecanismos de captura de expressões reais como o *motion capture*, que consiste na criação de uma representação 3D, a partir de uma performance real humana utilizando emissores (ópticos, mecânicos, magnéticos ou acústicos) que geram sinais captados por sensores estrategicamente distribuídos. Softwares especializados, que podem receber e “entender” os dados captados, geram as coordenadas tridimensionais.

3.7.4. Animação Baseada em Pseudo-Músculo (*Pseudomuscle-Based Animation*)

A complexa interação entre a pele, músculos e ossos da face resultam no que é normalmente chamado de expressões faciais, ou seja, sua interação resulta em uma vasta combinação de movimentos faciais.

A idéia básica deste paradigma é emular, sem grandes detalhes anatômicos, ações musculares básicas em modelos virtuais 3D.

3.7.5. Abstrações de Ações Musculares (*Abstract Muscle Actions*)

Magnenat-Thalmann desenvolveu um modelo de pseudo-músculo na qual os parâmetros possuíam controle em funções de abstrações de ações musculares (AMA) (LEE & MAGNENAT-THALMANN, 1998).

As funções AMA são similares, mas não as mesmas, que as funções FACS, pois as unidades FACS foram usadas como guia para a construção dessas funções AMA. Estas agem em regiões específicas da face aproximando a ação do músculo ou do grupo de músculos.

3.7.6. Deformações Livres de Formas (*Freeform Deformations - FFD*)

A FFD é uma técnica de deformar modelos de objetos sólidos. Pode ser usada para controlar mudanças de qualquer tipo e grau da forma da superfície como planos, quadrados, entre outros (PARKE & WATERS, 1996).

As simulações das ações de pseudomúsculos podem ser realizadas utilizando a técnica de deformações livre de formas.

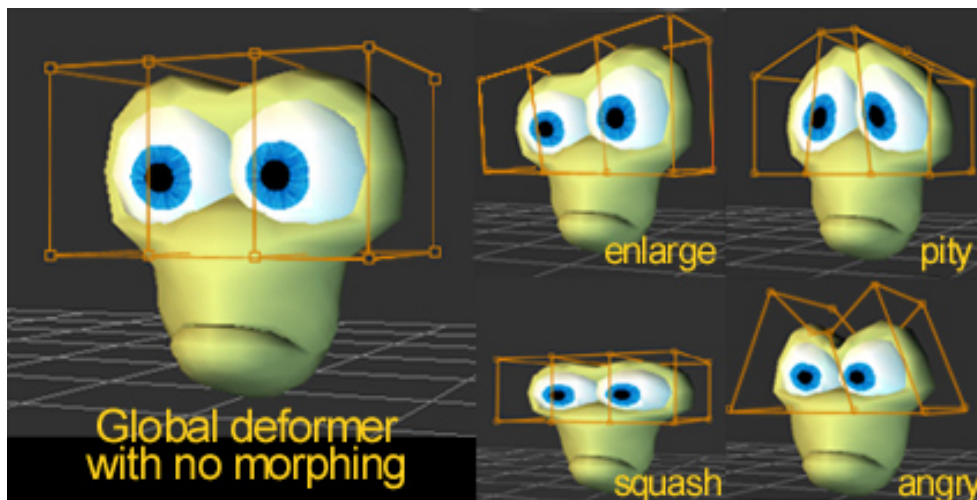


Figura 3.9: Exemplo de expressões utilizando a técnica de FFD.

(Fonte: http://www.xaloc.net/animations_bicho_tools.htm)

3.7.7. Animação Facial Baseada em Músculo (*Muscle-Based Facial Animation*)

O complexo detalhamento da anatomia da cabeça e da face é composto por ossos, cartilagens, músculos, nervos, vasos sanguíneos, glândulas, pele, junções e cabelo. Computacionalmente seria algo implementado com várias camadas.

A idéia básica deste paradigma é entender e modelar as ações dos músculos. Portanto, referências das *FACS* (seção 3.6) e *Mimic* (seção 3.5) são muito úteis para esta tarefa, além de novos estudos que podem ser feitos.

Um exemplo de um modelo baseado em músculo seria o de *Waters*, que desenvolveu um modelo dinâmico da face que incluía dois tipos de músculos: músculos lineares que puxavam e músculos *sphinchers* que contraíam. Seus músculos também tinham propriedades vetoriais de direção que eram independentes da estrutura óssea. Estes vetores fazem dos músculos modelados, independentes de uma topologia específica de face. Cada músculo possui uma zona de influência e são baseados nas *FACS* (PARKE & WATERS, 1996). Veja na figura 3.10, um exemplo de um modelo baseado em músculo usado no filme *Shrek*.



Figura 3.10: Modelo 3D na camada de músculos: Shrek (Fonte: Shrek, DVD Making Of, DreamWorks, 2002)

3.8. Sincronização Labial (*Lip Sync*)

Lip Sync pode ser definido como a sincronização dos rápidos movimentos labiais e da língua durante uma fala, determinando diferentes posições faciais.

O movimento labial e da língua durante a fala é considerado um componente importante da animação facial, pois de acordo com Frank, Hoch e Trogemann (1997), a habilidade de escutar e entender o que um indivíduo fala não vem somente do som, mas da associação deste com os movimentos labiais.

Portanto, para todos os indivíduos, a compreensão da fala é auxiliada pela informação visual, ou seja, o sincronismo labial é considerado uma atividade bi-modal. Contudo, para que esta bi-modalidade (informação visual mais acústica) se torne factível, há a necessidade de uma coerência espacial, temporal e da fonte, os quais estão descritos a seguir:

- **Coerência Espacial:** Capacidade humana de identificar coerência nos gestos faciais e sua informação acústica correspondente.
- **Coerência Temporal:** Velocidade da fala em sincronia com os movimentos labiais. Alguns problemas podem ocorrer, como por exemplo, em transmissões via satélite “ao vivo”, onde os sinais de imagens não estão sincronizados com os sinais da fala.
- **Coerência da Fonte:** As fontes de áudio e vídeo devem apresentar o mesmo conteúdo.

Os principais modos de realizar uma sincronização labial podem ser descritos como:

- **Data Driven:** Técnicas em que se utiliza uma seqüência de imagens reais para a implementação do sincronismo labial.
- **Speech Driven:** Técnicas que consistem no uso do som da fala como meio de determinar as poses labiais.
- **Híbrido:** Utiliza em conjunto as técnicas de *data driven* e *speech driven* para alcançar um resultado mais fidedigno (GUTIERREZ–OSUNA, 2005).
- **Text Driven:** Técnicas para aquisição de fonemas a partir de uma entrada de texto.

3.9. Áreas de Aplicação da Animação Facial

A grande motivação para o desenvolvimento de animação facial 3D, gerada por computadores, é a própria indústria de animação, a qual influencia sobremaneira o desenvolvimento da animação facial. Contudo, outras áreas que serão citadas aqui, também o fazem.

3.9.1. Indústria de Jogos

A indústria de jogos, a qual possui características de performance em tempo real, tem uma limitação correlacionada com o desenvolvimento e desempenho de processadores, arquiteturas e placas gráficas. Estas vêm se aperfeiçoando cada vez mais com a chegada de novas gerações de *consoles* de entretenimento digital, fazendo com que haja o aumento do uso de animações faciais realísticas.

3.9.2. Medicina

O planejamento de cirurgia crano-facial e simulação de costurar uma abertura na face são áreas da medicina que utilizam a animação facial. Modelos computacionais são normalmente gerados por tomografias da cabeça. O objetivo sempre é de simular a reação das componentes da cabeça ao passar por essas situações, ou seja, emular a resposta da pele e dos músculos faciais depois de terem sido cortados e reorganizados (LARRABEE, 1986).

3.9.3. Teleconferência

A habilidade de transmitir e receber sinais de imagens faciais é o núcleo de uma teleconferência, porém, a largura de banda para transmitir sinais pode ocasionar problemas de velocidade e performance. Esquemas de animação facial fazem com que uma câmera capte as expressões faciais de um usuário, cujos dados são compactados e enviados pela rede, chegando ao computador do outro usuário da teleconferência, que decodifica os dados da expressão recebida e realiza o processamento necessário para a obtenção de uma reprodução da expressão facial (PARKE & WATERS, 1996).

3.10. Exemplos de Sistemas de Animação Facial Existentes

3.10.1. CrazyTalk (Reallusion)

O *software Crazytalk* gera animações faciais a partir de fotos e um arquivo de som. Ao utilizar a textura de uma foto facial escolhida pelo usuário, o sistema aplica modificadores de *morphings* nesta imagem para a realização do sincronismo labial, a partir de um arquivo de som contendo uma fala. O reconhecimento de fonemas é a base do funcionamento deste sincronismo labial.



Figura 3.11: Face animada pelo *Software CrazyTalk* (Fonte: *Reallusion.com*).

3.10.2. Facial Studio (Digimation)

O *Facial Studio* trata-se de um *plug-in* do *software 3D Studio Max* contendo paradigmas musculares e de *morphings*, sendo utilizado para a modelagem da face 3D e configuração dos controles de animação facial de todas as partes do rosto como olhos e rugas.



Figura 3.12: Face modelada e animada pelo Facial Studio (Fonte: *digimation.com*).

3.10.3. LifeStudio: Head - Engine de Animação Facial do XBOX 360 (Lifemode Interactive)

O *engine* facial, denominado *LifeStudio: Head*, é capaz de produzir uma modelagem facial por parâmetros como sexo, cor, formato, entre outros. Ele também possui um banco de expressões faciais a serem usadas no modelo 3D, ferramentas de textura, sincronismo labial independente de língua que gera movimentos labiais com apenas um arquivo de som. Vale destacar ainda que, o *software* consegue comunicar com os principais pacotes gráficos e animar faces pelo paradigma muscular em tempo real.

Em fevereiro de 2006, a empresa produtora do *LifeStudio: Head* teve sua *engine* incorporada à arquitetura de multi-núcleos do console de terceira geração de games, o XBOX 360 da *Microsoft* (MICROSOFT TM, 2006)

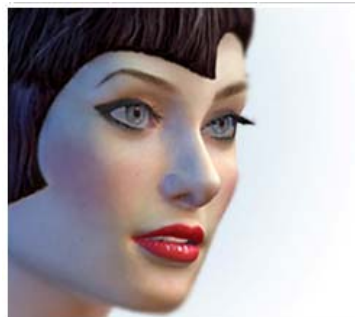


Figura 3.13: Face construída pelo *software* da *Lifemode Interactive* (Fonte:*lifemode.com*).

3.10.4. *Motion Capture* Óptico da Universidade de Ohio

O sistema por *motion capture* óptico da Universidade de Ohio é usado para simular os principais músculos faciais. A face, da figura 3.14, possui 85 fontes que serão captadas por 14 câmeras e depois repassadas para um software que manipulará os dados (SOMASUNDARAM, 2005).

Muitos avanços na indústria de animação por *motion capture* vem sendo realizado como mais pesquisas, melhoramento nos softwares e a especialização do equipamento, contudo, este ainda possui um alto custo financeiro, em torno de US\$ 20,000 a US\$ 150,000 (SILVA, 1997). Mais informações sobre os sistemas de *motion capture* encontram-se na seção 4.6.4.

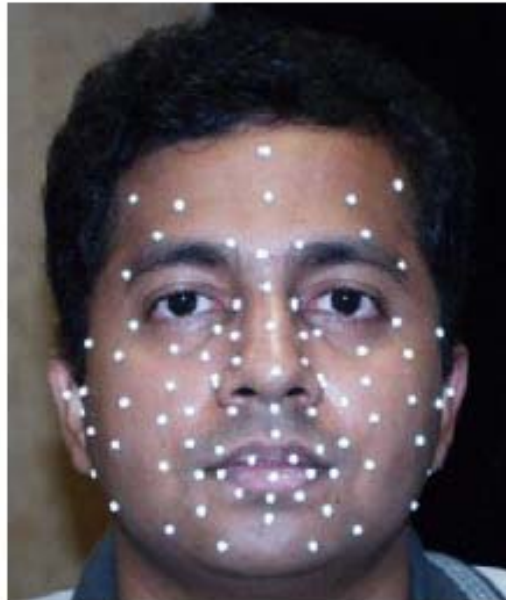


Figura 3.14: Face com fontes do *Motion Capture* (Fonte: SOMASUNDARAM, 2005).

3.10.5. Modelo Muscular de Parke

Este é um modelo de paradigma muscular que permite a seleção individual e a interação dos músculos. A definição de expressões faciais ocorre através de valores pré-definidos ou pela inserção manual de valores em cada músculo simulado.

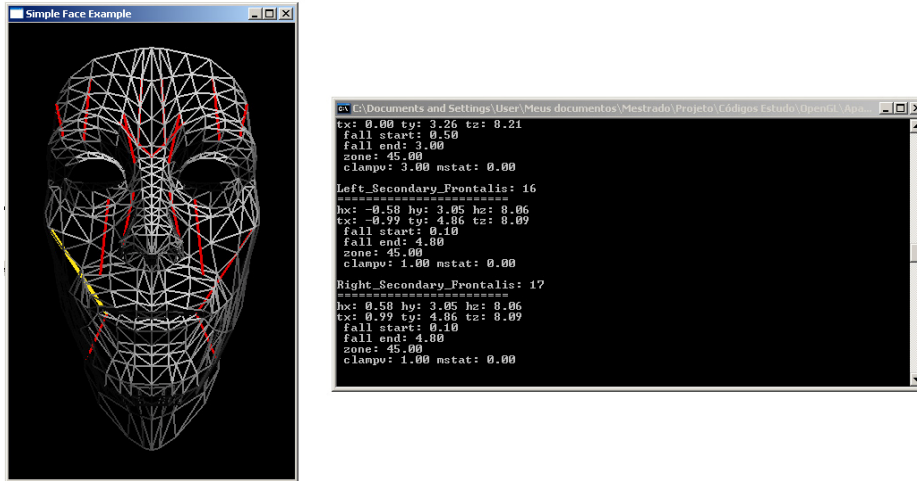


Figura 3.15: Modelo muscular utilizado por Parke (Fonte: PARKE & WATERS, 1996).

3.10.6. Modelo Muscular do “Max-Planck-Institut Informatik” (Alemanha)

O modelo muscular, do Instituto de Informática “Max-Planck”, é caracterizado por ser um sistema de paradigma muscular, que realiza a modelagem 3D de forma automática, através de seu *scanner 3D*. Vale ressaltar que, este modelo possui um sistema de massa-mola (*mass-spring system*) para camadas entre ossos, músculos e peles, realiza a simulação de músculos, aceita texturas e produz a sincronia labial a partir de um arquivo de som.

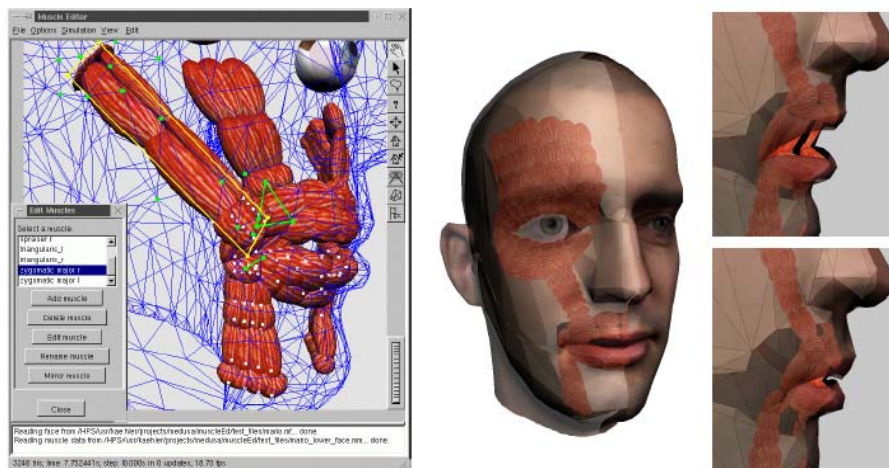


Figura 3.16: Modelo muscular posicionado na face 3D usado para a animação (Fonte: KAHLER, HABER, SEIDEL, 2001).

3.10.7. Magpie (Third Wish Software and Animation)

O *Magpie* é um *software* que utiliza o paradigma de *morphing*, com sincronização labial por reconhecimento de voz através de fonemas ou pela amplitude do som. Ele também

utiliza a técnica de roscopia (animação produzida referente a um vídeo real) para auxiliar a animação facial e o sincronismo labial, além de possuir um editor de curvas para o uso da técnica de interpolação. O *software* também exporta animação para outros formatos e aceita *scripts* no auxílio da animação facial.

3.10.8. Game Engine: Jupiter 3D (TouchDown Entertainment) – Módulo de Animação Facial

O *Jupiter 3D* é caracterizado por ser uma *engine* para *games*, o qual utiliza o sincronismo labial por fonemas (*speech driven*) e possui um banco de emoções por *morphing*.

3.10.9. Facade (Stanford Facial Animation System)

O foco do sistema *Facade* é a realização do sincronismo labial pelo paradigma *speech driven*. Contudo, este *software* também possui opções de câmera, luz, textura, modelagem e animação por parâmetros.



Figura 3.17: Resultado da animação facial pelo *Facade*: Na ordem da esquerda para a direita, expressão de bravo, aversão, medo, alegria, triste e surpreso.

(Fonte: www.dipaola.org/stanford/facade)

3.10.10. The Expression Toolkit

O *Expression ToolKit* realiza uma animação facial, com sistema de animação baseado em músculos (6 tipos de músculos), em tempo real. Este sistema também trata da animação dos olhos utilizando parâmetros e aceita *scripts* para o controle de músculos.



Figura 3.18: Resultados de expressões pelo *Toolkit*: Da esquerda para a direita, expressão neutra, bravo, medo e com a sobrancelha levantada. (Fonte: *expression.sourceforge.net*)

Capítulo 4: Estereoscopia

Este capítulo introduz e faz uma revisão à descoberta da visão estereoscópica. Este fato é algo fascinante e surpreendente, porém não é difícil notá-la na experiência do cotidiano. Para tal experiência, faz-se necessário o fechamento de um dos olhos por alguns minutos e por conseguinte, a sensação de um mundo mais plano, podendo a pessoa esbarrar em portas e perder o senso de distância (PARKER, 2004).

A estereoscopia é um dos métodos de obtenção de informação volumétrica, entre outros utilizados para a mesma função, como por exemplo, os de sensores ativos e passivos. A captação de informações volumétricas é de grande importância na animação facial, pois atua como instrumento na captura de movimentos de uma face real para uma virtual.

Este capítulo introduzirá e fará uma revisão sobre estereoscopia e seus métodos de busca por pontos correspondentes. Outros métodos de se obter informações tridimensionais também serão abordados neste capítulo.

4.1. Introdução a Estereoscopia

A estereoscopia é fundamentada no sistema de visão humano, que através do uso concomitante de dois olhos consegue perceber a profundidade.

Usando soluções geométricas entre os olhos e o valor da disparidade (nome dado à separação física de pontos semelhantes em duas imagens estereoscópicas), pode ser calculada a profundidade (eixo Z) dos pontos da cena.

A estereoscopia é uma técnica passiva, não sendo necessário o uso de sensores para detecção de profundidade. Seu uso estende-se a diversas áreas como cartografias, reconhecimento de objetos industriais e navegação de robôs portáteis.

4.2. O Olho Humano

Sem os olhos, não seria possível ter a noção de perspectiva, portanto eles são as principais ferramentas para a estereoscopia. Através deles, é possível registrar e levar ao cérebro as imagens dos objetos, com todas as suas características.

Em um sistema de estereoscopia, duas câmeras com a mesma capacidade e características são usadas para simularem os olhos humanos, tornando-se assim factível a obtenção de informações de profundidade a partir de fotos.

A figura 4.1 esquematiza um olho humano com suas partes principais, sendo que (a) mostra como a imagem é formada na retina, fazendo uma comparação com uma câmera e (b) mostra o esquema de transmissão das imagens dos olhos para o cérebro. A primeira função dos olhos é focalizar a luz. Seu funcionamento é similar ao de uma câmera fotográfica, onde os raios de luz penetram pela córnea, que possui um grande poder de focalização. A íris regula a quantidade de luz que entra nos olhos, aumentando ou diminuindo o tamanho da pupila (abertura central da íris). A luz então, viaja através do cristalino, que faz o ajuste fino na focalização sobre a retina, localizada na parte posterior do olho, atuando como se fosse o filme da câmera. A retina transforma a luz em impulsos elétricos, que são levados pelo nervo óptico até o cérebro, possuindo de 6 a 7 milhões de fotorreceptores, denominados cones, sensíveis a níveis altos de luminosidade e cores, e 75 a 150 milhões de fotorreceptores, denominados bastonetes sensíveis a níveis baixos de luminosidade (CARLBOM & PACIOREK, 1978).

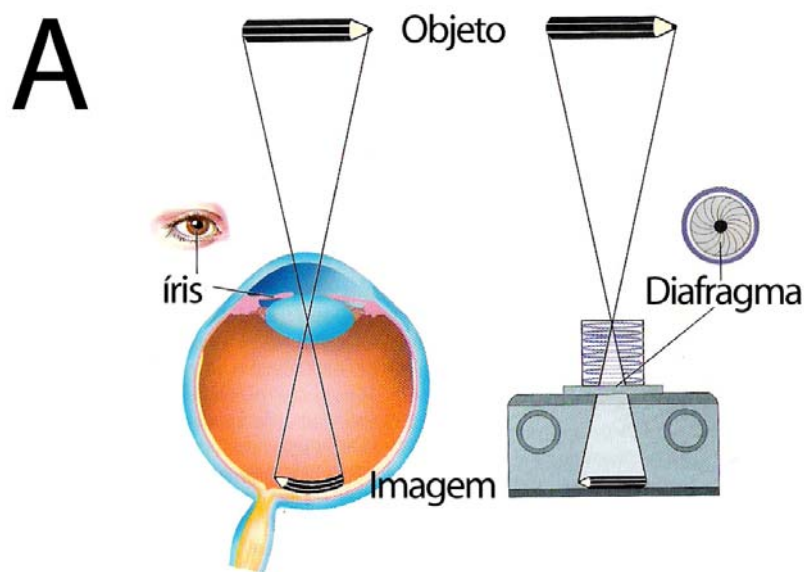


Figura 4.1 (a): O olho humano (Fonte: LENT, 2001).

A figura mostra como a imagem é formada na retina, fazendo uma comparação com uma câmera.

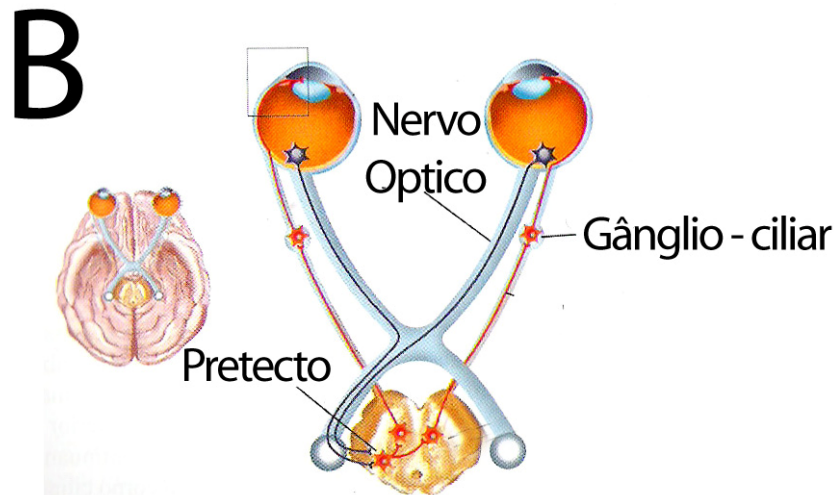


Figura 4.1 (b): A figura mostra o esquema de transmissão das imagens dos olhos para o cérebro (Fonte: LENT, 2001).

4.3. Projeção Perspectiva

Dado que a exibição de um objeto 3D em uma tela de computador ou em uma folha de papel exige o mapeamento de um sistema de coordenadas 3D em um 2D, operações de projeção são requeridas, da mesmo modo como no sistema de percepção de uma imagem pelo olho humano. Em geral, entende-se como projeção, o processo de mapear um sistema de coordenadas de dimensão “ n ” em um de dimensão menor ou igual a “ $n-1$ ” (DONALD & BAKER, 1997).

Uma transformação perspectiva projeta pontos 3D em um plano, ou seja, ela realiza uma aproximação do mundo 3D real. É importante ressaltar que a formação da imagem na retina do olho, ou na câmera fotográfica utiliza o processo da projeção perspectiva.

De acordo com a figura 4.2, percebe-se que o sistema de coordenadas da câmera (x, y, z) tem o plano da imagem coincidente ao plano xy e ao eixo óptico coincidente ao eixo z . O centro do plano da imagem localiza-se na origem e o centro da lente é a coordenada $(0, 0, \lambda)$. Sendo λ a distância focal das lentes, é assumido que o sistema de coordenadas da câmera está alinhada com o sistema de coordenadas do mundo (X, Y, Z) .

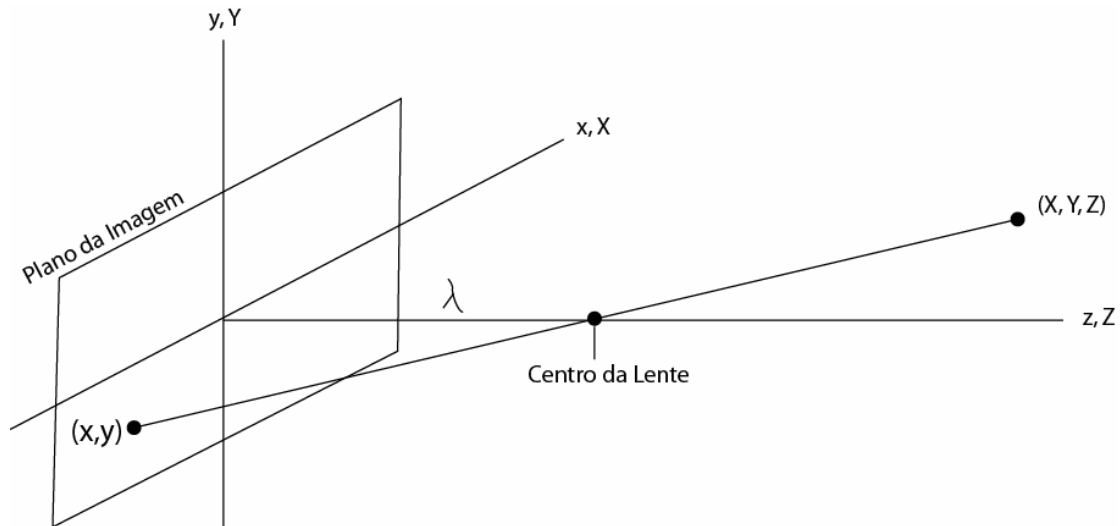


Figura 4.2: O sistema de coordenadas da câmera (x, y, z) está alinhado com o sistema de coordenadas do mundo (X, Y, Z) .

Considera-se $Z > \lambda$, ou seja, todos os pontos de interesse estão na frente da lente.

Com o objetivo de descobrir as coordenadas (x, y, z) a partir das coordenadas (X, Y, Z) , é feita a relação de semelhança de triângulos:

$$\frac{x}{\lambda} = -\frac{X}{Z - \lambda} = \frac{X}{\lambda - Z} \quad (1)$$

$$\frac{y}{\lambda} = -\frac{Y}{Z - \lambda} = \frac{Y}{\lambda - Z} \quad (2)$$

onde os sinais negativos que acompanham X e Y indicam que estes pontos estão invertidos.

Isolando x e y , pode-se deduzir que:

$$x = \frac{\lambda X}{\lambda - Z} \quad (3)$$

$$y = \frac{\lambda Y}{\lambda - Z} \quad (4)$$

Estas equações são não lineares pois elas envolvem divisão pela variável Z .

Embora seja possível usar as equações do jeito que se encontra, é mais conveniente expressá-las numa forma de matriz linear. Isto é facilmente alcançado usando as coordenadas homogêneas.

As coordenadas homogêneas de um ponto com coordenadas cartesianas (X, Y, Z) são definidas como (kX, kY, kZ, k) , onde k é um número arbitrário diferente de zero.

Obviamente, para converter coordenadas homogêneas para coordenadas cartesianas, é necessário a divisão das três primeiras coordenadas pela quarta.

Um ponto no sistema de coordenadas cartesianas pode ser expresso na forma de vetor como:

$$w = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (5)$$

e na forma de coordenadas homogêneas por:

$$w_h = \begin{bmatrix} kX \\ kY \\ kZ \\ k \end{bmatrix} \quad (6)$$

A matriz de transformação perspectiva é dada como:

$$p = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{-1}{\lambda} & 1 \end{bmatrix} \quad (7)$$

O produto Pw_h gera um vetor c_h (Vetor de Coordenadas Homogêneas da Câmera):

$$c_h = Pw_h \quad (8)$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{-1}{\lambda} & 1 \end{bmatrix} \begin{bmatrix} kX \\ kY \\ kZ \\ k \end{bmatrix} \quad (9)$$

$$= \begin{bmatrix} kX \\ kY \\ kZ \\ \frac{-kZ}{\lambda} + k \end{bmatrix} \quad (10)$$

os elementos de c_h são as coordenadas da câmera em forma homogênea.

A partir das equações 3 e 4, é possível deduzir as fórmulas 11 e 12:

$$X = \frac{x}{\lambda}(\lambda - Z) \quad (11)$$

$$Y = \frac{y}{\lambda}(\lambda - Z) \quad (12)$$

e portanto, estas mostram que não é possível achar z com uma imagem apenas, pois não é possível obter a variável z .

As equações 13, 14 e 15 mostram que tanto X , Y e Z dependem da variável desconhecida z .

$$X = \frac{\lambda x}{\lambda + z} \quad (13)$$

$$Y = \frac{\lambda y}{\lambda + z} \quad (14)$$

$$Z = \frac{\lambda z}{\lambda + z} \quad (15)$$

4.4. Princípios Básicos da Estereoscopia

4.4.1 Alinhamento de Câmera e Sistema de Coordenadas

Para realizar a transformação perspectiva (citada na seção 4.3), faz-se necessário o alinhamento entre o sistema de coordenadas da câmera e as coordenadas do mundo.

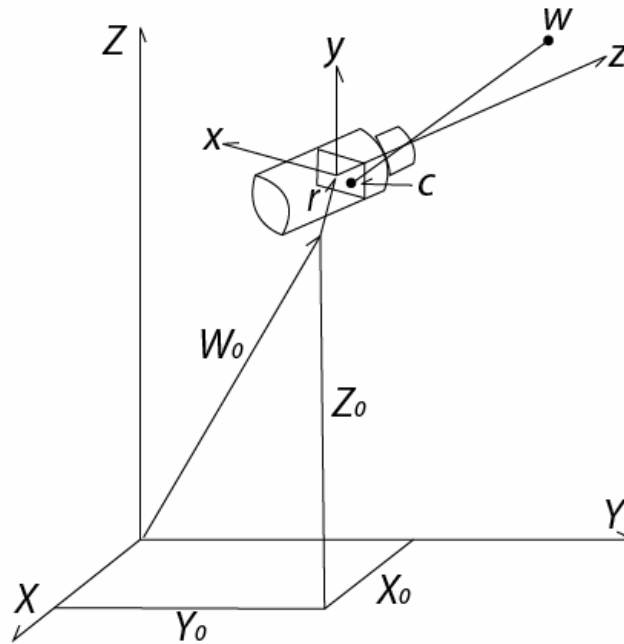


Figura 4.3: Sistemas desalinhados (Fonte: GONÇALEZ & WINTZ, 1987).

Sendo X , Y e Z o sistema de coordenadas do mundo, r o vetor de deslocamento da origem do plano da imagem, W_0 o vetor de deslocamento da câmera, X_0 , Y_0 e Z_0 as coordenadas da câmera no mundo, x , y e z o sistema de coordenadas não alinhado da câmera, W o ponto do mundo e c , o ponto captado pela imagem da câmera.

Tendo G , uma matriz de translação,

$$G = \begin{bmatrix} 1 & 0 & 0 & -X \\ 0 & 1 & 0 & -Y_0 \\ 0 & 0 & 1 & -Z_0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (16)$$

e R , uma matriz de rotação em torno dos eixos XZ ,

$$R = \begin{bmatrix} \cos \theta & \sin \theta & 0 & 0 \\ -\sin \theta \cos \alpha & \cos \theta \cos \alpha & \sin \alpha & 0 \\ \sin \theta \sin \alpha & -\cos \theta \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (17)$$

e finalmente, o deslocamento da origem do plano da imagem pelo vetor r é alcançado pela matriz de transformação C .

$$C = \begin{bmatrix} 1 & 0 & 0 & -r_1 \\ 0 & 1 & 0 & -r_2 \\ 0 & 0 & 1 & -r_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (18)$$

Chegando na fórmula final de : $c_h = PCR G w_h$, onde P é a matriz de transformação perspectiva.

4.4.2 Sistema de Estereoscopia

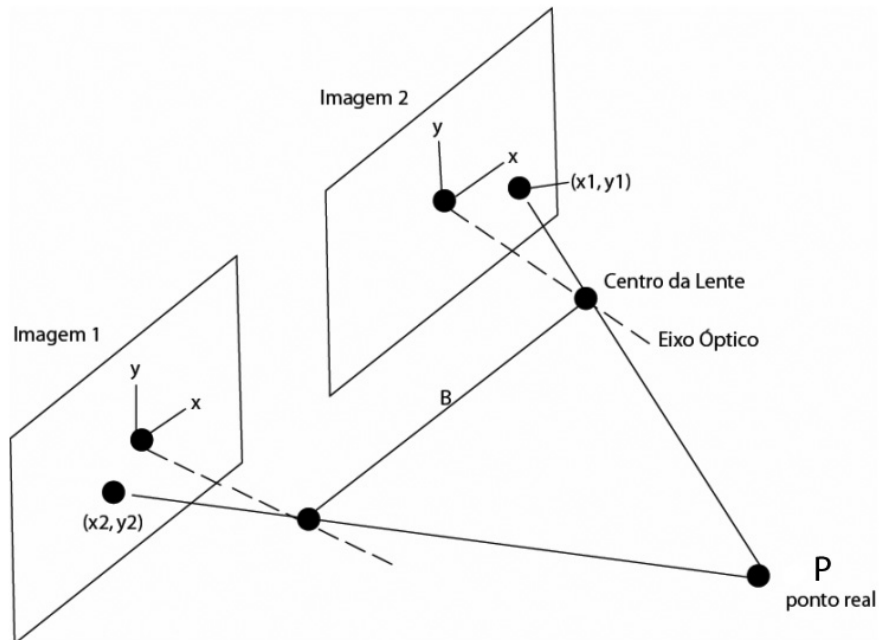


Figura 4.4: Modelo de processamento de estereoscopia.

A estereoscopia envolve o uso de no mínimo duas imagens separadas. Um único ponto \mathbf{P} (pertencente ao objeto real) estará em discussão. A distância entre os centros das duas lentes é chamado de linha base (*baseline*), e o objetivo é encontrar as coordenadas (X, Y, Z) do ponto \mathbf{P} tendo dois pontos de imagens $P_1(x_1, y_1)$ e $P_2(x_2, y_2)$.

Assume-se que as câmeras são iguais e que o sistema de coordenadas de ambas as câmeras estão perfeitamente alinhadas, diferindo apenas na localização de suas origens. Portanto, a coordenada Z de \mathbf{P} é exatamente a mesma para ambos os sistemas de coordenadas das câmeras.

Então, é conhecido pelas equações 11 e 12, que:

$$X_1 = \frac{x_1}{\lambda}(\lambda - Z_1) \quad (19)$$

$$X_2 = \frac{x_2}{\lambda}(\lambda - Z_2) \quad (20)$$

Entretanto, por causa da separação entre as câmeras, e visto que a coordenada Z de \mathbf{P} é semelhante para ambos sistemas de coordenadas das câmeras, segue que,

$$X_2 = X_1 + B \quad (21)$$

$$Z_2 = Z_1 = Z \quad (22)$$

sendo B a linha base, substituindo as equações 21 e 22 nas equações 19 e 20 tem-se:

$$X_1 = \frac{x_1}{\lambda}(\lambda - Z) \quad (23)$$

$$X_1 + B = \frac{x_2}{\lambda}(\lambda - Z) \quad (24)$$

e finalmente, tem-se a fórmula final de Z :

$$Z = \lambda - \frac{\lambda B}{x_2 - x_1} \quad (25)$$

Assim, pela equação 25, uma vez obtidas as coordenadas x_1 e x_2 das duas imagens e conhecendo-se λ e B , é fácil a obtenção de Z .

A tarefa mais difícil no processo de obtenção de informações de profundidade numa visão estereoscópica é achar dois pontos correspondentes nas duas diferentes imagens da mesma cena.

4.4.3 Calibração da Câmera

O alinhamento da câmera envolve várias variáveis como a distância focal, ângulos de rotação, deslocamento, entre outras. Embora estes parâmetros possam ser medidos diretamente, determinar um ou mais parâmetros usando a própria câmera como um dispositivo de medida é geralmente mais conveniente (especialmente quando as câmeras movem-se freqüentemente), porém, isto requer uma configuração de pontos previamente conhecidos do sistema mundo.

Portanto, o procedimento computacional usado para obter os parâmetros da câmera usando estes pontos conhecidos é chamado de calibração da câmera.

Sendo $A = PCRG$, (matriz perspectiva P, matriz transformação C, matriz rotação R e matriz de translação G), os elementos de A contém os parâmetros da câmera, e admitindo $k=1$ numa representação homogênea, então $\mathbf{c}_h = A\mathbf{w}_h$. Resultando assim através de substituições:

$$A = a_{11}X + a_{12}Y + a_{13}Z - a_{41}xX - a_{42}xY - a_{43}xZ - a_{44}x + a_{14} = 0 \quad (26)$$

$$B = a_{21}X + a_{22}Y + a_{23}Z - a_{41}yX - a_{42}yY - a_{43}yZ - a_{44}y + a_{24} = 0 \quad (27)$$

O procedimento da calibração consiste em:

1. Obter $m \geq 6$ (GONÇALEZ & WINTZ, 1987), para uma melhor precisão, pontos do mundo com coordenadas conhecidas (X_i, Y_i, Z_i) , $i = 1, 2, \dots, m$.
2. Imaginar esses pontos com a câmera numa dada posição para obter os pontos da imagem correspondente (x_i, y_i) , $i = 1, 2, \dots, m$.
3. Usar estes resultados nas equações 26 e 27 originados do alinhamento das coordenadas, para solucionar os coeficientes não conhecidos.

4.5. Buscando Pontos Correspondentes (*Matching Process*)

Denomina-se *matching* o processo de identificação dos pontos em ambas as imagens que resultem da projeção do mesmo ponto 3D real. Devido a fatores de oclusão (partes escondidas da câmera), brilho, iluminação entre outros, este processo é um problema fundamental na estereoscopia.

Neste trabalho, serão analisados alguns métodos de *matching* existentes.

4.5.1. Método LSM

Uma das maneiras de se buscar pontos correspondentes é baseada no método adaptativo conhecido como *least squares* (GRUEN, 1985), também chamado de algoritmo *least square method* (LSM).

Adota-se uma imagem de referência que define o ponto a ser buscado (*seed point*), e o processo de busca utiliza as imagens correspondentes para aplicação do LSM.

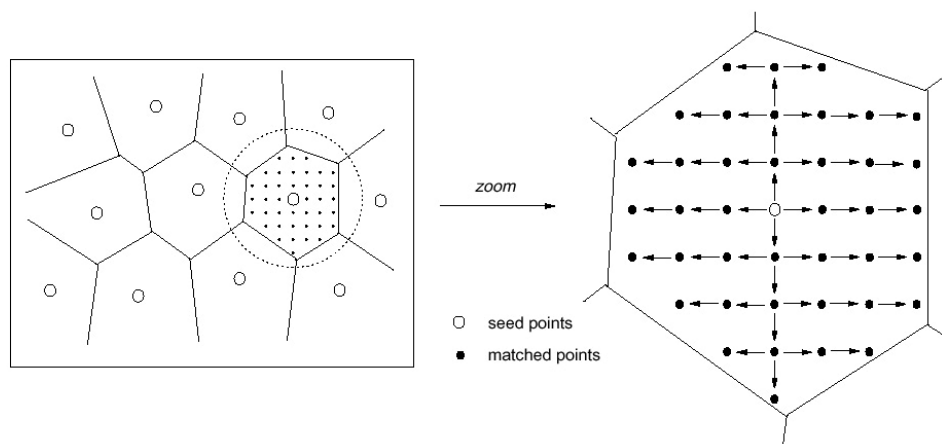


Figura 4.5: A partir dos *seed points* (pontos brancos), é aplicado o *voronoi tessellation* e a busca por pontos correspondentes (pontos pretos) é iniciada dentro das limitações da geometria (Fonte: GRUEN, 1985).

A busca automática produz uma densa e robusta configuração de pontos correspondentes, começando de apenas alguns *seed points*. Os pontos correspondentes podem ser manualmente definidos em cada imagem, gerados semi-automaticamente (definidos em apenas uma imagem) ou serem gerados de forma totalmente automático.

O método manual é apenas para casos onde o método automático falha. Os pontos correspondentes (*seed points*) têm de ser selecionados manualmente com uma aproximação de pelo menos 2 *pixels* em cada imagem e assim, o LSM é aplicado para localizar a posição exata.

No método semi-automático, os pontos correspondentes (*seed points*) têm de ser selecionados manualmente apenas na imagem de referência. Os outros pontos são gerados automaticamente como resultado da busca do melhor ponto correspondente ao longo da linha *epipolar* (interseção do plano da imagem com o plano que contém o ponto 3D procurado e o centro das duas lentes). Este modo é considerado o mais apropriado para casos normais onde a superfície é estática.

O método automático é recomendado para casos em que a superfície é dinâmica, como exemplo, uma seqüência de vídeo. Primeiro gera-se os pontos correspondentes na imagem de referência e depois realiza o processo de comparação com as outras imagens correspondentes.

Depois da definição dos pontos correspondentes, a imagem de referência é dividida em regiões poligonais de acordo com quais pontos correspondentes estão mais próximos (*Voronoi Tessellation*) - decomposição de um espaço métrico determinada pelas distâncias de um número de pontos discretos - ou seja, dado N pontos em um espaço euclidiano, há sempre um ponto " x " mais próximo de um ponto N , definindo assim, uma área geométrica de influência de proximidade e linhas de co-influência) (GRUEN, 1985). Começando dos pontos correspondentes previamente obtidos, outra gama de pontos correspondentes cresce automaticamente até que a região do polígono fique tomada.

O buscador de pontos correspondentes utiliza a seguinte estratégia: o processo começa de um ponto correspondente (*seed point*), muda horizontalmente na imagem de referência e faz a busca nas imagens correspondentes aplicando o LSM. A tomada de todo o polígono por pontos correspondentes é alcançada pelas mudanças horizontais e verticais. O processo é repetido para todos os polígonos até que toda a imagem esteja tomada pelos pontos.

4.5.1.1 Algoritmo LSM adaptativo (GRUEN, 1985)

Assumindo que duas regiões de imagens são denotadas como funções discretas bidimensionais $f(x,y)$ e $g(x,y)$, que podem ser definidas como regiões conjugadas de um par estereoscópico fotográfico esquerdo e direito, respectivamente. Pode-se interpretar $f(x,y)$ como 'padrão', e $g(x,y)$ como 'amostra'. Assim, a correlação é estabelecida se

$$f(x,y) = g(x,y). \quad (28)$$

Devido aos efeitos aleatórios em ambas as fotografias, a equação (28) não é consistente. Portanto, um vetor ruído $e(x,y)$ é adicionado, resultando em

$$f(x,y) - e(x,y) = g(x,y). \quad (29)$$

A localização do valor $g(x,y)$ deve ser determinada para a obtenção do ponto de registro (*match point*). Isso é obtido minimizando uma função alvo que mede as distâncias entre os níveis de cinza no *padrão* e na *amostra*. A função alvo a ser minimizada nesta abordagem é a norma- L_2 dos residuais da estimação de mínimos quadrados.

No contexto dos mínimos quadrados, a equação (29) pode ser considerada como uma equação de observação não-linear, que modela o vetor de observações $f(x,y)$ com a função $g(x,y)$, onde a localização da fotografia direita deve ser estimada. A localização é descrita por parâmetros de deslocamento Δx , Δy , que são contados com respeito a uma posição inicial de $g(x,y)$, a aproximação da região da *amostra* conjugada $g^0(x,y)$.

Para considerar uma variedade de deformações de imagem e para obter um melhor registro, parâmetros de morfologia da imagem e correções radiométricas são introduzidas além dos parâmetros de deslocamentos.

Se os níveis de cinza são dados sobre uma grade, a morfologia da imagem é obtida por re-amostragem de $g^0(x,y)$ sobre os pontos de grade transformados. A transformação geométrica é modelada por um polinomial bivariado

$$x = t_y^T A t_x, \quad (30 \text{ a})$$

$$y = t_y^T B t_x,$$

$$\text{com } t_x^T = \{1 \ x_0 \ x_0^2 \ \dots \ x_0^{m-1} \}, \quad (30 \text{ b})$$

$$t_y^T = \{1 \ y_0 \ y_0^2 \ \dots \ y_0^{m-1} \}, \quad (31 \text{ a})$$

$$(31 \text{ b})$$

onde as matrizes A, B são dadas por

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{bmatrix}, \quad B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mm} \end{bmatrix} \quad (32)$$

e x_0 e y_0 apontam para as localizações na grade dos pontos $g^0(x,y)$.

Os parâmetros de transformação $a_{11}, \dots, a_{mm}, b_{11}, \dots, b_{mm}$ devem ser estimados de (29). Para que seja possível operar com a abordagem de mínimos quadrados convencional a função $g(x,y)$ em (29) deve ser linearizada. Isso permite escrever

$$f(x,y) - e(x,y) - g^0(x,y) + \frac{\partial g^0(x,y)}{\partial x} dx + \frac{\partial g^0(x,y)}{\partial y} dy. \quad (33)$$

$$dx = \frac{\partial x}{\partial p_i} dp_i,$$

$$dy = \frac{\partial y}{\partial p_i} dp_i,$$

onde $p_i = i$ -ésimo parâmetro em (30), especificando A e B como

(34 a)

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & 0 \end{bmatrix}, \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & 0 \end{bmatrix}$$

e obtendo a transformação

(34 b)

$$x = a_{11} + a_{12}x_0 + a_{21}y_0,$$

$$y = b_{11} + b_{12}x_0 + b_{21}y_0.$$

A equação (34 b) também inclui os parâmetros de deslocamento $\Delta x, \Delta y$, que são denotados por a_{11} e b_{11} .

A diferenciação de (34 b) resulta em

$$dx = da_{11} + x_0 da_{12} + y_0 da_{21}, \quad (35)$$

$$dy = db_{11} + x_0 db_{12} + y_0 db_{21}.$$

Usando as notações simplificadas

$$g_x = \frac{\partial g^0(x,y)}{\partial x}, \quad g_y = \frac{\partial g^0(x,y)}{\partial y}$$

e adicionando um parâmetro de deslocamento radiométrico r_s à equação (33) resulta com a equação (35) em

$$f(x,y) - e(x,y) = g^0(x,y) + g_x da_{11} + g_x x_0 da_{12} + g_x y_0 da_{21} + g_y db_{11} + g_y x_0 db_{12} + g_y y_0 db_{21} + r_s \quad (36)$$

Combinando os parâmetros em (36) como parâmetros vetoriais de x , onde

$$x^T = \{da_{11}, da_{12}, da_{21}, db_{11}, db_{12}, db_{21}, r_s\},$$

os coeficientes na matriz A , e a diferença vetorial $f(x,y) - g^0(x,y)$ em ℓ , as equações de observação são obtidas em notação clássica (com $e = e(x,y)$) como

$$\ell - e = Ax. \quad (37 a)$$

Com o operador espectância estatística E e assumindo que

$$E(e) = 0, \quad E(ee^T) = \sigma_0^2 P^{-1} \quad (37 b)$$

o sistema (37) é o modelo de estimação Gauss-Markov.

A estimação de mínimo quadrado no modelo (37 a), (37 b) leva aos estimadores de variância mínima

$$\hat{x} = (A^T P A)^{-1} A^T P \ell, \quad \text{vetor solução} \quad (38 a)$$

$$\hat{\sigma}_0^2 = \frac{1}{r} v^T P v, \quad \text{fator de variância} \quad (38 b)$$

$$v = A\hat{x} - \ell, \quad \text{vetor residual} \quad (38 \text{ c})$$

$$r = n - u$$

onde r é a redundância, u é o número de parâmetros de transformação e n é o número de observações.

Os valores residuais v_i podem ser interpretados como diferenças nos níveis de cinza entre a *amostra* e o *padrão*, na região do ponto de registro.

$$v_i = \hat{g}(x, y)_i - f(x, y)_i, \quad i = 1, \dots, n \quad (39)$$

onde o símbolo $(\hat{})$ significa estimador de mínimo quadrado.

Assim, $\hat{\sigma}_0$ é um estimador *a posteriori* para uma diferença de ruído do *padrão* e da *amostra*.

Se os valores $g(x, y)$ em (29) são realmente quantidades estocásticas, defrontamos com o caso em que os elementos da matriz A não são fixos, mas quantidades estocásticas. Um procedimento de estimação estrita deve levar em consideração esse fato. Evitando complicações desnecessárias as propriedades estocásticas de A são ignoradas. Isso permite a aplicação do algoritmo padrão computacionalmente eficiente para a estimação de mínimos quadrados como delineado para (37 a) e (37 b).

Devido a não-linearidade da equação (29), a solução final é obtida iterativamente. Com as primeiras aproximações

$$a_{11}^0 = b_{11}^0 = a_{21}^0 = b_{12}^0 = 0, \quad a_{12}^0 = b_{21}^0 = 1 \quad (40)$$

resulta no conjunto de coordenadas

$$x_i = x_{0i}, \quad y_i = y_{0i}, \quad i = 1, \dots, n$$

para o passo da primeira iteração (n = número de pontos de grade no *padrão* ou *amostra*).

Após o vetor solução (38 a) ser obtido, a transformação (34 b) é aplicada e $g^0(x,y)$ é reamostrada sobre o novo conjunto de coordenadas, e a matriz A é reavaliada. A iteração pára se cada elemento do vetor de alteração \hat{x} (38 a) for abaixo de um certo limite (c_1, \dots, c_6)

$$\begin{aligned} |da_{11}| < c_1, & |db_{11}| < c_2, \\ |da_{12}| < c_3, & |db_{12}| < c_4, \\ |da_{21}| < c_5, & |db_{21}| < c_6. \end{aligned} \tag{41}$$

4.5.2. *Matching* por Contornos com operadores morfológicos

A primeira etapa do método consiste em detectar os contornos das imagens estereoscópicas originais (esquerda e direita) obtendo assim duas imagens binárias. O motivo da detecção de bordas é devido ao excesso de informações irrelevantes que as imagens originais apresentam. Contudo, com as novas imagens binárias geradas, ainda é possível encontrar informações irrelevantes, que aumentam o espaço de procura no *matching*, como contornos “grossos” (vários pontos definido o mesmo contorno) e pontos isolados próximos ao contorno.

Dois operadores morfológicos correspondentes às operações de erosão (*erosion*) e dilatação (*dilation*) são usados em imagens binárias para a simplificação deste tipo de problema. A partir destes, é possível gerar outros operadores, como por exemplo, o operador *clean*, que elimina pontos isolados na imagem. Logo após a aplicação do *clean*, há uma necessidade de aplicar a operação *close* para fechar lacunas existentes, que por ventura ficaram abertas. A aplicação do operador *close*, embora tenha conciliado os *pixels* correspondentes ao mesmo contorno, adiciona *pixels* irrelevantes. Portanto, para reduzir a largura dos contornos e diminuir ainda mais o espaço de procura do *matching*, é necessário a aplicação do operador *thinner*.

Cada ponto do contorno corresponde a uma linha epipolar na outra imagem, porém essa mesma linha pode passar por outros N diferentes pontos da outra imagem. A solução encontrada para quando isto ocorre é a utilização de um parâmetro de brilho da imagem original, que deverá ficar centrada sobre o ponto que se deseja fazer o *matching*. Os valores mais próximos de brilho identificarão a correspondência desejada.

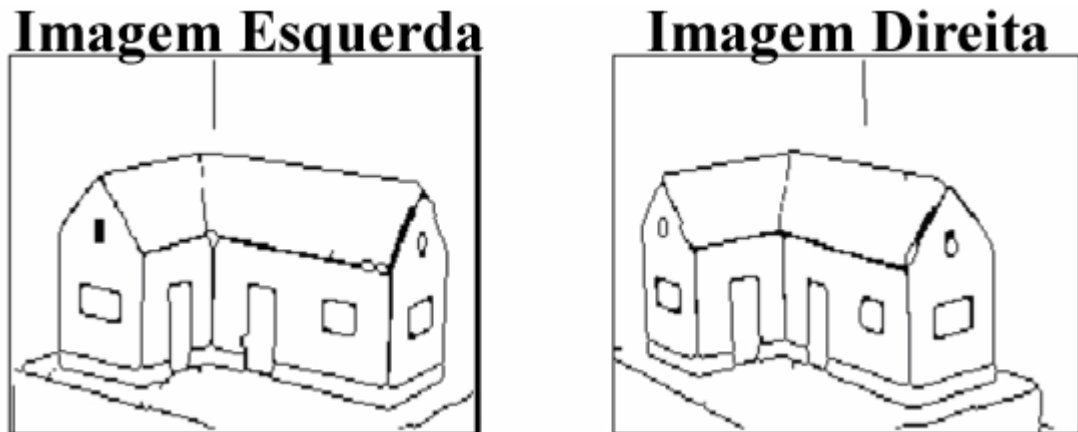


Figura 4.6: *Matching* por Contornos com operadores morfológicos (Fonte: VALLE, 2004).

4.5.3 *Matching* por Regiões

Atribui-se uma cor a um ponto inicial, o qual pode ser escolhido automaticamente ou aleatoriamente, e dentro da região do contorno é realizada a aplicação da técnica *flood fill* (implementada em qualquer software de tratamento de imagem), ou seja, iterativamente, os pontos vizinhos adquirem a mesma cor até chegarem ao contorno. Portanto, a região é definida pelos pontos coloridos.

Todas as regiões devem ser identificadas para que o processo identificado acima termine. Logo em seguida, realiza-se o *matching* de todas as regiões em ambas as imagens. Esta etapa consiste em pegar os pontos de contorno de uma região da imagem da esquerda e verificar quais pontos de contorno da imagem da direita de uma mesma região se combinam. Este problema pode ser subdividido tantas vezes quantas as regiões identificadas.

A forma de reconstrução é similar ao método anterior. Para cada ponto do contorno da região da imagem da esquerda, está associada uma reta epipolar que intersecta os contornos da região na imagem da direita.

Os pontos que definem os contornos das regiões em ambas as imagens, $C = \{(x,y)_1, (x,y)_2, \dots, (x,y)_N\}$, devem estar ordenados sequencialmente, de forma a minimizar as distâncias entre dois pontos consecutivos.

Ao percorrer o contorno da região na imagem da esquerda, devem ser apenas selecionados os pontos resultantes da intersecção das retas epipolares com os contornos da região na imagem da direita, que satisfaçam a continuidade, ou seja, pontos que se encontram numa dada vizinhança do ponto anterior. Finalmente, basta identificar a cor ou até mesmo o

padrão indicativo dessa região. Para tal, pode ser efetuada uma média dos valores em *RGB* dos pontos que se encontram no interior dessa região.

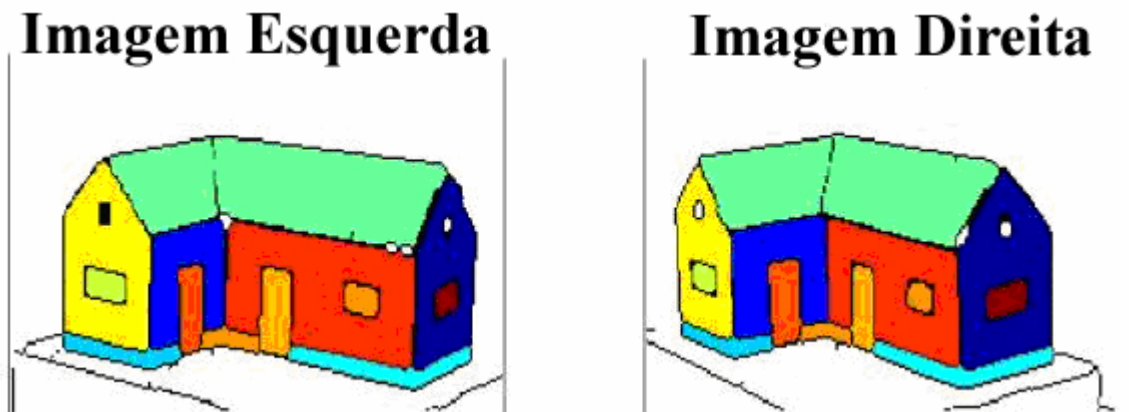


Figura 4.7: *Matching* por regiões (Fonte: VALLE, 2004).

4.6. Outros Métodos de se Obter Informações Tridimensionais

4.6.1. Introdução aos Sensores Ativos

O sensor ativo é geralmente definido como aquele que provê e controla sua própria iluminação. Irradia energia sobre as superfícies a serem medidas e computa a distância pela energia refletida. Tal medida pode ser feita por tempo de voo do iluminador ou por triangulação de um padrão projetado. Uma variedade de fenômenos físicos pode ser usada como fonte radiante (FRANÇA, 2002).

4.6.2. *Scanner* a Laser

Este sistema de aquisição volumétrica é constituído, basicamente, de duas partes: a) sensoriamento – onde a geometria de superfície da cena de interesse é capturada e b) remontagem – que usa como entrada os dados brutos do sensoriamento e os transforma em objetos poligonais tridimensionais (FRANÇA, 2002).

Os *scanners* a laser usam um feixe de luz focado e empregam o tempo de voo (TOF – *Time Of Flight*), ou diferença de fase, para encontrar a distância. Os de tempo de voo medem o intervalo de tempo entre a emissão de um pulso laser de menos de um nanosegundo e a recepção do eco ou feixe refletido. O relacionamento entre o tempo e a distância é:

$$r = \frac{ct}{2} \quad (42)$$

onde r é a distância, c é a velocidade da luz, e t é o intervalo de tempo entre emissão e recepção da onda.

Já os sensores de diferença de fase usam um feixe laser com onda contínua e amplitude modulada e medem a diferença de fase entre a onda emitida e a onda sensorizada. O relacionamento entre a diferença de fase $\Delta \Phi$ e a distância r é dada por:

$$r = \left(\frac{c}{4\pi f_{am}} \right) \Delta \Phi \quad (43)$$

onde f_{am} é a frequência modulada.

Um aspecto importante que deve ser considerado no uso de *scanners* a laser é a potência do sensor, pois sendo potente, o sinal refletido é mais forte e conseqüentemente gera uma melhor precisão. No entanto, para aplicações em faces humanas, eles são considerados lentos, devido ao seu tempo de varredura.

4.6.3. Luz Estruturada

Sistemas de luz estruturada são sensores ativos baseados em triangulação. Uma luz em forma de ponto, malha, ou grade é projetada sobre o objeto, e essa forma é capturada por uma câmera. A distância até o objeto pode ser calculada simplesmente por geometria, dado o conhecimento a priori das posições do projetor emissor de luz e do detector.

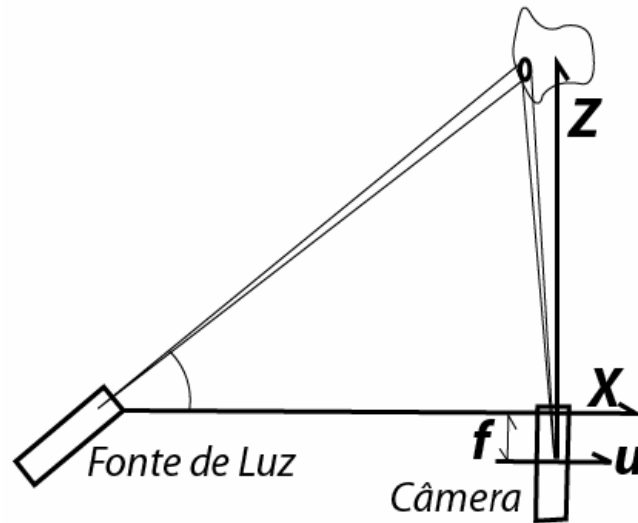


Figura 4.8: Típica geometria de luz estruturada (Fonte: FRANÇA, 2002).

As coordenadas reais dos pontos na luz estruturada são dadas por:

$$X = \frac{bu}{f \cot \Theta - u} \quad Y = \frac{bv}{f \cot \Theta - u} \quad Z = \frac{bf}{f \cot \Theta - u} C \quad (44)$$

As quantidades f , b , Θ são conhecidas pela configuração do equipamento, e a localização dos *pixels* u, v são encontradas por uma simples detecção de pico percorrendo a linha de varredura da imagem.

4.6.4. Sistemas de *Motion Capture*

Sistemas de *motion capture* são, por definição, a captura de movimentos reais por sistemas de computadores. O processo de captura consiste em marcações posicionadas pelo corpo (fonte emissora) e a captação destes sinais por sensores.

Os sistemas de *motion capture* podem ser classificados de acordo com a sua tecnologia, pois esta define o processo usado pela aquisição volumétrica.

Cabe ressaltar que a precisão de diferentes tipos de sistemas de *motion capture* varia numa escala de baixa a alta precisão. Para este trabalho será considerado um sistema de baixa precisão aquele que apresentar ruídos ou perda de dados capturados em 50% dos quadros capturados. Já, para um sistema de média precisão, considera-se a presença de ruídos ou perda

de dados capturados em um intervalo de 10% a 49%. Os sistemas de alta precisão são considerados nos casos em que houver uma perda menor que 10%.

4.6.4.1. Sistemas Acústicos

Emissores sonoros são espalhados pelo corpo, ou pela região a ter o movimento captado, e três receptores sensíveis são posicionados no local de captura.

Uma triangulação entre as distâncias dos emissores em relação aos receptores é a forma como o sistema capta a informação volumétrica. Para isso, o sistema utiliza como dados, o tempo decorrido entre a emissão do ruído pelo transmissor e o seu recebimento pelo receptor e a velocidade do som no ambiente.

As desvantagens deste sistema são: os ruídos presentes no ambiente gravado, que pode atrapalhar a captura dos dados; e a utilização de poucos transmissores, que podem ser usados simultaneamente, não fornecendo uma descrição precisa do movimento captado, devido à presença de ruídos. Uma vantagem deste sistema é a não ocorrência do problema da oclusão.

4.6.4.2. Sistemas Mecânicos

Sistema com equipamento de medidas absolutas sem nenhuma interferência do ambiente a ser captado, não necessitando de um processo longo de calibragem. A captura é realizada por meio de potenciômetros ou *sliders*, que são posicionados nos locais possíveis (preferencialmente articulações) do corpo.

Conforme a tabela 1, o sistema mecânico permite uma alta taxa de amostragem (tempo real), porém o seu equipamento não dá muita liberdade de movimento ao ator.

A tabela 1 mostra as principais características dos sistemas mecânicos.

Tabela 1: Características dos sistemas mecânicos (Fonte: SILVA, 1997).

	Sistemas Mecânicos
Resolução espacial	Independente do alcance, ~ 0.5 - 1 graus
Acurácia espacial	≤ 5 graus
Precisão	de média a elevada
Taxa de amostragem	> 120 quadros por segundo
Campo de trabalho	Ilimitado (depende apenas do comprimento dos cabos de transmissão)
Exemplos de sistemas	Animatton, Digital Monkey, Power Glove

Softwares compatíveis	Alias Wavefront, Jack, 3D Studio Max, SoftImage
Custo	US\$ 1,000 a US\$ 35,000

4.6.4.3. Sistemas Ópticos

Os emissores deste sistema, em geral, são refletores (emissores LED) que são posicionados em qualquer parte do corpo para a captação do movimento. Diversas câmeras (em torno de doze) são posicionadas no ambiente para a captação do movimento e a geração das coordenadas 2D. Após este procedimento, *softwares* irão realizar soluções geométricas (de acordo com a disposição das câmeras de captura) para fornecer as coordenadas 3D.

As taxas de amostragem destes sistemas ópticos podem atingir até 200 quadros por segundo, porém seu custo financeiro é muito elevado (US\$ 20,000 a US\$ 150,000). Outra característica do sistema óptico, é a liberdade do equipamento, ou seja, o usuário que simulará movimentos reais não ficará coberto por fios e outros equipamentos que prejudiquem seu movimento (SILVA, 1997).

Algumas desvantagens podem ser destacadas como a oclusão, que pode ser solucionado com o uso de várias câmeras. Além disso, ruídos e erros de precisão podem ocorrer se o sistema não estiver bem configurado.

A tabela 2 mostra as principais características dos sistemas ópticos.

Tabela 2: Características dos sistemas ópticos (Fonte: SILVA, 1997).

	Sistemas Ópticos
Resolução espacial	~ 0.0015 a 0.2 % do campo de visão (CV)
Acurácia espacial	~ 0.004 a 0.5 % do CV
Precisão	~ 0.0055 a 0.02 % do CV
Taxa de amostragem	> 200 quadros por segundo
Campo de trabalho	raio de 1 a 4 m
Exemplos de sistemas	ExpertVision HiRES 3-D System, Multitrax Motion Capture System, OPTOTRAK
Softwares compatíveis	ZoeTrax, The Creative Motion Editor, PowerAnimator V7

Custo	US\$ 20,000 a US\$ 150,000
--------------	----------------------------

4.6.4.4. Sistemas Magnéticos

Os equipamentos deste sistema utilizam receptores, que são posicionados no corpo do usuário, e uma antena que transmite um sinal de pulso para estes receptores, para medir suas posições 3D.

Os sistemas magnéticos são de baixo custo financeiros (US\$ 5,000 a US\$ 70,000), possui baixo custo computacional, alta precisão pois não há o problema da oclusão e a alta taxa de amostragem (em torno de 100 quadros por segundo) (SILVA, 1997).

Sua maior desvantagem são os cabos que conectam os receptores às antenas, dificultando o ator de executar movimentos e a interferência causada por objetos de metal próximos ao local de captura.

A tabela 3 mostra as principais características dos sistemas magnéticos.

Tabela 3: Características dos sistemas magnéticos (Fonte: SILVA, 1997).

	Sistemas Magnéticos
Resolução espacial	~ 0.005 a 8 mm; ~ 0.025 a 0.1 graus
Acurácia espacial	~ 0.8 a 5 mm; 0.1 a 3 graus
Precisão	Elevada
Taxa de amostragem	~ 100 quadros por segundo
Campo de trabalho	raio de 1 a 2 m
Exemplos de sistemas	A Flock of Birds, ULTRATRAK, FASTRAK, INSIDETRAK
Softwares compatíveis	Alias, 4Dvision, XSI, Kinemation 3.0, 3D Studio, PowerAnimator V7, Photo4D
Custo	US\$ 5,000 a US\$ 70,000

Capítulo 5: Definição de Métodos para um Sistema de Animação Facial baseado em Estereoscopia

Este capítulo abrange a metodologia envolvida na pesquisa e os materiais utilizados. Vale ressaltar que a construção dos métodos para um sistema de animação facial, usando estereoscopia, foi focada no uso de materiais de baixo custo financeiro.

5.1. Sistema de Animação Proposto

O sistema de animação proposto baseia-se no paradigma de performance humana real por um sistema *motion capture* óptico (seção 3.7.3) semelhante ao da universidade de Ohio (seção 3.10.4).

A figura 5.1 demonstra as etapas propostas para a conclusão de uma animação facial 3D, baseada na demarcação de pontos de controle faciais em uma face humana real (paradigma de performance humana real), que é previamente modelada em 3D utilizando a técnica de *box modeling*.

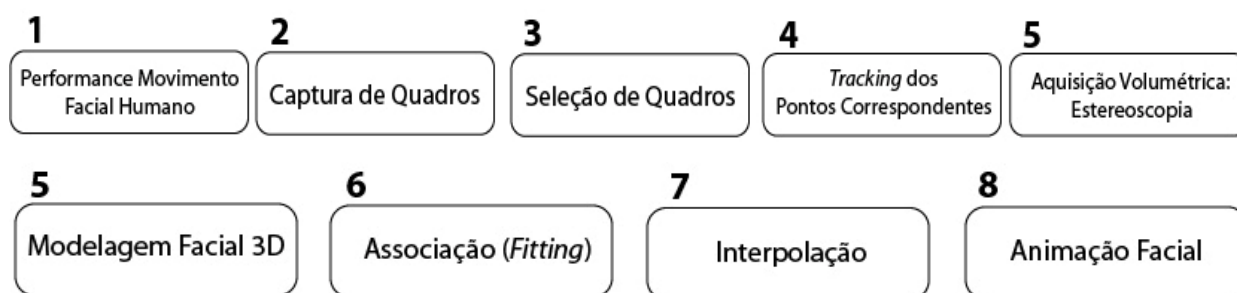


Figura 5.1: Métodos propostos para o sistema de animação facial.

Para a captação dos movimentos faciais do modelo real humano (etapa 1), faz-se necessário a demarcação de pontos de controles na face desta pessoa. Neste trabalho considerou-se 66 pontos distribuídos na face, os quais abrangem todos os principais músculos do movimento facial.

Após a demarcação dos pontos de controle, dá-se o início da etapa 2, que é a de captura de quadros a partir do sistema de duas câmeras sincronizadas. A saída desta gravação gera imagens ou quadros correspondentes (da câmera da esquerda e da direita) de expressões faciais feitas pelo modelo humano.

A seleção de quadros (etapa 3), que ocorre logo após a captura, é considerada uma etapa de extrema importância para a animação final, pois é nela que se escolhe as poses-chave. A escolha de poses certas resulta em uma animação com poucos quadros apenas. Contudo, para animações mais complexas, é necessário a escolha de um maior número de quadros, suficientes para garantir que a animação gerada corresponda ao movimento real.

A escolha dos quadros é realizada visualmente, através da habilidade técnica do animador, caracterizando um processo artístico.

O *tracking* dos pontos correspondentes (etapa 4) é o processo de localização destes pontos durante a performance humana.

Tendo as localizações dos pontos correspondentes (X e Y) em todas os quadros correspondentes, faz-se necessário a aquisição volumétrica desses, através do método de estereoscopia. O algoritmo estereoscópico recebe dados de entrada como a distância focal das câmeras, a distância física entre elas e a resolução das imagens geradas, para seus cálculos. Sua saída produz um arquivo contendo as localizações (X , Y e Z) dos pontos demarcados da face humana, de cada par de imagens correspondentes.

A etapa 5, consiste na modelagem de uma face 3D, utilizando a técnica de *box modeling*, a partir da referência do rosto proposto (Figura 5.14). A face construída é constituída de 734 polígonos (triângulos) em sua modelagem e conta com 66 pontos, ou vértices, de controle para a animação facial. Vale ressaltar que estes vértices de controle são associados às demarcações feitas no rosto humano (etapa 6), ou seja, todo o movimento realizado por um determinado ponto facial deve-se ser aplicado ao seu correspondente vértice da face 3D.

A interpolação, técnica da etapa 7, é utilizada para a transição entre os quadros ou poses-chave geradas pelas coordenadas do algoritmo de estereoscopia. É nesta etapa em que o animador consegue controlar o tempo de animação a ser usado.

Para a animação final (etapa 8) propõem-se a utilização da técnica de interpolação entre os quadros-chave, gerados pelas coordenadas do algoritmo de estereoscopia (etapa 7). Esta técnica proporciona o uso de modificadores *morphings* para expressões faciais e para sincronismo labiais.

5.2. Materiais

Para a gravação de movimentos faciais reais, foram utilizadas duas câmeras do tipo segurança, monocromáticas, com duas saídas RCA de vídeo e som e com uma entrada de força para cada uma. As lentes das câmeras possuem uma distância focal de 3 mm e suas saídas de vídeo possuem uma resolução de 320 x 240 *pixels*.

De acordo com a figura 5.2, o cabo denominado A, de cor vermelha, é caracterizado pela entrada de energia, o cabo B, de cor branca, é caracterizado pela saída do áudio e por fim, o cabo C, de cor amarela, é caracterizado pela saída de vídeo.



Figura 5.2: Duas câmeras de segurança monocromáticas.

A captação e digitalização da seqüência de quadros (*frames*) foram realizadas por meio de duas placas gráficas com sinal de entrada de vídeo RCA. As placas utilizadas foram:

1) Placa aceleradora *MSI FX5600XT VTDR 128 MB*: De acordo com a figura 5.3, esta possui uma interface AGP 8X e três conectores de entradas e saídas. Os conectores denominados de A e C, da figura 5.3, são caracterizados pela saída do sinal de vídeo do tipo VGA e o conector denominado B é caracterizado pela entrada e saída do sinal de vídeo do tipo DVI-I.

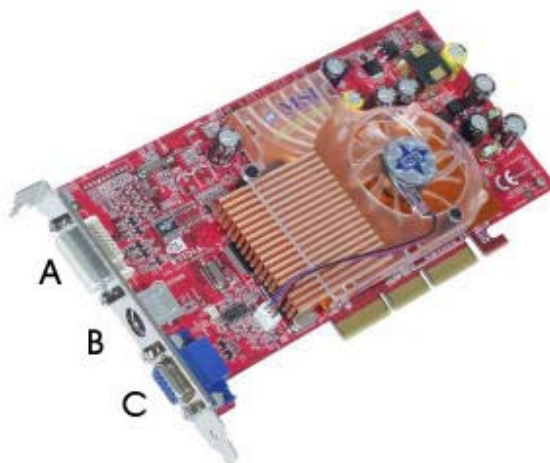


Figura 5.3: Placa gráfica para captura de vídeo *FX5600*.

Contudo, para que a entrada de vídeo do conector B do tipo DVI-I, da figura 5.3, receba o sinal proveniente da saída do sinal de vídeo, do tipo RCA, de uma câmera (cabo C da figura 5.2), fez-se necessário o uso de um adaptador para a conversão de sinais do tipo DVI-I para RCA.

Como mostra a figura 5.4, o adaptador possui 4 conectores: A, B, que são caracterizados pela entrada de sinal de vídeo por diferentes interfaces; e C, D, que são caracterizados pela saída de sinal de vídeo por diferentes interfaces. Ambos os conectores convergem para uma saída DVI-I. A entrada A foi utilizada para a realização da captura do sinal de vídeo por meio da interface RCA.



Figura 5.4: Adaptador que realiza a conversão de interfaces DVI-I para RCA.

2) Placa de captura *PixelView BT878*: Esta placa é caracterizada pelas entradas de sinais de vídeos do tipo RCA, pela entrada de TV do tipo coaxial, pela entrada de sinal de rádio FM e pelas entradas e saídas de áudio (vide figura 5.5).

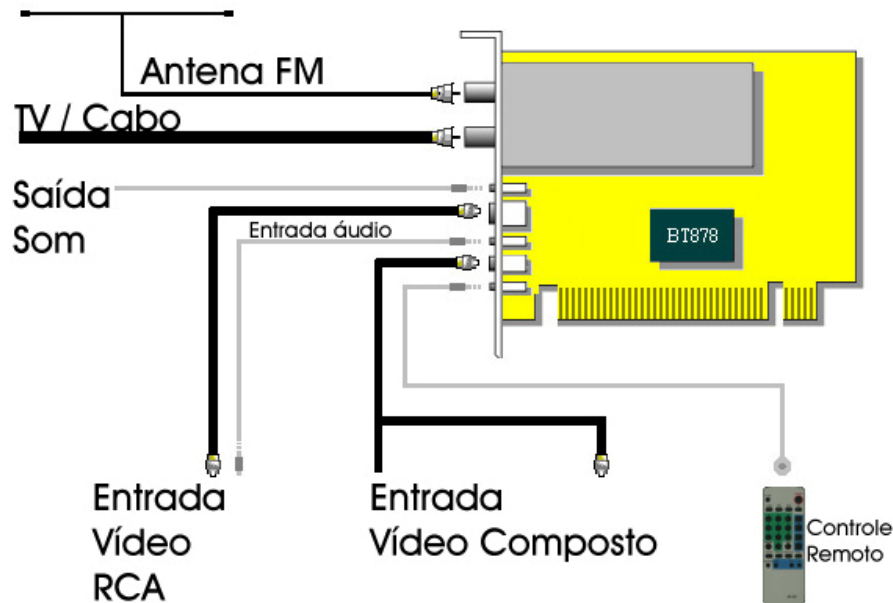


Figura 5.5: Esquema da placa *PixelView* para captura de vídeo.

O computador (PC) utilizado foi um processador AtlonXP 2200+, com 1.8GHz de *clock*, 512MB de memória RAM e com duas placas de vídeo citadas nas figuras 5.3 e 5.5.

5.3. Desenvolvimento do Sistema

O desenvolvimento dos métodos do sistema deu-se em duas principais etapas:

5.3.1 Hardware

O método estereoscópico avalia dois quadros iguais, tirados no mesmo intervalo de tempo, para a realização de seus cálculos, porém, sem uma sincronização, quadros semelhantes podem ser captados em tempos diferentes. Para evitar este problema, fez-se necessário uma sincronização entre as câmeras, ou seja, para que ambas gerassem vídeo ao mesmo tempo.

De acordo com a figura 5.6, percebe-se a sincronia entre as duas câmeras, pois ambas geram quadros ao mesmo tempo.

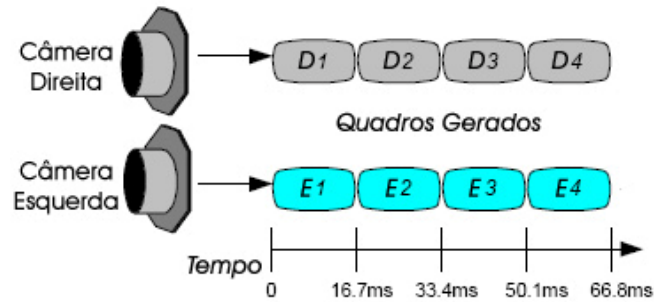


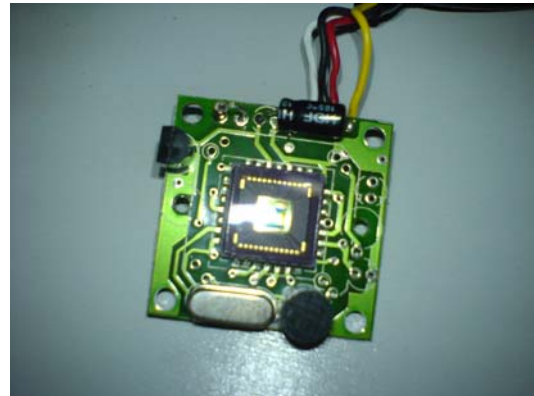
Figura 5.6: Esquema de câmeras sincronizadas (Fonte: VICKERS, 2002).

Optou-se pelo procedimento de sincronização por *hardware* devido aos fatores de desempenho, pois se esta sincronia fosse realizada via *software*, a execução desta etapa não teria o mesmo resultado.

Portanto, o método de sincronização das câmeras tem início na abertura de seus compartimentos (figura 5.7 item (a)) e na localização do núcleo do *hardware* das mesmas (figura 5.7 item (b)).



(a)



(b)

Figura 5.7: Câmera sem o compartimento de proteção (a) e com seu *hardware* (b).

De acordo com o item (b) da figura 5.7, é possível notar que o núcleo do *hardware* possui um formato quadrangular com 7 pequenos conectores em cada um de seus quatros lados.

Portanto, conclui-se que um destes 28 conectores possui a característica de saída do sinal de sincronia (*FIELD SYNC OUT - FSO*) e um outro a característica de entrada do sinal de sincronia (*FIELD SYNC IN - FSI*). Vale ressaltar que para o funcionamento da sincronia, é obrigatória a presença destes dois conectores. Contudo, como não há marcações a esse respeito

no *hardware* das câmeras e nem um manual técnico, fez-se uma pesquisa em um artigo de eletrônica (VICKERS, 2002), para descobrir qual desses conectores seria o de FSO e de FSI.

Os campos de saída e entrada de sincronia (FSO e FSI) possuem a seguinte característica de sinal pelo tempo (Figura 5.8):

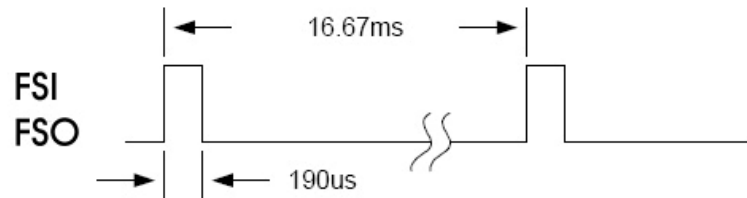


Figura 5.8: Tempo dos sinais de FSI e FSO (Fonte: VICKERS, 2002).

Outra propriedade destacada no FSO, é a relação entre este e o sinal vídeo, ou seja, o FSO é um pulso de sincronia vertical que ocorre durante o período de atualização vertical do sinal de vídeo (Figura 5.9).

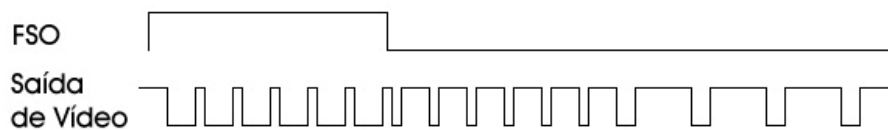


Figura 5.9: Relação entre os sinais de FSO e a saída de vídeo (Fonte: VICKERS, 2002).

Então, um osciloscópio foi utilizado para o descobrimento dos conectores FSO e FSI das respectivas câmeras tendo como informações à priori, as características já citadas dos sinais FSI, FSO e saída de vídeo. Após a localização dos conectores, uma ligação entre eles foi executada por meio da solda de um fio condutor. Conforme a figura 5.10, a identificação e a ligação dos conectores é esquematizada:

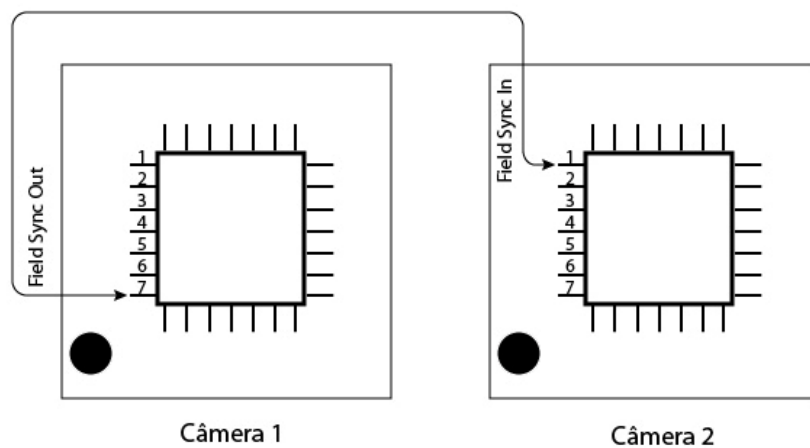


Figura 5.10: Identificação do FSO e FSI e a ligação entre eles.

Como o canal de áudio da câmera não seria utilizado pela pesquisa, o fio deste foi interligado entre os conectores FSI de uma câmera ao FSO da outra câmera.

A figura 5.11 mostra a solda dos fios nos respectivos conectores FSO e FSI e a utilização do fio de áudio para uma maior extensão.

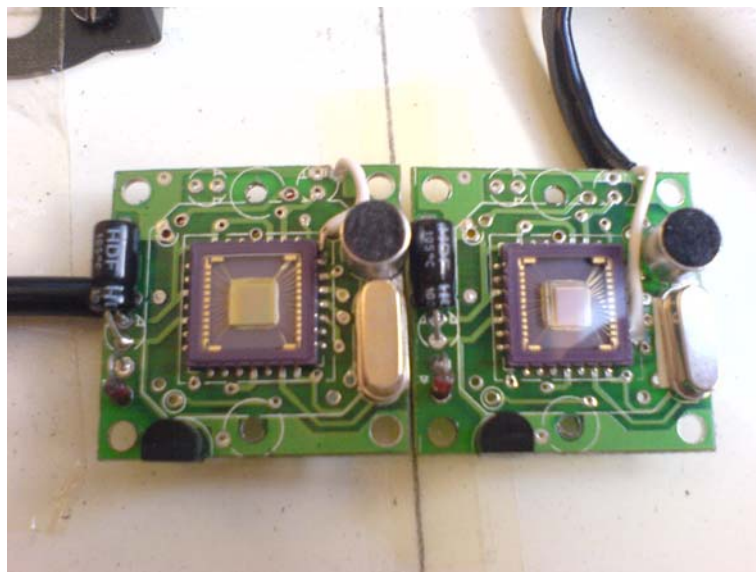


Figura 5.11: Câmeras sincronizadas.

Após o término destes procedimentos, a montagem das câmeras foi realizada, assim como a determinação de uma distância física de 40mm entre suas respectivas lentes ao longo de um plano. Esta distância foi determinada levando-se em consideração a resolução da

câmera e a capacidade de enquadramento de uma face inteira em ambas as câmeras, evitando assim o problema da oclusão.

Para a sustentação e disposição das 2 câmeras, foi construído um suporte de madeira facilitando assim o manuseio das mesmas (Figura 5.12).

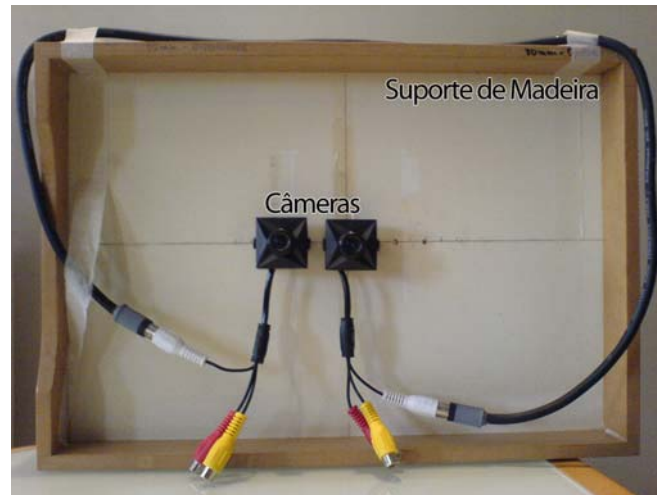


Figura 5.12: Disposição das duas câmeras com uma distância de 40mm entre seus centros ópticos.

Finalmente, o sistema de câmeras juntamente com o computador e placas de aquisição caracterizam-se no esquema da figura 5.13.

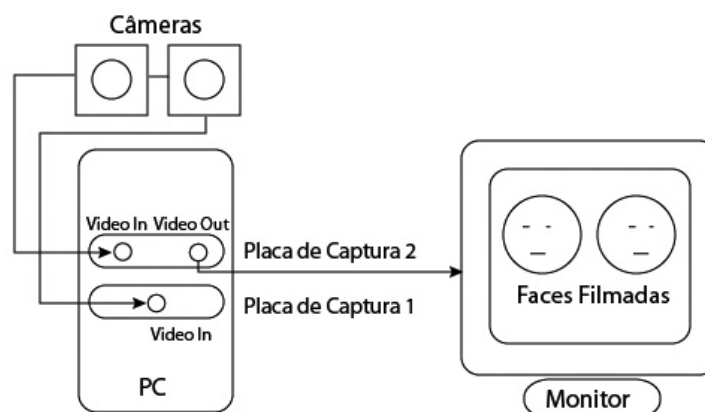


Figura 5.13: Esquema do sistema de animação facial montado.

5.3.2 Software

Os métodos da etapa de *software* são desenvolvidos nas seguintes sub-etapas:

5.3.2.1. Performance do Movimento Facial Humano

Para a captação dos movimentos faciais a partir de uma face real, diversas marcações foram feitas no rosto de uma pessoa para auxiliar a determinação dos pontos correspondentes entre as imagens da esquerda e da direita. Para a distribuição destes, foi levado em consideração, a importância da ação muscular nestas regiões.



Figura 5.14: Pontos que serão capturados pela estereoscopia na face humana.

A realização e captura dos movimentos faciais do modelo real foram realizadas a uma distância fixa de 40cm (sem mobilidade) da face para com o sistema de câmeras, tendo como relevâncias, o enquadramento total das faces em ambos canais esquerdo e direito, e o problema da oclusão. qual a distância??

No total, foram escolhidos 66 pontos de controle para a determinação da animação facial, os quais correspondem aos principais músculos faciais.

5.3.2.2. Seqüência de Quadros (*Frames*) Capturados

A realização da captura dos quadros de uma animação facial, proveniente de uma face humana, ocorre após a demarcação dos pontos correspondentes na mesma.

Foi utilizado o *software* de edição de vídeo, *Virtual Dub 1.6.10*, para o acesso às duas placas de aquisição e a gravação de um arquivo contendo uma seqüência de quadros captados (arquivo do tipo *avi*, padrão *windows*, contendo 30 quadros por segundo).

Com o uso deste *software*, a gravação gerou dois arquivos de vídeo provenientes das câmeras da esquerda e da direita, contendo 30 quadros por segundo. Tendo a sincronia por *hardware* entre as câmeras e os vídeos gerados em forma de quadros, selecionam-se então os principais quadros da animação (quadros-chave) para o cálculo da estereoscopia e de sua interpolação.

A seleção de quadros-chave é importante para evitar o cálculo estereoscópico de quadros intermediários (quadros entre os quadros-chave), que seria facilmente alcançado pelo processo de interpolação.

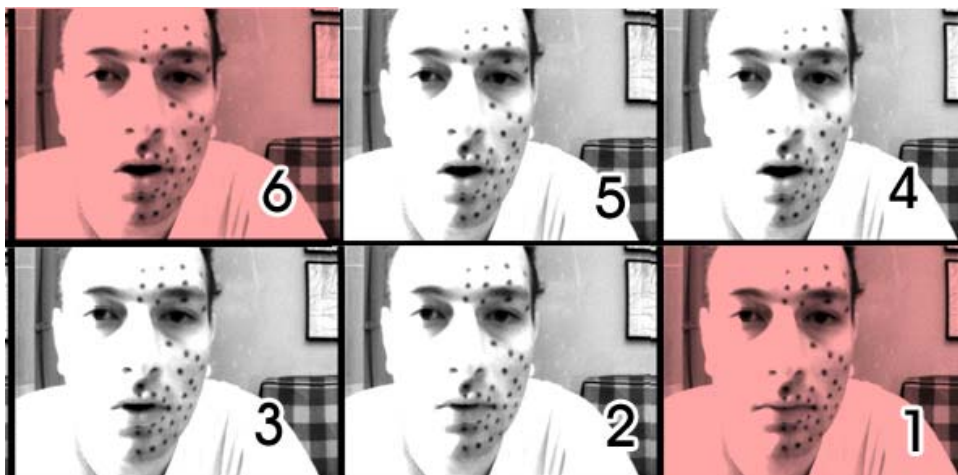


Figura 5.15: Seleção dos quadros 1 e 6 apenas para o cálculo estereoscópico.

5.3.2.3. Estereoscopia

O cálculo estereoscópico para determinar os valores X, Y e Z dos pontos correspondentes é explicado pelo esquema da figura 5.16, na qual contém oito passos. É importante ressaltar que a saída gerada pelo algoritmo é usada na etapa de interpolação dos pontos correspondentes. O algoritmo foi desenvolvido em *MatLab*.

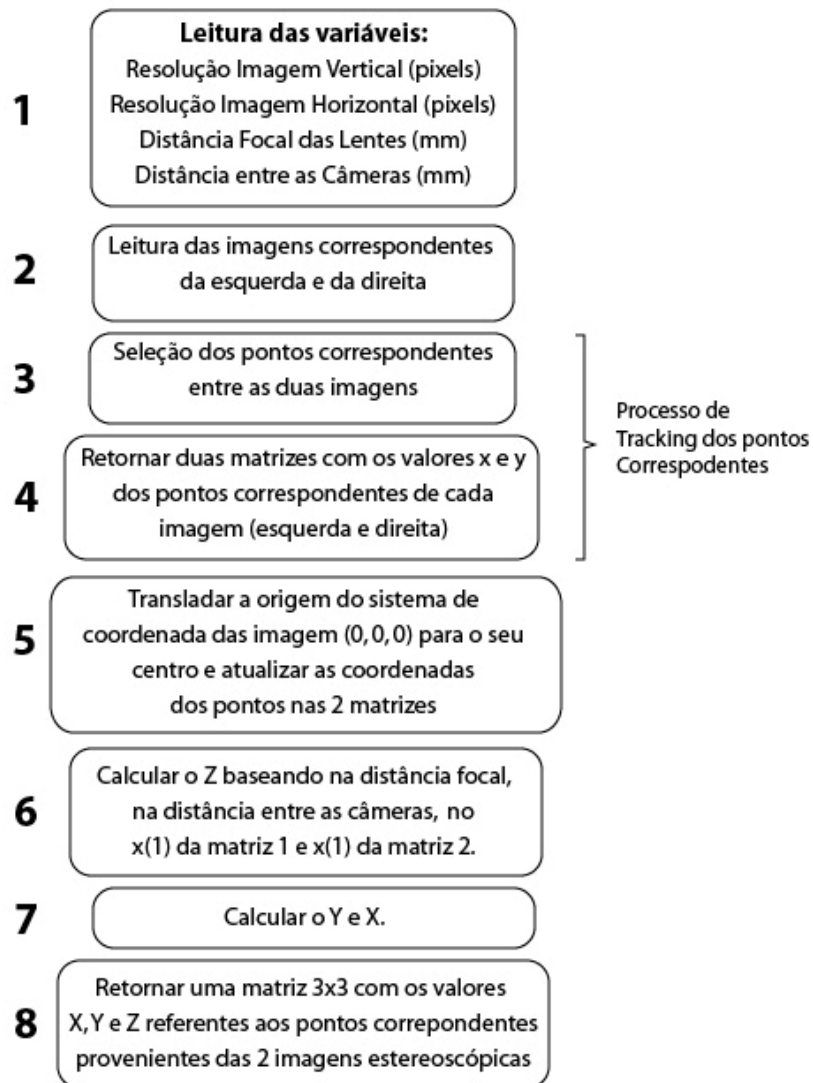


Figura 5.16: Esquema da estereoscopia em alto nível.

5.3.2.4. Modelo Facial 3D

O modelo facial 3D foi modelado com base na técnica de *box modeling*, usando como referência, uma foto real. Os pontos faciais são usados para referência também no processo de modelagem e disposição dos vértices na face. Conforme a figura 5.17, percebe-se o uso da referência e a modelagem pronta.

Os vértices da face modelada são associados com os valores resultados do algoritmo de estereoscopia para a realização da interpolação.

A face apresenta 734 polígonos (triângulos) em sua modelagem e conta com 66 pontos, ou vértices, de controle para a animação facial. Utilizou-se a técnica de subdivisão de superfícies para obter a aparência de uma superfície suave da face, resultando num modelo

com 11696 polígonos. Esta técnica permite a geração automática de novos polígonos em cima da superfície aplicada, além da geração de uma tensão nas quinas dos mesmos, deixando um formato mais arredondado.



Figura 5.17: Fotos de referência e o resultado modelado.

5.3.2.5. Associação (*Fitting*)

As associações dos pontos resultantes pela estereoscopia, na face 3D virtual, são realizadas para que os valores X, Y e Z gerados sejam usados na etapa da interpolação. Como se observa na figura 5.18, no item (a), a face humana foi demarcada com pontos pretos que se movimentam durante a performance humana. Contudo, para que a face 3D virtual consiga realizar a simulação destes movimentos captados pela estereoscopia, faz-se necessário a identificação ou associação dos pontos demarcados na face real (item (a)), para com seus correspondentes vértices na face 3D (item (b)).

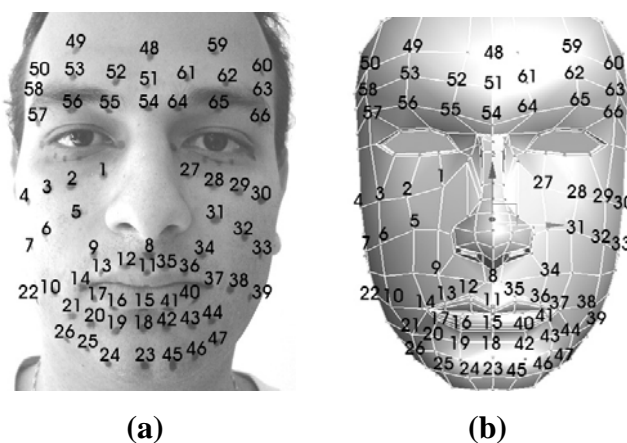


Figura 5.18: (a) Pontos identificados na face real; (b) Mesmos pontos associados nos vértices da face 3D. Total de 66 pontos de controle.

Portanto, a identificação dos pontos faz com que o sistema consiga entender a ordem de escolha dos pontos correspondentes e, assim, prepará-los para uma interpolação na face 3D virtual. O processo de estereoscopia gera todas as posições X, Y e Z de cada ponto em cada quadro captado durante uma performance humana.

5.3.2.6. Interpolação

Tendo a posição base, ou seja, a pose inicial no formato (X, Y, Z) e as poses dos quadros-chave (selecionados em uma etapa anterior e respeitando o mesmo formato resultante da estereoscopia), faz-se necessário uma interpolação entre esses pontos para que se obtenha um resultado de animação facial.

A interpolação permite a criação de curvas de aceleração e desaceleração, movimentos lineares, quadros intermediários, parametrização por *morphing* e movimento contrário. Esta etapa foi realizada dentro do *software 3D Studio Max*.

5.3.2.7. Animação Facial

Após a conclusão das etapas anteriores, torna-se possível a implementação final da animação facial. Esta comporta diversas saídas como vídeo, imagens entre outras.

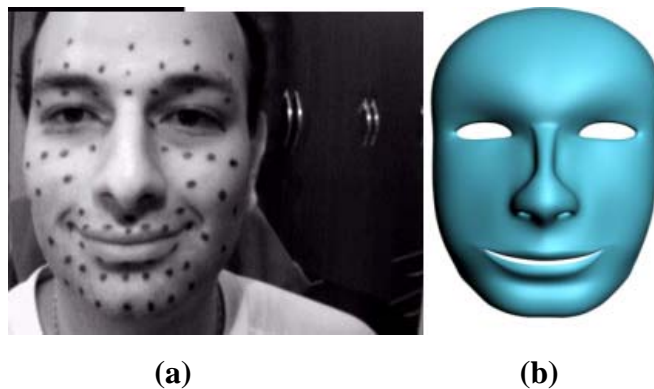


Figura 5.19: Animação Facial concluída baseada no vídeo capturado. (a) uma das imagens faciais captadas e (b) a imagem correspondente do modelo.

Capítulo 6: Resultados e Discussão

Neste capítulo serão apresentados os resultados e a discussão do trabalho no que diz respeito às expressões faciais e sincronismo labial aplicados em uma face 3D.

6.1. Expressões

Neste estudo, são utilizadas as expressões faciais de surpreso, sorriso e emburrado.

6.1.1. Surpreso

A figura 6.1 mostra a captação do movimento facial humano pela câmera da esquerda (13 quadros), em ordem cronológica, da esquerda para a direita, e de cima para baixo, da expressão de surpreso.



Figura 6.1: Imagens faciais de expressão de surpreso pela câmera esquerda.

A captação foi realizada levando-se em consideração o alinhamento entre as câmeras, evitando erros de cálculos na etapa de estereoscopia. Uma possível correção do problema de alinhamento é o deslocamento da seqüência das imagens ao longo de um determinado eixo.

A figura 6.2 mostra a captação do movimento facial humano pela câmera da direita (13 quadros), na mesma ordem da figura 6.1.



Figura 6.2: Imagens faciais de expressão de surpresa pela câmera direita.

Observa-se que a animação do movimento facial foi realizada com 13 quadros, porém, nem todos necessitam estar presentes no cálculo de estereoscopia, uma vez que a interpolação é uma solução mais eficaz para este problema. Contudo, como é observado no caso da expressão surpresa, o vídeo indica que o movimento de sobrancelhas ocorreu antes do movimento de abertura de boca. Este fenômeno não pode ser compreendido na interpolação, caso os quadros-chave escolhidos sejam os extremos (o primeiro e o último). Porém, se mais quadros-chave forem escolhidos indicando este movimento, este problema pode ser solucionado. Vale ressaltar que a interpolação aproxima o movimento realizado entre os quadros-chave. Portanto, para uma maior proximidade do movimento real, deve-se adotar todos os quadros-chave captados na filmagem.

A figura 6.3 mostra a seleção de quadros-chave da câmera esquerda (2 quadros extremos):



Figura 6.3: Quadros da câmera esquerda que foram selecionados para a interpolação. (a) quadro inicial e (b) quadro final.

Seleção de quadros-chave da direita (2 quadros extremos):



Figura 6.4: Quadros da câmera direita que foram selecionados para a interpolação. (a) quadro inicial e (b) quadro final.

Após a seleção dos quadros-chave, inicia-se a etapa de estereoscopia. Esta abrange a seleção dos pontos correspondentes previamente marcados na face humana entre os pares de quadros equivalentes. Assim, o primeiro quadro-chave proveniente da câmera esquerda e da direita devem ser as entradas para o cálculo da estereoscopia.

É importante ressaltar que a escolha dos pontos correspondentes entre os quadros foi realizada de forma manual, ou seja, o usuário seleciona os pontos da imagem da esquerda e os associa com os mesmos na imagem da direita um a um.

De acordo com a figura 5.17, a ordem de escolha dos pontos leva em consideração o processo de *fitting*, ou seja, a indexação dos pontos na face.

6.1.1.1 Estereoscopia dos pontos correspondentes de cada par de quadros equivalentes (Processo de *Matching*).

Para o processo de cálculo estereoscópico, escolhe-se primeiramente o ponto 1 (Figura 5.17) da imagem da esquerda e o seu correspondente da direita. Faz-se o mesmo com os demais, até os últimos pares de pontos. A figura 6.5 mostra um par de quadros, imagem da direita (a) e esquerda (b), destacando através de subimagens, direita (c) e esquerda (d), uma instância de obtenção de pontos correspondentes baseados na indexação pré-estabelecida (Figura 5.17).

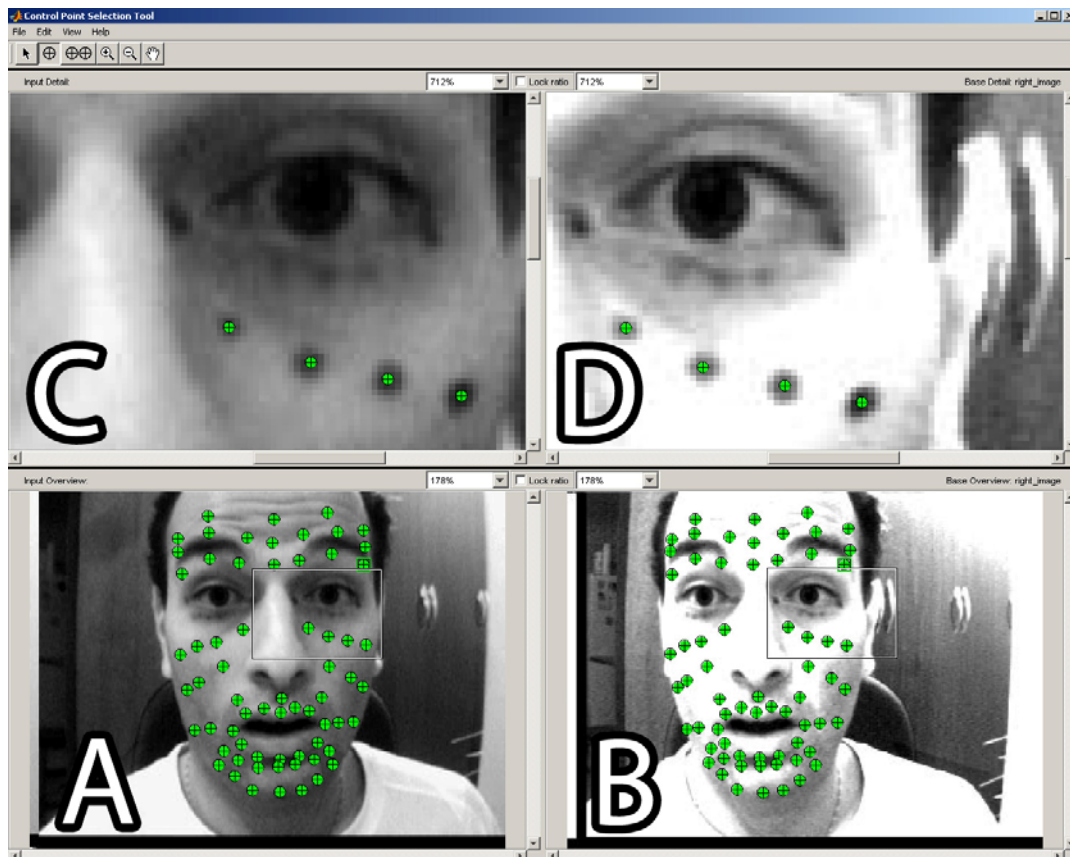


Figura 6.5: Encontro de pontos semelhantes em imagens correspondentes da esquerda e da direita no par de quadros.

Os pontos resultantes do algoritmo de estereoscopia dos quadros extremos, previamente selecionados foram obtidos conforme o Apêndice.

Tendo os pontos resultantes da estereoscopia, na ordem pré-estabelecida pelo *fitting*, deve-se associá-los em uma face 3D e realizar as interpolações entre os quadros extremos, atingindo o resultado observado na figura 6.6.

A figura 6.6 mostra a associação no modelo 3D e a interpolação:

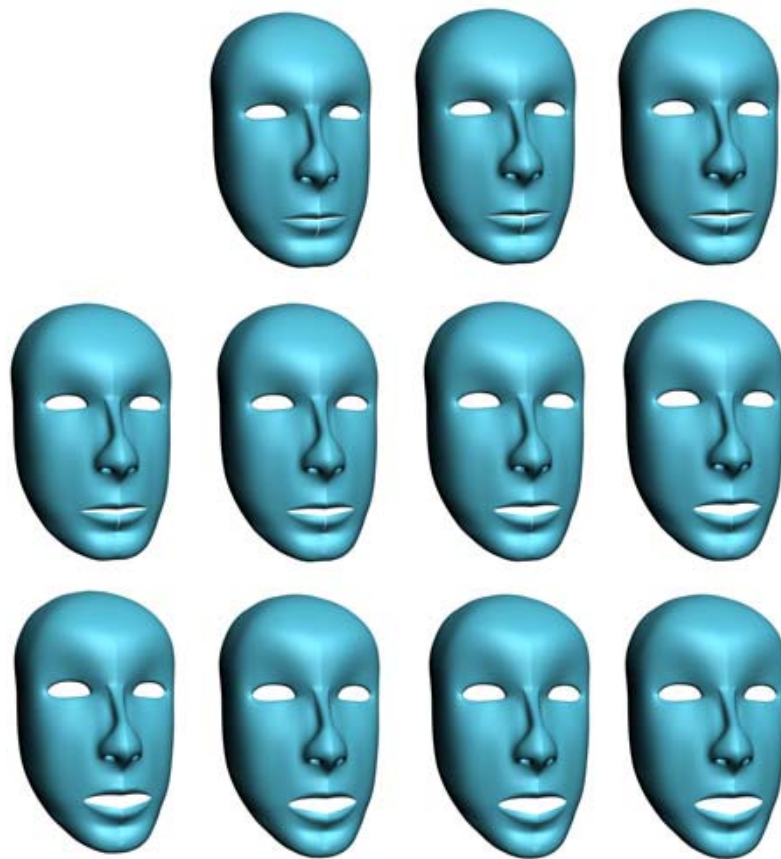


Figura 6.6: Associação no modelo e interpolação entre as poses-chave.

Para as demais expressões faciais e sincronismo labial, são apenas demonstrados as principais etapas do processo, pois utilizou-se a mesma metodologia de desenvolvimento da expressão de surpresa já descrita.

6.1.2. Sorriso

A figura 6.7 mostra a captação do movimento facial humano pela câmera da esquerda (18 quadros), em ordem cronológica, da esquerda para a direita, e de cima para baixo, da expressão de sorriso.

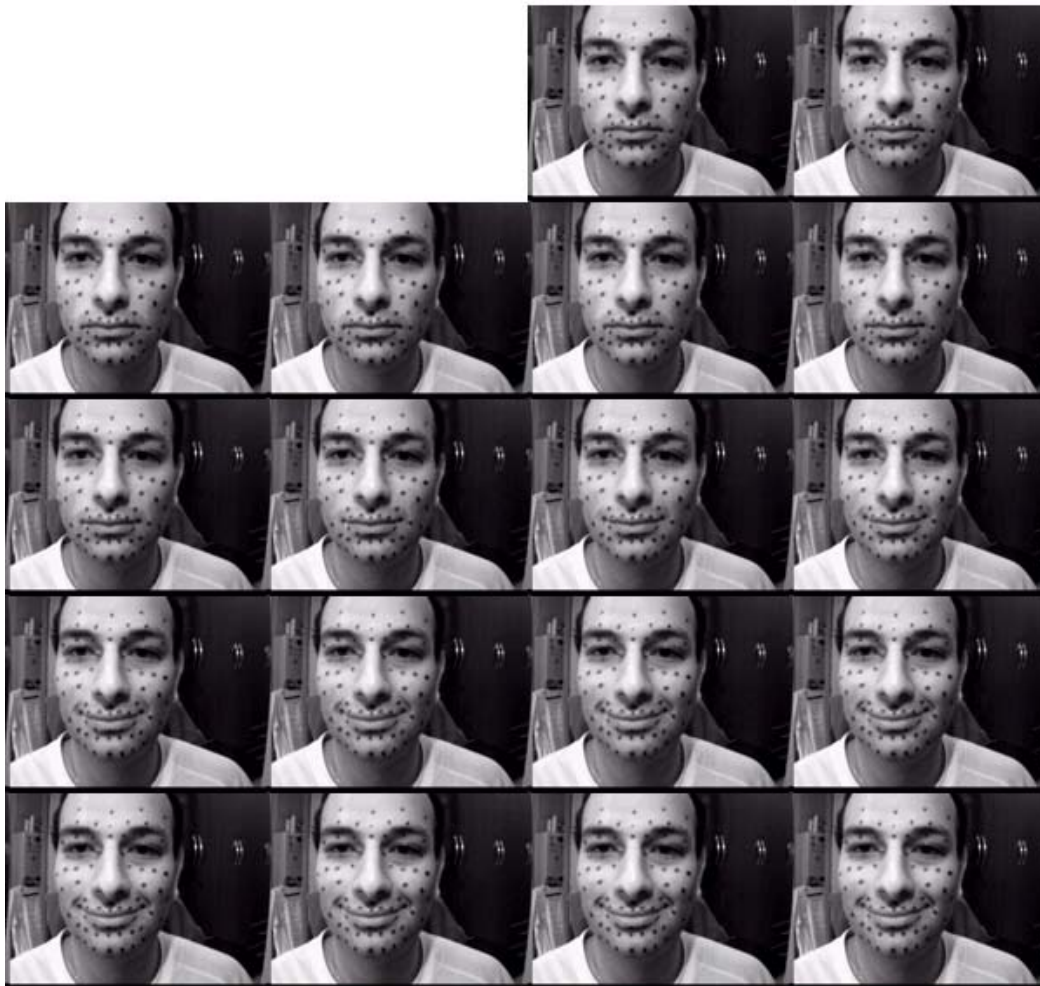


Figura 6.7: Imagens faciais de expressão de sorriso obtidas pela câmera esquerda.

A figura 6.8 mostra a captação do movimento facial humano pela câmera da direita (18 quadros), na mesma ordem da figura 6.7, da expressão de sorriso.



Figura 6.8: Imagens faciais de expressão de sorriso obtidas pela câmera direita.

A figura 6.9 mostra a seleção de quadros-chave da esquerda (3 quadros: 2 extremos e 1 intermediário), das imagens de sorriso, usados para interpolação e o cálculo estereoscópico.

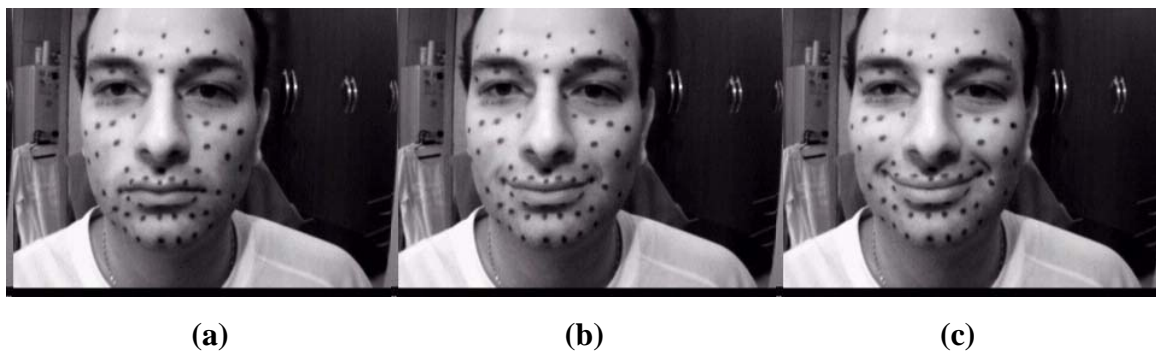


Figura 6.9: Quadros da câmera esquerda que foram selecionados para a interpolação. (a) quadro inicial, (b) quadro intermediário e (c) quadro final.

A figura 6.10 mostra a seleção de quadro-chave da direita (3 quadros: 2 extremos e 1 intermediário), usados para a interpolação e o cálculo estereoscópico.

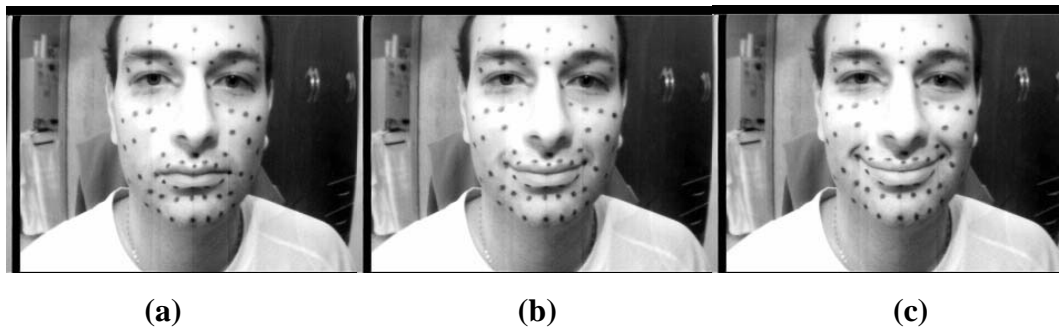


Figura 6.10: Quadros da câmera direita que foram selecionados para a interpolação. (a) quadro inicial, (b) quadro intermediário e (c) quadro final.

6.1.2.1. Estereoscopia dos Pontos Correspondentes de cada Par de Quadros Equivalentes (Processo de *Matching*)

O processo de estereoscopia dos pontos correspondentes de cada par de quadros equivalentes foi realizado da mesma forma que a expressão de surpresa (figura 6.5), como mostra a figura 6.11.

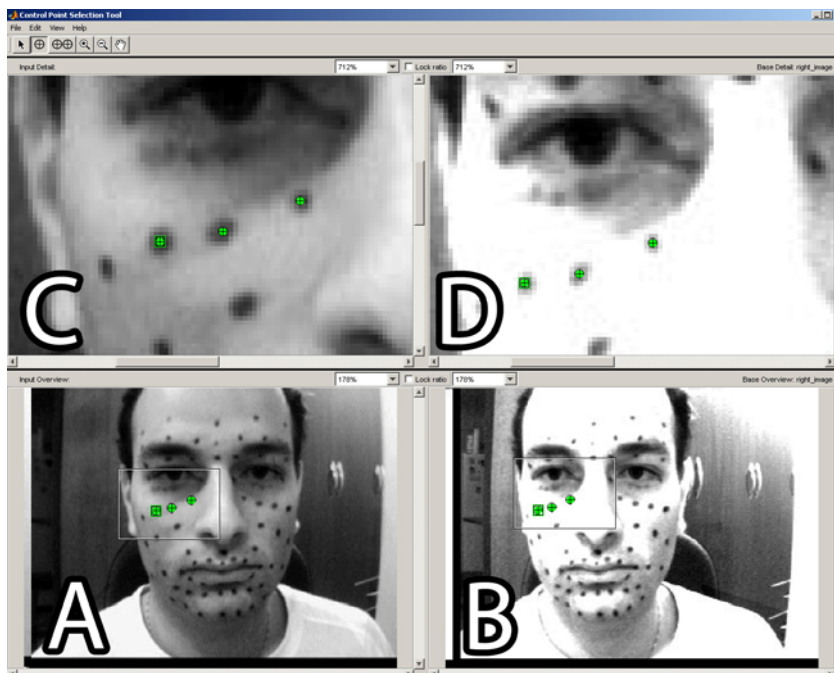


Figura 6.11: Encontro de pontos semelhantes, em imagens correspondentes da esquerda e da direita, nos 3 pares de quadros.

O resultado da associação no modelo 3D e a interpolação, para a expressão de sorriso é mostrado na figura 6.12.

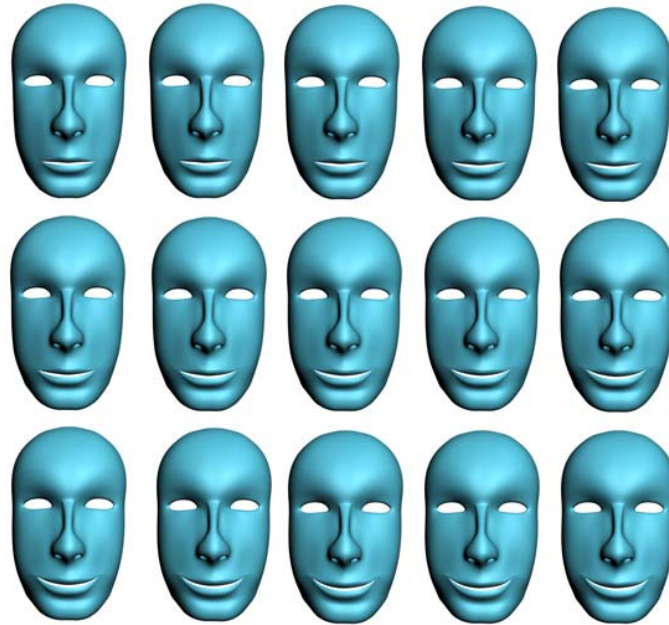


Figura 6.12: Associação no modelo e interpolação entre as poses chaves.

6.1.3. Emburrado

Apesar da expressão emburrado não constar das expressões universais, esta é considerada neste trabalho para demonstrar que é possível a utilização de outras expressões, além das já citadas, no sistema proposto.

A figura 6.13 mostra a seleção de quadro-chave da câmera esquerda (2 quadros extremos), para a expressão de emburrado, usados para interpolação e o cálculo estereoscópico.

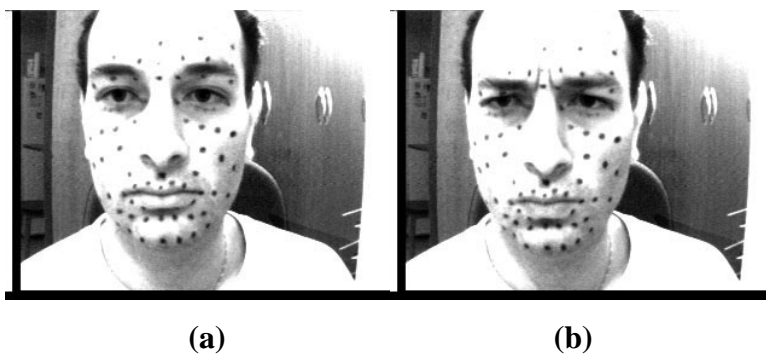


Figura 6.13: Quadros da câmera esquerda que foram selecionados para a interpolação, no caso da expressão facial de emburrado. (a) quadro inicial e (b) quadro final.

A figura 6.14 mostra a seleção de quadros chaves da câmera direita (2 quadros extremos), para a expressão de emburrado, usados para interpolação e o cálculo estereoscópico.

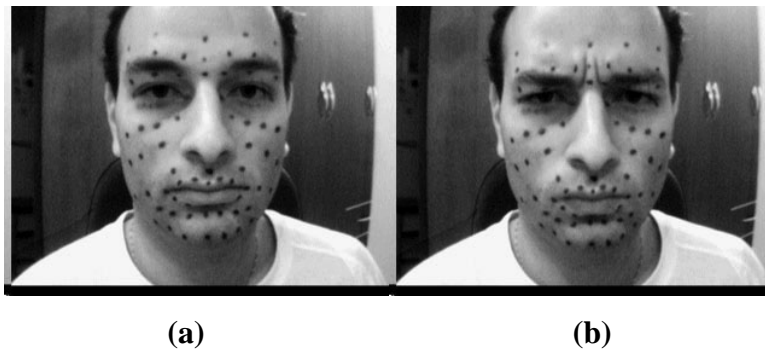


Figura 6.14: Quadros da câmera direita que foram selecionados para a interpolação, no caso da expressão facial de emburrado. (a) quadro inicial e (b) quadro final.

6.1.3.1. Estereoscopia dos pontos correspondentes de cada par de quadros equivalentes (Processo de *Matching*)

A figura 6.15 mostra alguns aspectos de estereoscopia dos pontos correspondentes no par de quadros equivalentes, no caso de expressão de emburrado.

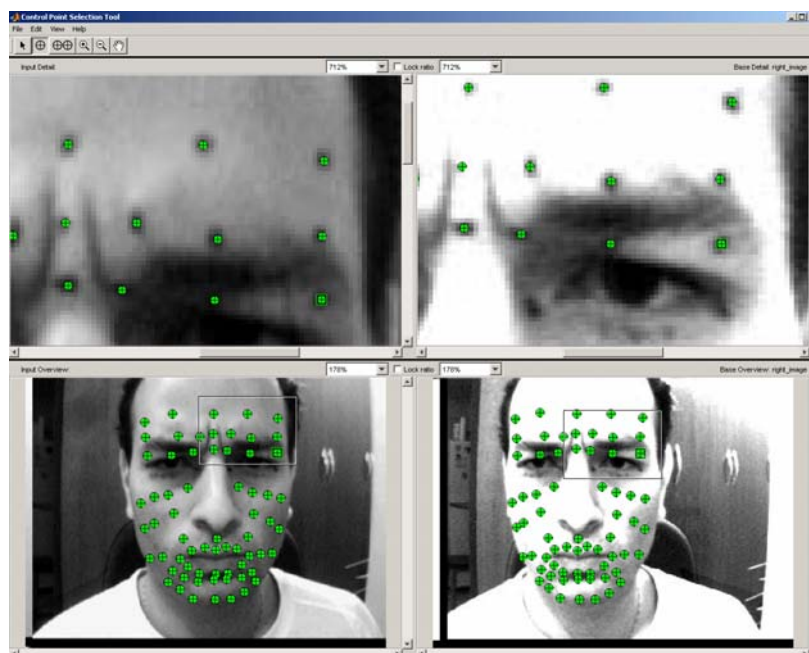


Figura 6.15: Encontro de pontos semelhantes em imagens correspondentes da esquerda e da direita no par de quadros, no caso de expressão de emburrado.

Finalmente, é mostrada na figura 6.16, a associação no modelo 3D e a interpolação:



Figura 6.16: Associação no modelo e a interpolação entre as poses-chave para o caso de expressão de emburrado.

O aparecimento de eventuais rugas percebidas nas imagens do vídeo ocorre devido ao elevado número de pele dobrável na região. Porém, para que estas sejam capturadas para a face 3D, faz-se necessário o aumento de pontos de controle na região de dobra na face 3D e o aumento de marcações na face real.

6.2. Sincronismo Labial (*Lip Sync*)

O estudo da animação correspondente ao sincronismo labial, cuja descrição se encontra na seção 3.8, foi desenvolvido levando-se em consideração o exemplo de fala: “*Olá como vai você?*”, que está demonstrada através do gráfico da figura 6.17. É importante destacar que esta fala foi gravada juntamente com a captação de vídeo, agilizando o processo de produção de animação facial.

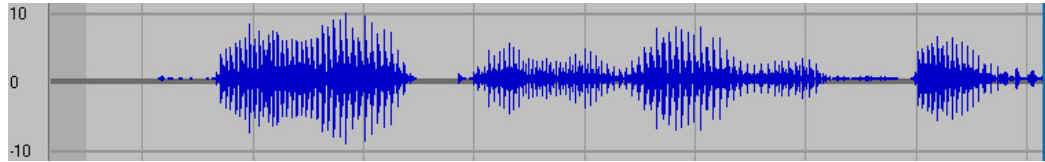


Figura 6.17: Gráfico da fala usado para referência no sincronismo labial.

O *Lip Sync* foi desenvolvido em uma face 3D referente ao modelo real (obtido a partir da imagem facial humana) e em um modelo de personagem 3D (*cartoon*).

6.2.1. Modelo Aplicado em Face Humana 3D

A seleção de quadros-chave levou em consideração, as diferentes aberturas da região labial e os momentos em que estas se fecham, abrangendo assim, as principais poses da animação.

Neste estudo, assumiu-se que os movimentos de um lado da face podem ser espelhados no outro, quando um movimento for considerado simétrico. Desta forma, apenas a metade do rosto foi demarcada devido à simetria da animação facial, reduzindo o cálculo estereoscópico pela metade.

A figura 6.18 mostra a seleção de quadro-chave da câmera direita (10 quadros), para o caso de sincronismo labial.



Figura 6.18: Quadros da câmera direita que foram selecionados para a interpolação, objetivando o estudo do sincronismo labial.

A figura 6.19 mostra a seleção de quadros-chave da câmera esquerda (10 quadros), para o caso de sincronismo labial.



Figura 6.19: Quadros da câmera esquerda que foram selecionados para a interpolação, objetivando o estudo do sincronismo labial.

A figura 6.20 mostra os aspectos de estereoscopia dos pontos correspondentes de um par de quadros equivalentes (Processo de *Matching*).

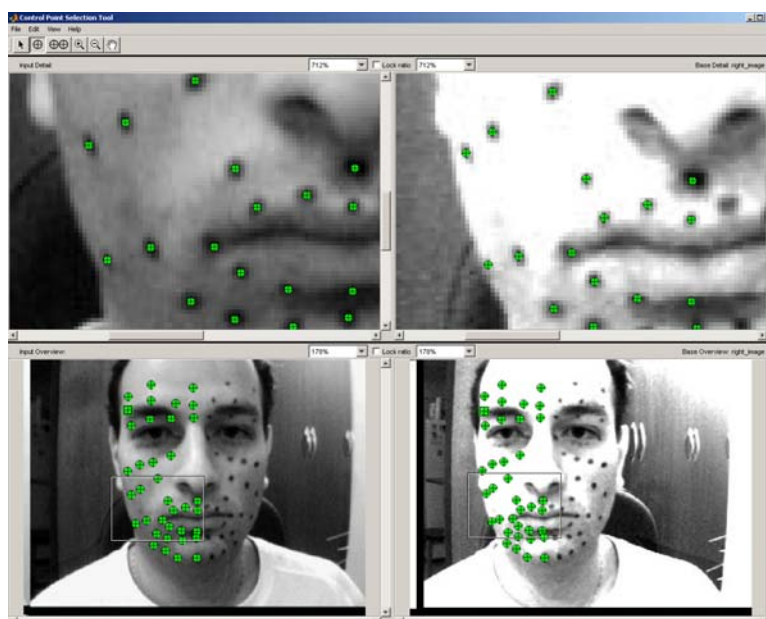


Figura 6.20: Encontro de pontos semelhantes em imagens correspondentes da esquerda e da direita no par de quadro.

Finalmente, a figura 6.21, mostra a associação no modelo de animação 3D e a sua interpolação:

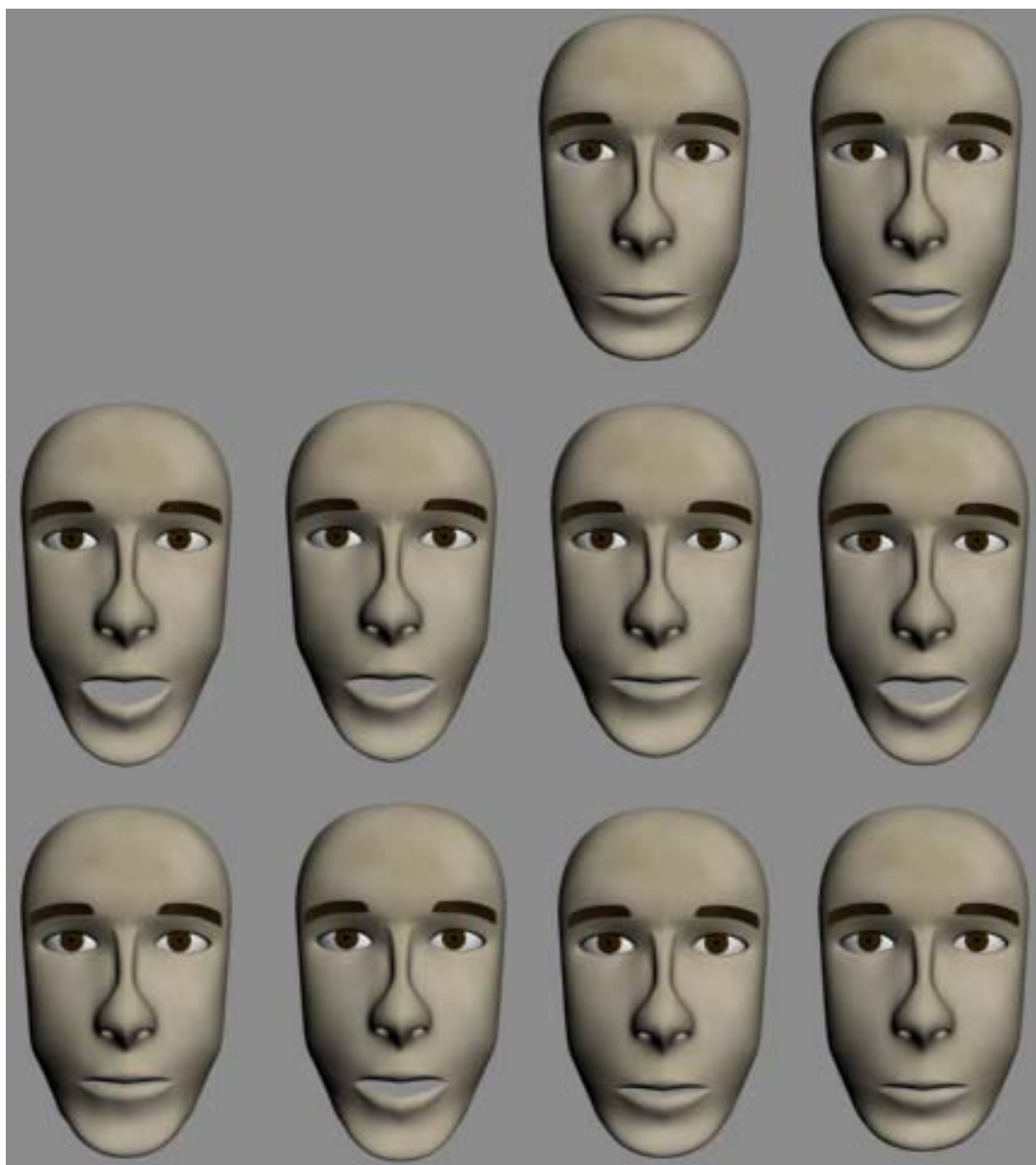


Figura 6.21: Associação no modelo e a interpolação entre as poses-chave.

6.2.2. Modelo Aplicado em uma Face *Cartoon*

A implementação da estereoscopia em uma outra face 3D, como a da figura 6.22, é realizada da mesma forma, porém, dependendo da topologia da boca, o resultado pode não ser consistente com as características do *cartoon* devido à escala utilizada.

Neste caso, apenas os pontos da região da boca são utilizados para o sincronismo labial da personagem.

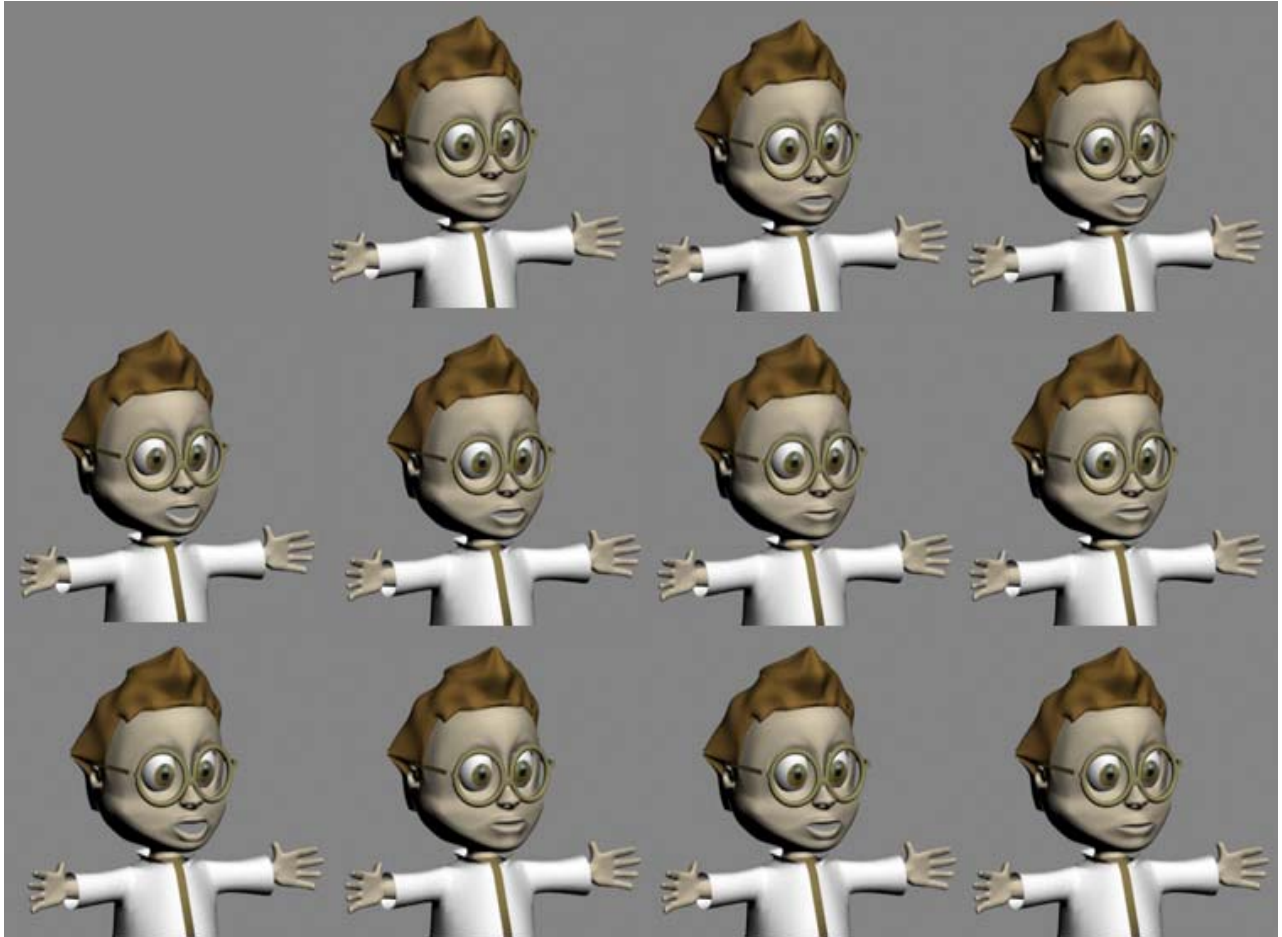


Figura 6.22: Associação no modelo e interpolação entre as poses-chave em uma face *cartoon*.

6.3. Discussão

Para cada expressão facial realizada neste trabalho, foram usados não mais do que três quadros para se atingir uma animação realística.

Contudo, o número de quadros a ser usado para a realização de uma animação facial, pode variar para cada caso, de acordo com a necessidade do animador. Como exemplo, na seção 6.2, o sincronismo labial da frase (“*Olá como vai você?*”) exigiu a escolha de determinadas poses, para que a animação atingisse a sincronia labial, tendo no final, um total de dez quadros.

Como há esta interação com o sistema de animação facial, a taxa de amostragem deste método de captura não chegaria a mais de 30 quadros por segundo, devido à limitação das câmeras utilizadas. A grande vantagem desta característica, é que o sistema não produz informações exageradas e inúteis, facilitando a tarefa de retirar estes quadros, em excesso, produzidos na animação antes dos cálculos de estereoscopia.

O método proposto de estereoscopia mostrou-se eficaz (de acordo com os resultados visualmente obtidos do capítulo 6) para a obtenção de informações volumétricas a partir de imagens, mesmo utilizando câmeras de baixa resolução (320 x 240 *pixels*).

Objetivando um sistema de baixo custo financeiro e de experimentação, as câmeras utilizadas no projeto, custaram em torno de R\$ 40,00. Contudo, a precisão deste método pode alcançar melhores níveis com o uso de câmeras de maior qualidade.

A tabela 4 mostra uma comparação entre o sistema de animação facial por *motion capture* óptico da Universidade de Ohio (SOMASUNDARAM, 2005) com o sistema proposto.

Tabela 4: Comparação entre o sistema óptico de *motion capture* da Universidade de Ohio (SOMASUNDARAM, 2005) com o sistema proposto.

	<i>Motion Capture</i>	Sistema Proposto
Características principais dos sistemas	Conjunto de câmeras, fazendo o <i>tracking</i> 2D dos refletores para posterior conversão para 3D	2 Câmeras que capturam os pontos marcados na face humana.
Dados obtidos	Posição <i>XYZ</i> dos refletores	Posição <i>XYZ</i> dos pontos marcados no rosto
Taxa de amostragem	200 quadros por seg. [60]	Até 30 quadros por seg.
Processamento de dados	<i>Tracking</i> via software dos conjuntos de posições 2D para posterior geração de coordenadas 3D, não sendo um sistema de tempo real.	Idem
Principais vantagens	Alta taxa de amostragem (até 200 quadros por segundo) (captura de	Baixo custo financeiro, R\$1.000,00, (valor dos materiais utilizados), captura

	movimentos mais complexos com excelente precisão), número de refletores ilimitado, liberdade dos atores durante o processo	de movimentos com precisão (de acordo com os resultados visualmente obtidos do capítulo 6), liberdade dos atores durante o processo de produção, número alto de marcadores sem atrapalhar o movimento (entre 1 a 100) e geração de quadros-chave apenas
Principais desvantagens	Oclusão de refletores, necessita <i>tracking</i> via <i>software</i> para geração de coordenadas 3D, alto custo do equipamento e excesso de quadros produzidos	Baixa taxa de amostragem (até 30 quadros por segundo), necessita <i>tracking</i> via <i>software</i> para geração de coordenadas 3D, oclusão dos pontos marcados no corpo
Preço médio equipamento	US\$ 20.000,00 a 150.000,00 (SILVA,1997)	R\$ 1000,00 (valor dos materiais utilizados)
Interatividade com o sistema	Após a captura de informações volumétricas gerando excesso de dados a serem excluídos na fase de produção.	Antes dos cálculos das informações volumétricas, produzindo apenas quadros-chave.

Capítulo 7: Conclusões, Contribuições e Trabalhos Futuros

Este capítulo trata da conclusão desta dissertação de mestrado apontando as contribuições dos métodos do sistema, as limitações e os trabalhos futuros.

7.1. Contribuições

O desenvolvimento do método de estereoscopia, na área de animação facial, contribui para uma automatização e facilidade de implementação na mesma, permitindo uma precisão suficiente para uma animação facial realista.

O sistema estereoscópico pode ser caracterizado pelas contribuições de:

- Re-uso e portabilidade: a captura da seqüência da animação facial é dada na forma de dados, ou seja, o uso destes para um outro personagem ou uma outra cena é possível ao carregar o arquivo de dados.
- Escolha de quadros chave: a produção de quadros chave na animação se dá previamente ao cálculo estereoscópico. Esta característica faz com que o sistema produza apenas os quadros escolhidos, economizando cálculos na fase do *tracking* e na aquisição tridimensional. O método de escolha dos quadros chave é visual fazendo com que este processo se torne artístico.
- Criação de *Targets Morphing*: os fonemas e as expressões faciais podem ser capturados e armazenados em formas de *Target Morphings*, ou seja, uma animação pode ser repetida, ou modificada, diversas vezes. É possível também criar um banco de fonemas a serem utilizados pela face do personagem construindo assim sincronismos labiais através da utilização do *morphing*.
- Criação de *Keyframes*: os *keyframes* podem ser manipulados, acelerando ou desacelerando uma animação e também possuem a característica de terem suas curvas ajustáveis, podendo ter um movimento oposto, ou seja, do *frame* posteriormente obtido ao anteriormente obtido.
- Sistema de baixo custo financeiro: a utilização de duas câmeras de segurança monocromáticas e de duas placas de aquisição de vídeo, como principais ferramentas, torna o sistema de baixo custo financeiro. Vale ressaltar também, que a precisão da animação atingiria níveis mais altos com a utilização de melhores câmeras,

aumentando o custo do sistema, porém mantendo-se ainda muito abaixo de outros sistemas ópticos existentes (de acordo com tabela 2).

- Facilidade de uso e de mudanças: os métodos utilizados neste projeto facilitam a qualquer usuário implementar uma animação facial, ou uma sincronia labial, mesmo que este não conheça os princípios da animação. A mudança de um movimento facial pode ser facilmente alcançada pela translação dos vértices dos movimentos já capturados, alcançando assim um resultado satisfatório sem a necessidade de novas gravações.
- Simetria: o desempenho do sistema pode ser melhorado com o uso de expressões simétricas, diminuindo assim, o cálculo realizado pelo sistema.
- *Sistema Lip Sync* multilingüe: o sistema de *Lip Sync*, alcançado pelo uso da estereoscopia, é caracterizado por ser independente do locutor a ser utilizado na animação, ou seja, a não dependência de fonemas de uma certa língua, mas sim, do uso de uma seqüência de imagens dos movimentos labiais. Vale destacar que o processo de gravação de voz e o da filmagem para o *Lip Sync* pode ser realizado em uma única etapa, aprimorando assim o *workflow* do sincronismo labial.

7.2. Limitações

Durante a execução dos métodos foram evidenciadas algumas limitações, que estão destacadas a seguir:

- Alta quantidade de dados: pode ocorrer se não houver uma seleção de quadros-chave, tornando a animação mais próxima da performance humana, porém, mais carregada para se editar.
- Baixa qualidade das câmeras: O uso de câmeras de baixo custo financeiro resultou em imagens com diferentes e poucos contrastes, prejudicando assim a utilização de um algoritmo de busca. Como teste de precisão utilizou-se um algoritmo que usa o método de mínimos quadrados para a busca da trajetória dos pontos correspondentes. Porém, seu resultado não foi considerado satisfatório com o uso das duas câmeras monocromáticas.

Deste modo, constatou-se que, para a automatização da etapa de seleção dos pontos correspondentes, é necessário o uso de câmeras de maior qualidade (acima de 320 x 240 *pixels* e colorida).

- Seleção de quadros-chave errados: a seleção de quadros-chave não apropriados pode gerar uma animação não condizente com a proposta e sem fluidez. Esta etapa exige um profissional artístico especializado que consiga determinar a mínima quantidade de quadros a ser escolhida para a economia nos cálculos do sistema proposto.
- Alinhamento das câmeras: para o cálculo estereoscópico funcionar, é necessário que as câmeras estejam alinhadas por uma linha horizontal, porém este problema é contornado por soluções via *software* com a translação das imagens geradas em um determinado eixo.
- Movimento da cabeça: o sistema de câmeras fixas pode gerar problemas devido à movimentação natural da cabeça. Os pontos faciais devem ser movimentados para expressar movimentos faciais e não translações e rotações da cabeça. Para isso, deve-se adicionar três pontos de referência no rosto (um no centro do nariz e dois posicionados em cima de cada sobrancelha) para calcular os movimentos de rotação e translação da cabeça toda.
- Automatização do *Fitting*: para um melhor desempenho, faz-se necessário a automatização da etapa de *fitting*. Como a face possui muitos vértices de controle, a inserção automática das novas posições destes na face 3D, pouparia enormes esforços despendidos nesta etapa.
- Precisão do algoritmo de *tracking*: prejudica o desempenho do sistema, pois quando o mesmo não consegue achar a posição de um marcador, ele entra em modo manual pedindo ao usuário a seleção do ponto.
- Imperfeições da realidade: assim como outros sistemas de captação de movimentos reais, este método de estereoscopia tem uma excelente simulação do real. Contudo, para que o espectador compreenda algumas expressões, ainda pode ser necessário o uso de conceitos de animação como o de exagero (seção 2.2). Portanto, os dados provenientes da performance humana real captada em vídeo devem ser editados.

7.3. Trabalhos Futuros

A continuação deste projeto, descritas como trabalhos futuros, incluem os seguintes itens:

- Melhoria no algoritmo: melhoria do algoritmo de *tracking*, dando maior precisão e implementação do algoritmo de *fitting* para um maior desempenho do sistema.
- Portabilidade dos dados: os dados dos quadros-chave provenientes da etapa da estereoscopia devem comunicar-se com o sistema, podendo ser carregado em outras faces 3D, manipulados e salvos com uma extensão proprietária.
- Melhoria na precisão da estereoscopia: para a melhoria da precisão do sistema, faz-se necessário a aquisição de câmeras de melhor qualidade e em maior número destas.
- Sistema de *Lip Sync* híbrido: implementação de um sistema híbrido (*data e speech driven*) para o *Lip Sync*, utilizando parâmetros de variáveis provenientes de imagens e da onda sonora do locutor para a melhoria da precisão.

Glossário

Disparidade	Nome dado à separação física ao longo de uma linha epipolar entre características equivalentes de duas imagens estereoscópicas.
Linha Epipolar	É a interseção do plano de imagem com o plano epipolar, que por sua vez é o plano que contém o ponto 3D procurado e o centro das duas lentes. Assim, no caso das câmeras estarem dispostas com seus eixos ópticos em paralelo e os planos de imagens terem sofrido deslocamento horizontal a linha epipolar é uma reta horizontal, e portanto os pontos candidatos estão sobre ela.
Correspondência	Processo na qual consiste em associar características equivalentes de duas imagens estereoscópicas.
Calibração Estéreo	Nome dado a algoritmos que retornam precisas informações métricas considerando a estrutura 3D real da cena.
Retificação de Imagem	Aplicação de técnicas para alinhar as linhas epipolares em duas imagens estereoscópicas.
Oclusão	A oclusão ocorre quando o formato da imagem real cobre outros próprios pontos num determinado ponto de vista, gerando buracos na conversão de pontos 3D.
Rotoscopia	Rotoscopia é um processo que consiste em redesenhar quadros (<i>frames</i>) de um vídeo. Animar ou desenhar seguindo uma referência filmada. As imagens em questão podem servir de referência e o autor pode também, “fantasiar” em cima desta.
Box Modeling	A técnica de <i>Box modeling</i> consiste em começar a modelagem de um objeto qualquer, a partir de uma primitiva básica como um cubo ou uma esfera chegando

	em sua forma final desejada através de extrusões de arestas e translações dos vértices das faces poligonais da mesma.
Lip Sync (Sincronismo Labial)	É a sincronização dos rápidos movimentos labiais e da língua durante uma fala, determinando um vasto número distinto de posições faciais.
Quadros Chaves (Keyframes)	Quadros chaves são as principais poses de uma animação, ou seja, são a partir delas que um animador produz os quadros intermediários.
<i>Motion Capture</i>	<i>Motion capture</i> é a criação de uma representação 3D, a partir de uma performance real humana, utilizando emissores que geram sinais captados por sensores estrategicamente distribuídos ou fixos.
<i>Morphing</i>	São efeitos visuais produzidos por computador, no qual um objeto / imagem se transforma em outro objeto / imagem.
<i>Game Engine</i>	Game Engine é uma estrutura de programação, muitas vezes acompanhado de um ambiente de desenvolvimento integrado (<i>Integrated Development Environment - IDE</i>) utilizado para acelerar o desenvolvimento de jogos. Esta contém módulos de renderização, cálculos físicos e matemáticos, e pode ainda fornecer facilidades para programação de inteligência artificial e acesso à rede.
<i>Matching</i>	Processo de identificação dos pontos correspondentes em imagens estereoscópicas para resultar na projeção 3D de um objeto real.
Subdivisão de Superfície	Técnica que permite a geração automática de novos polígonos sobre a superfície aplicada, além da geração de uma tensão nas quinas dos mesmos, deixando um formato mais arredondado na geometria 3D.

Apêndice

Quadro 1:

>> X

X =

17.2819	36.3478	54.0913	73.5247	31.0781	53.8745	70.0271	-12.0405	9.3835
46.2017	-11.0483	0.4099	13.6148	26.6510	-10.4116	6.1273	17.9316	-10.2711
4.3788	19.9843	33.8423	66.4601	-11.7004	9.2392	23.9504	39.1130	-29.9416 -
44.0725	-57.1367	-71.5704	-45.0686	-61.4066	7.8939	-39.4429	-20.2184	-30.4306 -
44.4748	-55.9557	-69.0078	-23.9288	37.0249	-21.8327	-35.8667	-46.4853	-26.1460 -
37.9782	-50.8830	-6.8044	40.6420	77.1643	-6.1014	11.9700	39.1969	-7.0394
15.8340	39.9841	67.1688	75.5982	-41.6145	-70.9732	-26.5404	-48.1571	-69.5351 -
22.9718	-43.6632	-68.9446						

>> Y

Y =

-24.9241	-18.0843	-16.4952	-10.6517	-0.0599	11.3166	18.2112	21.1644	23.1003
47.4986	30.8437	28.4971	32.8321	43.3908	50.9194	52.5268	48.0025	58.6984
62.8207	62.4593	61.3472	55.3271	79.7928	83.3908	79.4408	77.3246	-25.6034 -
20.7802	-18.1847	-16.3531	-0.9218	5.2186	12.6692	18.8024	26.8419	28.5579
35.3405	35.3678	38.0801	48.7711	42.3654	57.2533	57.1284	52.7377	76.5864
71.2170	64.2440	-93.6359	-108.5405	-107.4898	-73.8237	-78.4095	-85.7297	-60.4134
-62.6994	-66.6746	-65.0014	-88.2892	-96.6376	-95.2312	-76.2643	-77.3405	-75.0381 -
60.4765	-61.3871	-60.8684						

>> Z

Z =

6.2973	6.3755	6.5436	6.8296	6.3359	6.5732	6.8643	5.8851	6.2105	6.5882
5.9448	6.0177	6.1630	6.3490	6.0830	6.1985	6.2471	6.1282	6.2348	6.3622
6.5436	7.0094	6.1985	6.3490	6.5289	6.7130	6.1168	6.0942	6.0719	6.1630
6.0830	6.1282	6.2348	6.0500	5.9147	5.9653	6.0719	6.1513	6.3101	6.0391

```

6.0391  6.1055  6.1630  6.1397  6.1168  6.1055  6.2226  6.1513  6.5144  7.0094
6.0942  6.1985  6.3622  6.0609  6.2226  6.4024  6.6970  6.8995  6.1055  6.2720
5.9965  6.0500  6.1513  6.0070  5.9965  6.1513

```

Quadro 2:

```
>> X
```

```
X =
```

```

17.0415  37.6126  55.0871  72.7577  32.0467  54.2689  67.9641 -10.0481  20.9660
45.7949 -9.1604  1.5139  14.3855  25.0236 -9.2454  6.8812  18.5891 -8.5865  6.0077
21.5100  33.6763  63.9814 -10.0154  10.6190  25.2552  39.8079 -28.8551 -42.1418 -
56.3913 -70.1618 -43.1685 -59.2088  -70.0916 -37.4700 -18.6531 -27.8219 -40.4493 -
51.0849 -65.0739 -22.7055  -35.5217 -20.5393 -34.5606 -44.4593 -25.0337 -37.1145 -
48.9385  -5.1618  46.1245  82.9073  -4.3433  13.5698  44.3387  -5.1942  19.7585
44.0161  71.3480  79.0512 -42.6232 -73.9602 -25.5217 -48.0423 -74.9565 -22.2676  -
45.2130 -68.5018

```

```
>> Y
```

```
Y =
```

```

-27.9824 -19.8235 -17.6570 -12.0499  -1.4604  10.8599  17.8685  19.1913  22.5836
48.9943  28.3436  26.1968  31.2170  46.7113  63.5336  64.4623  56.0646  70.3183
73.1454  70.1392  65.1646  54.6264  90.8641  92.1614  85.7936  80.4571 -28.3164 -
21.7565 -19.3309 -16.6740  -1.7754   5.9512  13.0712  19.1684  24.9819  27.6187
38.4952  37.8770  39.1621  60.8897  51.3937  68.5568  65.8510  58.1654  86.8340
80.8575  70.0949 -104.9828  -121.3084 -121.5382 -89.9826 -94.7748 -105.3038 -71.7818
-76.8126 -84.9812 -80.3319 -103.9783 -108.0990 -105.9575 -93.4614 -92.2371 -92.3136 -
73.0822 -79.4055 -73.6359

```

```
>> Z
```

```
Z =
```

```

6.2226  6.3755  6.5436  6.7621  6.3230  6.5882  6.8296  5.9347  6.2226  6.6496
5.9147  5.9653  6.1055  6.3622  6.1282  6.2720  6.3101  6.2105  6.2973  6.4297
6.5436  6.9721  6.2846  6.3889  6.5732  6.7293  6.0942  6.0284  6.0942  6.1397

```

6.0609	6.1282	6.2595	6.0177	5.9048	5.9247	6.0942	6.1513	6.3230	6.1055
6.0942	6.1985	6.2348	6.1630	6.1985	6.2595	6.2846	6.1397	6.5436	7.1659
6.1866	6.2226	6.4435	6.0719	6.2226	6.4435	6.7621	6.9721	6.0942	6.4024
6.0942	6.0070	6.3755	6.0070	6.0830	6.1513				

Referências Bibliográficas

BLAIR, P. Cartoon Animation. Walter Foster Publishing, 1994.

BERGERON, P.; LACHAPELLE, P. Controlling facial expressions and body movements. Em: Advanced computer animation, SIGGRAPH '85 Tutorials, ACM, New York, v. 2, p. 61-79, 1985.

BRENNAN, S. E. Caricature generator. Dissertação de Mestrado, Massachusetts Institute of Technology, Cambridge, 1982.

CARLBOM, I.; PACIOREK, J. Planar Geometric Projections and Viewing Transformations. Computing Surveys, v.10, n.4, 1978.

CHERNOFF, H. The use of faces to represent points in n-dimensional space graphically. Technical Report Project NR-042-993. Washington: Office of Naval Research, 1971.

CURLESS, B.; LEVOY, M. Better Optical Triangulation through Spacetime Analysis, 1995. Disponível em: www.graphics.stanford.edu/papers/spacetime/spacetime_full.pdf. Data de acesso: 08/02/2005

CYBERWARE LABORATORY. 3D Scanner with Color Digitizer. Monterey, CA, 1990.

D'APUZZO, N. Modeling Human Faces with Multi-Image Photogrammetry. ETH Zurich, Suíça: Institute of Geodesy and Photogrammetry, 2002. Disponível em: www.photogrammetry.ethz.ch/projects/face/publications. Data de acesso: 17/11/2004

DENG, X. Q. A finite element analysis of surgery of the human facial tissue. Tese de Pós-Doutorado, Columbia University, New York, 1988.

DONALD, H.; BAKER, M. P. Computer Graphics C Version, 2ª Ed., 1997.

DUDA, R.; HART, P. Use of the Hough Transformation to Detect Lines and Curves in Pictures. *Communications of the ACM*. v. 15, n. 1, 1972.

EKMAN, P. The argument and evidence about universals in facial expressions of emotion. Em: WAGNER, H.; MONSTEAD, A. *Handbook of Social Psychophysiology*, p. 143-146, 1989.

EKMAN, P.; FRIESEN, W. V. *Manual for the facial action coding system*. Consulting Psychologists Press, Palo Alto, CA, 1978.

FRANÇA, J. G. D. M. Desenvolvimento de um sistema de aquisição de informações volumétricas usando o método de triangulação a laser e campo de visão lateral. Trabalho apresentado no Exame de Qualificação do Programa de Pós-Graduação em Ciência da Computação, Universidade Federal de São Carlos, Programa de Pós-Graduação em Ciência da Computação, 2003.

FRANK, T.; HOCH, M.; TROGEMANN, G. *Automated Lip-Sync for 3D-Character Animation*. Alemanha: Academy of Media Arts, 1997.

GONZALEZ, R. C.; WINTZ, P. A. *Digital Image Processing*. Addison – Wesley, 1987, 2ª Ed.

GREWE, L. L.; KAK, A. C. *Handbook of pattern recognition and image processing: Computer vision*. Academic Press, 1994.

GRUEN, A. Adaptive least squares correlation: A powerful image matching technique. *South African Journal of Photogrammetry*, v. 14, n. 3, p. 175-187, 1985.

GUTIERREZ-OSUNA, R. *Speech-Driven Facial Animation With Realistic Dynamics* - *IEEE Transactions on Multimedia*, v. 7, n. 1, 2005.

HJORTSJO, C. Man's face and mimic language. Suécia: Student-literatur, 1970. Disponível em: <http://face-and-emotion.com/dataface/facs/guide/FACSIV1.html>. Data de acesso: 10/05/2004.

HGN PRODUÇÕES Ltda. Apostila do curso de animação. São Paulo, 1997.

KAHLER, K.; HABER, J.; SEIDEL, H. P. Geometry-based Muscle Modeling for Facial Animation. Max-Planck. Alemanha: Institut fur Informatik , 2001.

KURIHARA, T.; ARAI, K. A transformation method for modeling and animation of the human face from photographs. Em: State of the Art in Computer Animation, Springer-Verlag, p. 45–57, 1991.

LARRABEE, W. A finite element model of skin deformation, 1986.

LEE W. S.; MAGNENAT-THALMANN, N. Head modeling from pictures and morphing in 3d with image metamorphosis based on triangulation. University of Geneva, 1998.

LEE, Y.; TERZOPOULOS, D.; WATERS, K. Constructing Physics-Based Facial Models of Individuals. Em: Graphics Interface '93 Conference, Toronto, Canadá, p. 1-8, 1993.

LENT, R. Cem bilhões de neurônios: Conceitos fundamentais de neurociência. São Paulo: Atheneu, 2001.

LI, Z. et. al. Spacetime Faces: High Resolution Capture for Modeling and Animation. University of Washington, 2005.

LUCENA JUNIOR, A. Arte da animação: Técnica e estética através da história. São Paulo: Editora SENAC, 2002.

LUTZ, E. G. Animated cartoons: How they are made, their origin and development. Bedford: Applewood Books, 1998.

MAGNENAT-THALMANN, N.; THALMANN D. Abstract muscle actions procedures for human face animation. *Visual Computer*, v. 3, n. 5, p. 290-297, 1988.

MARTINEZ, A.M.; BENAVENTE, R.. The AR Face Database. CVC Technical Report #24, June, 1998.

MICROSOFT TM. XBOX 360, 2006. Disponível em: <http://www.microsoft.com>. Data de acesso: 07/02/2006

MORRISON, M. *Becoming a computer Animator*. Indianápolis: Howard W. Sans, 1994.

MURDOCK, K. L. *3D Max 6 Bible*, Wiley Publishing Inc. 2004.

NAHAS, M. et. al. Facial image synthesis using skin texture recording. *Visual Computer*, v. 6, n. 6, p. 337–343, 1990.

PARKE, F. Parameterized models for facial animation revisited. Em: *SIGGRAPH Facial Animation Tutorial Notes*, p 43–56. ACM SIGGRAPH, 1989.

PARKE F. I.; WATERS, K. *Computer Facial Animation*. Massachusetts: A K Peters, 1996.

PARKER, S. *Como a mente Funciona*. Companhia das Letras, 2004, 2ª Ed.

PINGHIN, F. et. al. Synthesizing Realistic Facial Expressions from Photographs. 25th Annual Conference on Computer graphics and interactive techniques, 1998. Disponível em: <http://portal.acm.org/citation.cfm>. Data de acesso: 21/10/2004

PLATT S. M. *A system for computer simulation of the human face*. Dissertação de Mestrado, The Moore School, University of Pennsylvania, Philadelphia, 1980.

PLATT, S.; BADLER, N. Animating facial expression. *Computer Graphics*, v. 15, n. 3, p. 245–252, 1981.

- SANTOS- VITOR, J. Notas do Curso de Visão. Portugal: Instituto Superior Técnico, 2004.
- SILVA, F. Motion Capture: Introdução à Tecnologia. Laboratório de Computação Gráfica, LCG. COPPE, Universidade Federal do Rio de Janeiro, UFRJ, 1997.
- SOLOMON, C. (ORG) The Art of the Animated Image: An Anthology. Los Angeles: The American Film Insitute, 1987.
- SOMASUNDARAM, A. Realistic Facial Animation During Speech. CSE-ACCAD, The Ohio State University, 2005. Disponível em: http://accad.osu.edu/research/mocap/papers/Arun_ACCADSummary.pdf. Data de acesso: 07/07/2005.
- THOMAS, F.; JOHNSON, O. The Illusion of Life. New York: Abbeville Press, 1981.
- TOMMASELLI, A. M.G.; HASEGAWA, J. K.; GALO, M. Modernas Tecnologias de Aquisição de Imagens em Fotogrametria. Universidade Estadual Paulista, 2000.
- ULRICH, P.; ASHDOWN, K. S.; WORONKA, D. Body scan analysis of posture and body shape for reengineering. Nation Textile Center, 2003. Disponível em: www.ntcresearch.org/pdf-rpts/AnRp04/S01-AC27-A4.pdf. Data de acesso: 26/09/2004
- VALLE, A. Projeto stereo binocular de Alberto Vale. Portugal: Instituto Superior Técnico, 2004. Web: <http://lrm.isr.ist.utl.pt/vale/projects/visao/visao.html>. Data de acesso: 12/06/2005
- YAU, J.; DUFFY, N. 3-D facial animation using image samples. Em: New Trends in Computer Graphics, Springer-Verlag, p 64–73, 1988.
- YUENCHENG, L.; DEMETRI, T; WATERS, K. Realistic Modeling for Facial Animation. University of Toronto and Digital Equipment Corporation, 1993.
- VICKERS, P. How to Generate Stereoscopic 3D Video Using Cheap Board Cameras, 2002.

WATERS, K. A muscle model for animating three-dimensional facial expression. *Computer Graphics*, v. 22, n. 4, p 17–24, 1987.

WILLIAMS, L. Performance-driven facial animation. Em: SIGGRAPH 24, *ACM Computer Graphics*, p. 235–242, 1990.

ZHANG, L. et. al. *Spacetime faces: High resolution capture for modeling and animation*, University of Washington, 2003.