

ALGORITMO DE ANIMAÇÃO FACIAL BASEADO EM DEFORMAÇÃO DE MALHA

Edigleison F. CARVALHO (1); Ricky PAZ (2); Mauro GADELHA (3); Marcos FILHO (4); José MARQUES (5)

(1) CEFET-CE, Rua Pierre Luz, 3286-3337, e-mail: edigleison.carvalho@gmail.com

(2) CEFET-CE, e-mail: rickypaz@gmail.com
(3) UFC, e-mail: mauro.gadelha@hotmail.com
(4) FIC, email: marques.soares@gmail.com
(5) CEFET-CE, email: marques.soares@gmail.com

RESUMO

A animação de faces virtuais de modelos tridimensionais exige considerável poder computacional, o que dificulta a aplicação em dispositivos com baixa capacidade de processamento. Com o objetivo de criar técnicas de animação eficientes para uso em dispositivos móveis, foi desenvolvido um sistema que permite a configuração de parâmetros de animação facial e visualização da projeção do modelo 3D. O ambiente desenvolvido permite a parametrização do modelo e o teste de animação em tempo real. Para o desenvolvimento dessa ferramenta, foram realizadas pesquisas sobre algoritmos de animação baseados em deformações geométricas de malhas tridimensionais. Essas deformações são conseqüentes dos deslocamentos de pontos característicos que atuam em determinadas regiões da face. Estes pontos são definidos pelo padrão ISO/IEC 14496 (MPEG-4), que dedica parte de sua especificação à animação facial. Com o objetivo de animar faces virtuais de maneira sincronizada à fala humana, foram estudadas configurações de parâmetros para a representação dos fonemas. São utilizados quinze conjuntos de parâmetros, um para cada fonema, conhecidos como visemas. Adicionalmente, esta aplicação considera um conjunto de seis estados emocionais que são associados aos visemas para a produção de animações em tempo real.

Palavras-chave: Animação Facial, Deformação de Malha, FP, FAP, FAPU

1. INTRODUÇÃO

A animação facial tem como objetivo simular, de forma convincente, o comportamento da face humana. É possível representar, visualmente, as emoções humanas e as movimentações labiais referentes à fala, transmitindo com exatidão os sentimentos de um humano virtual. Ela é muito aplicada em filmes, jogos, televisão. Além disso, sua utilização em dispositivos móveis se faz possível, reproduzindo mensagens de texto, visualmente, tornando-a mais atraente e não ambígua, porque o estado emocional do remetente estará exposto na face do humanóide, impedindo duplas interpretações quanto à carga emocional posta no texto, já que as mensagens podem conter expressões que não permitem identificar qual o real sentido que o remetente quis passar. Entretanto, esses dispositivos, normalmente, possuem uma limitação em processamento. Para contornar esse problema, foi desenvolvido um algoritmo que utiliza parâmetros para animar um modelo facial tridimensional, exigindo baixo esforço de processamento. Esse algoritmo baseia-se nas diretrizes e especificações fornecidas pelo padrão MPEG-4. Ele é dividido em dois processos distintos que são responsáveis por preparar o modelo para a animação e animá-lo. O primeiro processo, que realiza a parametrização do modelo facial 3D, exige um considerável tempo operacional, entretanto, é necessário realizá-lo somente uma vez. Para não ser necessário parametrizar o modelo nos dispositivos móveis, foi desenvolvido uma ferramenta de parametrização para desktop, e a partir dela, obtêm-se os parâmetros necessários à animação. O processo de animação necessita que o modelo esteja parametrizado e é realizado constantemente, porém, exige pouco processamento. Utilizando parâmetros pré-calculados por essa ferramenta, foi possível obter animações em tempo real, em dispositivos móveis.

2. PARAMETRIZAÇÃO DO MODELO

É utilizado um modelo tridimensional contendo um conjunto de vértices que formam os polígonos de uma ou mais malhas que representam a face virtual. Nessa primeira etapa, selecionam-se as *Facial Animation Parameters Unit (FAPU)* na face neutra, cujas malhas devem estar com os vértices mapeados para permitir a seleção. No próximo passo, devem ser selecionados os *Feature Points (FPs)* na malha, de acordo com o estabelecido pelo padrão MPEG-4. Em seguida, deve ser calculada a região de influência desses pontos sobre os demais vértices ordinários, ou seja, sobre os vértices que não são FPs. No passo seguinte, deve ser calculado, para todos os vértices ordinários, o peso de influência sobre a sua movimentação, para cada FP que o influencia.

2.1 Unidade dos Parâmetros de Animação Facial (Facial Animation Parameters Unit – FAPU)

Com o intuito de permitir que os *Facial Animation Parameters* (FAPs) possam ser utilizados em qualquer modelo facial 3D, foram definidos, pelo padrão MPEG-4, cinco unidades de proporcionalidades. Essas unidades definem medidas importantes na face, que são utilizadas como fatores de aumento ou de queda de intensidade da movimentação, fornecida pelos parâmetros de animação, de um ponto característico. Esse fator permite que o deslocamento do ponto não seja nem brusco e nem imperceptível aos nossos olhos, e que seja o mesmo para qualquer modelo facial tridimensional.

Essas unidades, apresentadas na Figura 1, são definidas a seguir:

- IRISD IRIS *Diameter*: calculada de acordo com o tamanho da íris do modelo (IRISD0/1024).
- ES Eye Separation: calculada com base na distância entre os centros dos olhos do modelo (ES0/1024).
- \bullet ENS *Eye/Nose Separation*: baseada na distância entre a linha que liga os dois olhos e a base do nariz (ENS0/1024).
- MNS *Mouth/Nose Separation*: calculada com base na distância entre a base do nariz e a linha que une os cantos extremos da boca (MNS0/1024).
 - MW Mouth Width: baseada na distância entre os extremos da boca (MW0/1024).

A FAPU final é obtida dividindo-se a distância característica por 1024.

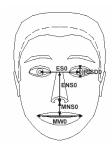


Figura 1 - Unidades de FAPU

2.2 Face Neutra

As unidades faciais de cada modelo são calculadas em relação à face neutra, assim como a seleção dos FPs. As seguintes características a definem:

- 1. Sistema de coordenas da mão direita (eixo X positivo para a direita, Y positivo para cima e Z positivo saindo do papel);
- 2. A face "olha" para a direção Z positiva.
- 3. Todos os músculos da face relaxados.
- 4. Pálpebras tangenciando a íris.
- 5. Pupilas medindo um terço de IRISD0.
- 6. Lábios juntos formando uma linha horizontal de um canto ao outro da boca.

2.3 Pontos Característicos (Feature Points – FPs)

Um dos principais conceitos fornecidos pelo padrão são os *Feature Points* (FPs), ou pontos característicos. Cada FP representa um ponto específico da face. Ao todo, são 84 pontos característicos espalhados pelo cabelo, olhos, nariz, bochechas e queixo. Dada uma malha 3D, deve-se localizá-los (ou um subconjunto deles), e "marcar" os pontos que ocupam suas posições como pontos característicos.

Na Figura 2 é mostrada a distribuição dos FPs em uma face humana:

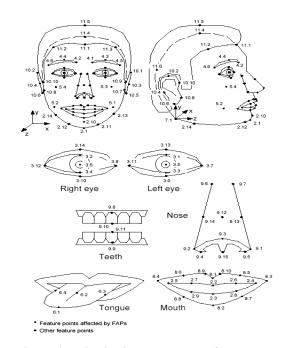


Figura 2 - Distribuição dos FPs da face humana

2.4 Região de Influência

Após a seleção dos FPs, faz-se uma varredura no modelo a fim de se obter a região de influência de cada ponto característico, ou seja, um conjunto de pontos que terão seus movimentos afetados pela movimentação dos FPs. Em vez de se descobrir quais pontos um FP influencia, descobre-se, para cada vértice ordinário, quais FPs o influenciam (GARCHERY, 2005). O processo de análise é feito para cada vértice, é automático e dividido em cinco etapas:

- Criam-se duas listas. Em uma delas, colocam-se todos os possíveis FPs que influenciarão o vértice. Inicialmente, todos os FPs selecionados deverão estar nessa lista. A segunda lista, inicialmente, vazia, conterá os FPs que influenciam o vértice;
- Localiza-se qual FP da primeira lista está mais próximo do vértice corrente. O mesmo é retirado da primeira lista e é colocado na segunda, caracterizando-o como influenciador. A distância entre dois pontos é superficial (não euclidiana), ou seja, ela é calculada seguindo as arestas da malha entre um ponto e outro, podendo-se usar o algoritmo A* para calculá-la (LESTER, 2004);
- Verificam-se quais FPs da primeira lista estão além dos limites definidos por um plano normal ao vetor que o une o vértice corrente e o FP posto na segunda lista na etapa anterior, e os eliminam da lista:
- Repete-se a segunda e a terceira etapa nos demais FPs da primeira lista até esvaziá-la;
- Adiciona-se o vértice corrente na região de influência de cada FP da segunda lista, e repete-se o processo nos demais vértices ordinários da malha.

2.5 Cálculo dos pesos

O movimento de um ponto característico faz com que os vértices que estão na sua região de influência se movimentem com intensidades diferentes. Quanto mais próximo está um vértice ordinário de um FP, maior influencia ele sofrerá quando ocorrer a movimentação desse ponto característico. O fator que determina a proporção de quanto o vértice se deslocará, quando um FP influenciador se deslocar, é chamado de fator peso.

Todos os pontos característicos possuem um fator peso associado a cada ponto de sua região de influência. No cálculo dos pesos, deve-se descobrir, para cada ponto característico, seus vizinhos, que são FPs que possuem pelo menos um vértice em comum em suas respectivas regiões de influência, e as distâncias superficiais até eles (KSHIRSAGAR, 2000).

Considere o ponto P exibido na Figura 3:

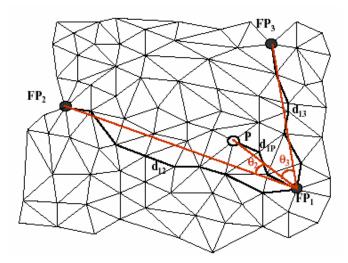


Figura 3 - Parâmetros para cálculo do peso de um vértice com relação a seus influenciadores

Ele é influenciado pelos pontos FP_1 , FP_2 e FP_3 , que são vizinhos entre si. Os valores d_{12} , d_{13} e d_{1P} são as distâncias superficiais entre FP_1 e seus vizinhos e o vértice ordinário P. θ_2 e θ_3 são os menores ângulos formados pelos vetores $FP1_FP2$ - $FP1_P$ e $FP1_P$ - $FP1_FP3$ e possuem valor menor que $\pi/2$, caso contrário, o FP vizinho não será considerado um ponto característico válido no cálculo do peso. Obtido esses valores, eles são computados no cálculo de um fator d, que é a média ponderada das distâncias superficiais d_{12} e d_{13} utilizando como peso os cossenos dos ângulos θ_2 e θ_3 , como mostrado a seguir:

$$d = \frac{d_{12}\cos\theta_2 + d_{13}\cos\theta_3}{\cos\theta_2 + \cos\theta_3}$$
 [Eq. 01]

Se houver apenas um FP vizinho, o FP₂, por exemplo:

$$d = \frac{d_{12}}{\cos \theta_2}$$
 [Eq. 02]

Caso não existam FPs vizinhos válidos, o cálculo de 'd' não poderá ser efetuado, tornando-se impossível prever o que irá acontecer durante a deformação da malha.

Genericamente, sendo n o número de FPs vizinhos, a equação do fator 'd' de um FP influenciador em relação a um ponto P é dada por:

$$d_P = \frac{\sum_{i=0}^{n} [d_i \cos(\theta_i)]}{\sum_{i=0}^{n} \cos(\theta_i)}$$
 [Eq. 03]

 d_i será a distância superficial entre o FP atual e o seu i-ésimo vizinho, e θ_i o ângulo formado entre os vetores FP atual P – FP atual FP_i.

Finalmente, o peso W_{i,P} do FP_i em relação ao ponto P é:

$$W_{i,P} = sen\left(\frac{\pi}{2}\left(1 - \frac{d_{i,P}}{d}\right)\right)$$
 [Eq. 04]

Como um ponto P é influenciado por mais de um FP, o peso dos demais pontos característicos sobre P deverá ser calculado da mesma forma, verificando se o critério da distância angular é satisfeito, e se a distância superficial entre o FP e P é menor que d; se não for, o FP não será considerado durante a deformação da malha.

Ao final, cada ponto terá uma lista de FPs influenciadores e os seus respectivos pesos de influência sobre ele.

3. DEFORMAÇÃO DA MALHA

Depois de parametrizada, a malha 3D está preparada para a deformação. Dado um conjunto de parâmetros de animação, cada FP terá seu deslocamento efetuado, movimentando, consequentemente, os vértices de sua região de influência.

3.1 Parâmetros de Animação Facial (Facial Animation Parameters – FAPs)

A partir de um modelo 3D devidamente parametrizado, é possível gerar uma animação facial por meio dos FAPs. Eles definem o valor dos deslocamentos a serem aplicados em cada ponto característico. O padrão MPEG-4 especifica 68 FAPs, divididos em dois grupos: os de alto-nível e os de baixo-nível. Pertencem aos de alto-nível as categorias "Visemas" e "Expressões", que são formados, cada um, por um conjunto de FAPs de baixo nível. Os demais 66 FAPs são de baixo-nível e estão associados a um único FP, movimentando-o em uma única direção; além disso, relacionam-se com um único FAPU. Quanto maior a intensidade de

movimento provida pelo FAP, maior será o deslocamento sofrido pelo FP ao qual ele está associado. A utilização de FAPs exige pouco esforço computacional e é aplicada em cada animação do modelo.

3.2 Visemas e Emoções

Visemas e Emoções são formados por conjuntos de FAPs de baixo nível de forma a representar um fonema visualmente ou uma das seis emoções básicas adotadas pelo padrão (PANDZIC, 2002): alegria, tristeza, raiva, medo, rejeição e surpresa. Para produzi-los em um modelo 3D, cada FAP do grupo deve ser movido com uma determinada intensidade.

Ao todo, são 15 visemas divididos como mostrado na Tabela 1:

Tabela 1 – Visemas definidos pelo padrão MPEG-4

Visema	fonemas	exemplos
0	none	na
1	p, b, m	put, bed, mill
2	f, v	far, voice
3	T,D	think, that
4	t, d	tip, doll
5	k, g	call, gas
6	tS, dZ, S	chair, join, she
7	S, Z	sir, zeal
8	n, 1	lot, not
9	r	red
10	A:	car
11	e	bed
12	I	tip
13	Q	top
14	U	book

As seis emoções básicas são definidas como mostrado na Tabela 2:

Tabela 2 – Emoções Humanas Primárias e suas características

Emoção	Descrição
Alegria	As sobrancelhas estão relaxadas. A boca está aberta e os cantos da boca direcionados para cima, na linha de segmento das orelhas.

Tristeza	A parte interior das sobrancelhas (perto do nariz) está elevada. Os olhos estão ligeiramente fechados. A boca está relaxada.
Raiva	A parte interior das sobrancelhas contrai-se simultaneamente para baixo. Os lábios pressionam-se um contra o outro ou abrem-se ligeiramente, mostrando os dentes.
Medo	As sobrancelhas elevam-se simultaneamente e as respectivas partes interiores aproximam-se. Os olhos estão tensos e em alerta.
Rejeição	As sobrancelhas e pálpebras estão relaxadas. O lábio superior eleva-se ligeiramente de modo assimétrico, fazendo uma pequena curva.
Surpresa	As sobrancelhas estão elevadas. As pálpebras superiores estão muito abertas e as inferiores relaxadas. A boca está aberta, descendo significativamente o maxilar inferior.

3.3 Algoritmo de Deformação

A deformação da malha é dada pela movimentação dos vértices que a compõem. Cada vértice possui seu movimento atrelado aos FPs que o influenciam, e estes deslocam-se através dos FAPs. Como cada FAP de alto-nível é formado por um conjunto de FAPs de baixo-nível, é possível que um determinado visema ou emoção queira deslocar um ponto característico que será deslocado por um FAP de baixo-nível. Nesse caso, o deslocamento do FP será fornecido pelo FAP de baixo-nível, pois estes possuem prioridade sobre o outro grupo de FAPs.

Durante a deformação, os deslocamentos dados pelos visemas e emoções são postos em seus conjuntos de pontos característicos. Caso um mesmo FP pertença aos dois conjuntos, o seu deslocamento será dado por (KARUNARATNE, 2006):

$$D_{FP} = 0.85 * D_V + 0.15 * D_E$$
 [Eq. 05]

Sendo D_V o deslocamento provido pelo visema e D_E o deslocamento provido pela emoção.

Em seguida, são analisados quais FAPs de baixo nível possuem intensidades diferentes de zero. Depois de descobertos, o deslocamento do FP será dado por:

$$D_{FP} = Intensidade * Valor_{APU}$$
 [Eq. 06]

O FAP fornece a direção do deslocamento do FP e a unidade de FAPU utilizada. O valor do FAPU será igual à unidade referenciada pelo FAP no modelo. É esta multiplicação pelas unidades de medida do modelo que permite que os FAPs sejam usados para qualquer modelo, pois ela torna o deslocamento dos pontos proporcional à malha a que pertencem.

Vale ressaltar que esse valor sobreporá qualquer deslocamento que o FP já possua, respeitando a prioridade que eles possuem sobre os FAPs de alto nível.

Computados os deslocamentos de cada FP, os deslocamentos dos vértices ordinários são calculados com base nos deslocamentos dos seus influenciadores. Devemos considerar os efeitos causados quando dois ou mais FPs movem um mesmo vértice simultaneamente. Seja FP_i, i = 1,2,..., N os FPs que controlam a movimentação de um ponto P da malha (WILEY, 2004).

D_i = deslocamento do FP_i

W_{i,P} = peso do FP_i em relação ao ponto P influenciado

O deslocamento D_P sofrido pelo ponto P será:

$$D_{P} = \frac{\sum_{i=0}^{N} W_{i, P} D_{i}}{\sum_{i=0}^{N} W_{i, P}}$$
 [Eq. 07]

Para cada ponto da malha, o cálculo será efetuado. Sendo conhecida a direção do deslocamento do FP, altera-se a coordenada, de cada ponto, nessa direção, somando-se seu valor atual ao deslocamento obtido.

Entretanto, durante a deformação, especificamente na aplicação de visemas e expressões, é necessário movimentar determinados FPs em direções não fornecidas pelo conjunto de FAPs. Por conta disso, em vez de movimentá-los fornecendo uma intensidade via FAP, é dado o deslocamento necessário de cada ponto característico, em todas as direções, a fim de se obter o efeito desejado. Esses deslocamentos foram obtidos através da comparação entre os vetores posição dos vértices das malhas de dois modelos com o mesmo conjunto de pontos, um com a face neutra e o outro com uma emoção ou visema que se deseja obter. Seleciona-se o mesmo conjunto de FPs nos dois modelos, nos mesmos pontos, e subtraem-se as suas respectivas posições, obtendo-se o valor do deslocamento que deve ser aplicado a cada FP quando se deseja ir da face neutra à expressão desejada. Esses valores podem ser aplicados em qualquer modelo, necessitando-se, porém, multiplicá-los pelos FAPUs de cada modelo aonde serão postos e dividir o resultado pelos FAPUs do modelo-base, ou seja, aquele que forneceu os deslocamentos necessários.

Em alguns casos, o somatório dos pesos no denominador poderá ser zero, resultando em uma indeterminação matemática, o que ocasionará o sumiço de algumas faces da malha, aparecendo buracos no modelo tridimensional. Para evitar isso, basta verificar, antes de efetuar a divisão, o valor do somatório, e se este for igual a zero, não se calcula a divisão e não se move o vértice.

Uma animação pode ser obtida aplicando todos os FAPs em intervalos de tempo constantes. Para cada quadro de uma animação, determina-se quais visema e emoção serão aplicados (se o forem), e movimentam-se, caso se queira, os FAPs de baixo nível. Entre a transição de visemas ou de emoções, se não houver uma suavização do movimento através de FAPs de baixo nível, os deslocamentos dos pontos serão grosseiros, não havendo uma passagem pelas posições intermediárias entre as posições inicial e final do movimento. Devido à dificuldade em se determinar valores de intensidade para cada FAP de baixo nível, utiliza-se um algoritmo de interpolação linear para se determinar as posições intermediárias (Coelho, 2008). Para cada quadro da execução, calculam-se os deslocamentos de todos os pontos, porém, antes de aplicá-los, é feita uma interpolação entre os valores inicial e final de cada vértice, e obtêm-se tantos valores intermediários quanto se queira. Antes da execução do quadro, há a execução dos valores intermediários, então cada ponto passará pelas posições entre o inicio e o fim do deslocamento.

4. PROTÓTIPO

Foi desenvolvido uma ferramenta que permite parametrizar o modelo 3D e animá-lo, posteriormente, através do algoritmo de deformação descrito nas seções anteriores. A seleção dos FAPUs e dos FPs é feita manualmente e são escolhidos quantos FPs se deseja, porém os cálculos da região de influência e do peso é automático. Na Figura 4, são mostradas, respectivamente, as subetapas: seleção de FAPUs, cálculo da região de influência e animação de um modelo facial.

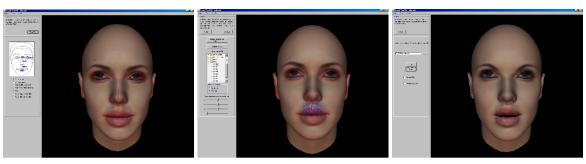


Figura 4 – Subetapas do processo de deformação

5. CONCLUSÃO

A aplicação dos dois FAPs de alto-nível simultaneamente pode resultar em expressões diferentes da imaginada. Não estabelecendo prioridades entre eles, é possível que o algoritmo proporcione imagens estranhas, como, por exemplo, alguém falando com um sorriso extremamente largo. Apesar de ser possível que ambos movimentem o mesmo FAP, os visemas, por sua função, movem apenas os que se situam em torno e na boca, enquanto as emoções deslocam os FAPs tanto da boca quanto da região dos olhos. Em virtude disso, estabeleceu-se uma prioridade do visema sobre a emoção quando ambos possuem FAPs em comum em seus conjuntos [Eq. 7], pois ainda será possível reconhecer qual a emoção aplicada, tendo em consideração que todas movimentam a região dos olhos de uma forma singular.

O algoritmo, na maioria dos testes, proporcionou deformações condizentes com as expectativas. Obteve-se, visualmente, a representação dos sentimentos humanos aliada aos visemas. A interpolação entre duas expressões humanas produziu animações realistas de fala associada à emoção. Quanto mais valores intermediários obtinham-se, mais perfeitos eram os movimentos de transição. Entretanto, a partir de 30 quadros intermediários, observa-se que a adição de mais um não melhora significativamente a qualidade da animação.

Quanto maior a complexidade do modelo 3D utilizado, maior deverá ser o poder computacional necessário para parametrizar o modelo em um tempo aceitável e animá-lo. Em virtude disso, as malhas complexas devem passar por um processo de simplificação que elimine vértices e faces que, se removidas, não causam grande perda de qualidade do modelo e diminuem o tempo gasto no cálculo da região de influência e do peso de cada ponto da malha.

A implementação do algoritmo de animação facial em dispositivos móveis pode ser feita utilizando-se vários recursos diferentes como OpenGL ES, M3G (Mobile 3D Graphics), DirectX ou Managed DirectX. O recurso a ser utilizado depende da necessidade e/ou da preferência de quem implementa. OpenGL ES e DirectX são bibliotecas de funções nativas, ou seja, que rodam diretamente no hardware gráfico. Já M3G ou Managed DirectX são bibliotecas de classes de determinadas plataformas. M3G é parte da plataforma Java e Managed DirectX é parte da .NET Compact Framework. Levando em consideração o desempenho da animação, aplicações nativas são preferíveis, já que aplicações gerenciadas, que rodam numa plataforma, têm uma camada a mais de software. Em nosso trabalho foram feitos testes utilizando OpenGL ES, M3G e Managed DirectX em dispositivos móveis e foi verificado o melhor desempenho na implementação com OpenGL ES.

REFERÊNCIAS

Coelho, J.C. **Programa para o Cálculo da Interpolação de Lagrange.** Disponível em: http://agata.ucg.br/formularios/NPI/pdf/ediiexemplo.pdf Acesso em: 09 jul 2008

Garchery, S., Egges, A., & Magnenat-Thalmann, N. Fast facial animation design for emotional virtual humans (2005). *Measuring Behaviour*. Wageningen, NL.

ISO/IEC 14496-2: Information technology – Generic coding of audio-visual objects.

Karunaratne, S. K., Yan, H. An Emotional Viseme Compiler for Facial Animation (2006).

Kshirsagar, S., Garchery, S., Magnenat-Thalmann, N. Feature Point Based Mesh Deformation Applied To Mpeg-4 Facial Animation (2000).

Lester, P. A* Pathfinding para Iniciantes (2004).

Pandzic, I. S., Forchheimer, R. MPEG-4 Facial Animation: The Standard, Implementation and Applications (2002).

Wiley, J., Thalmann, D., Magnenat-Thalmann, N. Handbook Of Virtual Humans (2004).