MC409/MO603 – Computação Gráfica

© Jorge Stolfi

Segundo Semestre de 1994

Notas de Aula – Fascículo 6

A transformação de perspectiva

6.1 Projeção cônica

6.1.1 O problema

O problema básico da síntese de imagens é produzir uma figura plana — a *imagem* — que, vista de determinado ponto, dê ao observador a ilusão de estar olhando para uma coleção de objetos tridimensionais dados — a *cena*.

A imagem pode ser apresentada ao observador de muitas maneiras: impressa em papel, pintada numa parede, mostrada num monitor gráfico, projetada numa tela de cinema, impressa num eslaide e examinada com um visor de eslaides, etc. Qualquer que seja o meio, o problema computacional é essencialmente o mesmo.

A técnica mais importante para produzir a ilusão de três dimensões com uma imagem plana — a projeção de perspectiva — foi extensamente estudada pelos pintores e arquitetos da Renascença. Séculos mais tarde, essa técnica foi mecanizada com a invenção da máquina fotográfica (e seus derivados, como a câmara de televisão), e do computador eletrônico. De novo, a geometria de uma imagem em perspectiva correta é independente da tecnologia usada na sua produção.

6.1.2 Limitações da perspectiva

A ilusão de profundidade produzida pelas técnicas de perspectiva é necessariamente limitada e imperfeita.

Em primeiro lugar, uma imagem plana pode produzir uma ilusão realista só para um observador que esteja colocado numa determinada posição em relação à imagem. Se a imagem for vista da posição errada, o objeto parecerá estar distorcido; e, quanto maior a distorção, mais difícil será conseguir a ilusão de três dimensões.

Na verdade, é teoricamente impossível produzir uma *única* imagem de um objeto que seja "realista" para um observador com visão binocular, pois pelo menos um dos olhos dele estará necessariamente fora da posição correta, e portanto verá uma imagem ligeiramente distorcida.

Mesmo que, por algum sistema de lentes e espelhos, conseguíssemos colocar os dois olhos ao mesmo tempo nessa posição ideal, a ilusão de relevo não seria perfeita. O problema é que quando olhamos para uma cena tridimensional de verdade, cada olho vê uma imagem diferente, e nosso cérebro automaticamente extrai dessas discrepâncias informações sobre a distância e forma dos objetos.

Outra limitação das imagens planas é que a lente do olho do observador tem que ajustar seu foco para distância da imagem, que em geral não corresponde á posição do objeto representado. Como o cérebro ¡também usa o foco da lente para estimar distâncias, a ilusão de relevo pode não ocorrer se esta distância for muito grande.

Por essas razões, uma imagem única só vai parecer razoavelmente realista se o objeto representado estiver relativamente longe do observador. Nesse caso, tanto a discrepância entre as imagens dos sois olhos quanto a diferença de foco entre a imagem e o objeto representado são pequenas o bastante para que a ilusão funcione.

Mesmo assim, a ilusão de relevo da projeção que pose ser conseguida é relativamente fraca. A projeção de perspectiva baseada numa única imagem (perspectiva monocular) reproduz apenas a forma aparente dos objetos, e suas dimensões e posições relativas, mas não permite julgar suas dimensões e distâncias absolutas. Por exemplo, a imagem de um cubo de 3cm de lado a 30cm de distância é idêntica à de um cubo de 2cm de lado a 20cm de distância, ou de um cubo de 3m de lado a 30m de distância. Esta ambigüidade só pode ser resolvida se a cena representada na imagem incluir objetos de tamanho conhecido.

Para uma ilusão de relevo mais perfeita, incluindo a sensação de distância absoluta, é necessário apresentar a cada olho uma imagem produzida especificamente para ser vista da posição desse olho. Este é o princípio básico da maioria dos esquemas de “terceira dimensão” usados em cinemas, fotogrametria, visores de eslaides, etc.

6.1.3 Geometria da perspectiva

A geometria básica da projeção de perspectiva é dada pela figura abaixo:

Nesta figura, o é a posição do olho do observador (a pessoa que vai contemplar a imagem), I é a imagem a ser produzida, e B é o objeto imaginário a ser representado nela. Cada ponto visível p do objeto deve ser desenhado na posição p^* da imagem, que é a intersecção do plano de I com a semireta que sai de o e passa por p.

Note que o objeto a ser representado pode estar tanto na frente da imagem, ou atrás dela, ou mesmo parte na frente e parte atrás. Como já observamos, a imagem final não permite distinguir estes três casos; e, do ponto de vista geométrico e algorítmico, tanto faz se o ponto p está na frente ou atrás do plano da imagem.

6.1.4 Coordenadas da imagem

Para descrever a imagem, precisamos estabelecer um sistema de coordenadas no plano da mesma.

Neste curso, usaremos para esse fim um *sistema de coordenadas da imagem* (SI) com eixo X apontando para a esquerda do observador, e Y apontando para cima, conforme a praxe em matemática. A origem desse sistema é por definição o ponto f do plano da imagem que está diretamente em frente ao olho do observador; ou seja, a projeção ortogonal do ponto o no plano da imagem. Na falta de nome melhor, chamaremos este ponto de *foco* da imagem.

Além dos eixos X e Y, para o calculo da projeção de perspectiva precisamos acrescentar a este sistema um eixo Z, perpendicular a ambos e apontando na direção de f para o. Assim, se d é a distância do observador à imagem, as coordenadas cartesianas do ponto o neste sistema são $(0, 0, d)$, e o plano da imagem é o plano $Z = 0$.

Na maioria das aplicações, podemos supor que o ponto f é o centro da janela I. Essa suposição é razoável, por exemplo, quando a imagem será exibida num terminal de vídeo, ocupando a tela inteira; ou quando a imagem será impressa numa revista, ou numa folha de papel avulsa.

Entretanto, em certas aplicações o foco pode estar em posição excêntrica, ou mesmo fora da janela. Por exemplo, se a imagem é um cartaz a ser pendurado na parede, a 2m do chão, o foco f (para um observador de estatura média, de pé) estará a uns 30–40cm abaixo da margem inferior da janela.

Para o cálculo da imagem, portanto, além da distância d do observador ao foco é necessário especificar a forma, o tamanho e a posição da janela I no plano X, Y, em relação ao foco f. Na grande maioria dos casos, a janela é um retângulo alinhado com os eixos, e portanto pode ser especificado por dois intervalos de coordenadas $[X_{\min} \_ X_{\max}] \times [Y_{\min} \_ Y_{\max}]$.

6.1.5 Sistema de coordenadas da cena

Na maioria dos casos, os objetos a serem representados na imagem são descritos e posicionados em termos de um sistema de coordenadas próprio, o *sistema de coordenadas da cena* (SC).

Por exemplo, a descrição de um prédio provavelmente usará eixos X e Y horizontais, alinhadas com a bússola ou com as paredes principais, e um eixo Z vertical. A origem provavelmente estará nalgum ponto do andar térreo, e as coordenadas serão dadas em metros.

Obviamente, este sistema não tem nada a ver com o sistema de coordenadas da imagem definido na seção anterior. Na verdade, se quisermos produzir várias imagens da mesma cena, vista a partir de pontos diferentes, para cada imagem teremos um sistema de coordenadas diferente.

Portanto, o primeiro passo no cálculo de uma imagem sintética é estabelecer a correspondência entre os dois sistemas de coordenadas, SC e SI.

6.1.6 A transformação de perspectiva

Há muitas maneiras de descrever a correspondência entre os dois sistemas; dependendo da aplicação, um pode ser mais conveniente do que o outro. De modo geral, é necessário especificar as coordenadas SC da origem do sistema da imagem (o ponto f), e dos vetores unitários r, s, e t paralelos aos eixos X Y e Z daquele sistema.

Esses parâmetros determinam a função do $\mathbb{R}^3$, que, dadas as coordenadas (x, y, z) de um ponto do objeto, no sistema da cena, devolve as coordenadas (x', y', z') desse mesmo ponto no sistema da imagem. Essa função consiste uma translação T_f do $\mathbb{R}^3$ que leva o ponto f para o ponto $(0, 0, 0)$, seguida de uma rotação $R_{r,s,t}$ que leva os vetores r, s, t para os eixos X, Y, Z.

Como vimos, em termos de coordenadas homogêneas a combinação dessas duas transformações pode ser descrita por uma única matriz 4×4

$$\mathsf{T}_f \mathsf{R}_{r,s,t} = \begin{bmatrix} 1 & -f_X & -f_Y & -f_Z \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & r_x & s_x & t_x \\ 0 & r_y & s_y & t_y \\ 0 & r_z & s_z & t_z \end{bmatrix} \qquad (6.1)$$

Uma vez obtidas as coordenadas do ponto p no sistema da imagem, o próximo passo é calcular sua projeção no plano da imagem, a partir da posição do observador o. Ou seja, precisamos determinar a função que, dadas as coordenadas SI do ponto p, retorna as coordenadas SI da intersecção da reta op com o plano $Z = 0$.

Lembremos que, pela definição do sistema da imagem, as coordenadas SI do ponto o são $(0, 0, d)$, onde d é a distância do observador ao plano da imagem. Uma análise geométrica elementar mostra que a projeção deve levar o ponto genérico do espaço com coordenadas cartesianas (X, Y, Z) para o ponto do plano $Z = 0$ com coordenadas cartesianas $(dX/(d - Z), dY/(d - Z))$. Ou seja, as coordendas X e Y do ponto devem ser ampliadas pelo fator $d/(d - Z)$, e a coordenada Z deve ser descartada.

Traduzindo esta fórmula para de coordenadas homogêneas, concluimos que a projeção leva a quádrupla $[w, x, y, z]$ para a tripla $[dw - z, dx, dy]$. Portanto, a projeção pode ser descrita pela matriz

$$\begin{bmatrix} d & 0 & 0 \\ 0 & d & 0 \\ 0 & 0 & d \\ -1 & 0 & 0 \end{bmatrix} \tag{6.2}$$

Como veremos mais tarde, para a eliminação de partes invisíveis da imagem é importante preservar a informação sobre a "profundidade" (distância do observador) relativa dos pontos da cena que projetam no mesmo ponto da imagem. Assim, é conveniente acrescentar uma quarta coluna à matriz acima, que preserva essa informação:

$$\mathsf{P}_d = \begin{bmatrix} d & 0 & 0 & 0 \\ 0 & d & 0 & 0 \\ 0 & 0 & d & 0 \\ -1 & 0 & 0 & d \end{bmatrix} \tag{6.3}$$

A função P_d que corresponde a esta matriz calcula não mais uma projeção no plano, mas sim uma transformação do espaço $\mathbb{T}_3$ para o espaço $\mathbb{T}_3$, que leva $[w, x, y, z]$ para $[dw - z, dx, dy, dz]$ — ou, em coordenadas cartesianas, (X, Y, Z) para $(dX/(d - Z), dY/(d - Z), dZ/(d - Z))$. Para obter as coordenadas do ponto da imagem, basta descartar a coordenada Z deste ponto — ou seja, basta projetá-lo na direção paralela ao eixo Z.

Geometricamente, a função $P(d)$ é uma transformação projetiva não afim de $\mathbb{T}^3$, que mantém os pontos do plano $Z = 0$ fixos, e leva o observador $[1, 0, 0, d]$ para o ponto no infinito $[0, 0, 0, 1]$ na direção do eixo Z. Por ser uma transformação projetiva, ela preserva colinearidade; ou seja, retas e planos do objeto continuam retas e planos depois da transformação P_d.

Em particular, depois da transformação P_d, as retas que passavam pelo ponto o tornam-se retas paralelas ao eixo Z. As retas que eram paralelas ao eixo Z tornam-se retas que convergem para o ponto $(0, 0, -d)$. (Isto corresponde à observação de que, numa imagem em perspectiva de um túnel, todas as linhas paralelas ao mesmo parecem se encontrar num ponto). Os planos paralelos à imagem continuam paralelos à mesma, mas a distâncias diferentes; o plano $Z = d$ (chamado de *plano do observador*) é levado para o plano no infinito Ω_2, e este é levado para o plano $Z = -d$.

Mais geralmente, os intervalos $(-\infty \_ 0)$, $(0 \_ d)$, e $(d \_ +\infty)$ do eixo Z são levados para $(-d \_ 0)$, $(0 \_ +\infty)$, e $(-\infty \_ -d)$, respectivamente; e o mesmo vale para todos os planos paralelos ao plano XY.

As três etapas da transformação de perspectiva podem ser compostas numa única projetividade $V_{f,r,s,t,d}$, cuja matriz é

$$\mathsf{V}_{f,r,s,t,d} = \mathsf{T}_f \mathsf{R}_{r,s,t} \mathsf{P}_d = \begin{bmatrix} 1 & -f_X & -f_Y & -f_Z \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & r_x & s_x & t_x \\ 0 & r_y & s_y & t_y \\ 0 & r_z & s_z & t_z \end{bmatrix} \begin{bmatrix} d & 0 & 0 & 0 \\ 0 & d & 0 & 0 \\ 0 & 0 & d & 0 \\ -1 & 0 & 0 & d \end{bmatrix} \tag{6.4}$$

Portanto, para calcular a imagem, basta em princípio tomar as coordenadas homogêneas SC de cada ponto da cena, multiplicá-las por esta matriz, descartar a coordenada z, e pintar o ponto resultante no plano XY da imagem, com a cor apropriada.

6.1.7 Especificação dos parâmetros da perspectiva

As fórmulas acima supõem conhecidas as coordenadas SC do ponto f e dos vetores r, s, t. Na prática, é geralmente mais fácil especificar os parâmetros indiretamente, dando-se a distância d entre o e f, e as coordenadas SC do observador o e de dois outros pontos c e u, tais que a linha co coincide com o eixo Z do sistema da imagem, e o plano uco coincide com o plano YZ desse sistema.

A idéia é que c seria um "ponto de interesse" da cena, cuja projeção na imagem final queremos que fique bem em frente ao observador — ou seja, no foco f; e u seria um ponto da cena cuja projeção na imagem, deve ficar diretamente acima da projeção de c - – ou seja, nalgum ponto do eixo Y do SI.

A partir destes dados, podemos facilmente calcular as coordenadas SC dos vetores r, s, t e do ponto f. O vetor t é, obviamente, a direção do vetor $o - c$; e o vetor s é a direção da componente vetor $u - c$ que é ortogonal a t. O vetor r é o produto vetorial de s e t, e o ponto f fica a distância d do ponto o na direção $-t$. Ou seja,

$$\begin{aligned} t &= \mathrm{dir}(o - c) \\ s &= \mathrm{dir}((u - c) - ((u - c) \cdot t)t) \\ r &= s \times t \\ f &= o - dt \end{aligned} \tag{6.5}$$

onde $\mathrm{dir}(v)$ é a direção do vetor v, isto é, $v/|v|$.

Uma outra possibilidade é especificar as coordenadas SI dos pontos o, f, e u, supondo-se implicitamente que $c = f$. Este método é especialmente prático quando queremos dar a ilusão que a distância entre observador e o objeto representado é mais ou menos igual à distância do observador à imagem; ou seja, de que a imagem é um buraco na parede, e o objeto está flutuando bem na boca do mesmo. (Esta escolha seria razoável, por exemplo, para mostrar a imagem de uma peça mecânica pequena, ou um utensílio de cozinha, na tela do computador.) Nesse caso, o "centro de interesse" c e o foco f podem ser freqüentemente colocados por "default" na própria origem do sistema da cena.

6.1.8 Um exemplo

Por exemplo, suponha que queremos produzir, a partir da descrição de um prédio, uma "fotografia sintética" do mesmo, a ser exibida numa tela do computador, tal como seria visto por um passageiro num helicóptero, voando a alguma distância do mesmo. Isto é, nosso objetivo é dar a ilusão ao usuário de que ele está no helicóptero, e que a tela é uma janela do mesmo, voltada para o prédio.

Para tornar o exemplo concreto, vamos supor que a tela é vertical e mede 30cm por 20cm, e que o usuário está sentado com o olho a 50cm de distância, bem em frente ao centro da mesma. Digamos que a posição do olho do passageiro hipotético, no sistema de coordenadas do prédio, é (40, 60, 40). Digamos também o centro de interesse c do prédio está em (0, 20, 20).

Note que, para que a imagem vista pelo usuário corresponda à a visão do passageiro, devemos supor que o plano de projeção — a janela imaginária do helicóptero que corresponde à tela — é perpendicular à linha que liga estes dois pontos, e portanto inclinada em relação á vertical.

Para o ponto u, uma escolha razoável neste caso é qualquer ponto que esteja diretamente acima do centro c, por exemplo (0, 20, 21). Esta escolha corresponde à suposição natural de que o lado inferior da janela imaginária, e a linha dos ombros do passageiro, são horizontais (paralelos ao plano XY do SC).

Aplicando as fórmulas (**??**) acima, temos

$$
\begin{aligned}
t &= \mathrm{dir}((40,\,60,\,40) - (0,\,20,\,20)) \;=\; \mathrm{dir}((40,\,40,\,20)) \;=\; (\frac{2}{3},\,\frac{2}{3},\,\frac{1}{3}) \\
s &= \mathrm{dir}((0,\,0,\,1) - ((0,\,0,\,1)\cdot t)\,t) \;=\; \mathrm{dir}((-\frac{2}{9},\,-\frac{2}{9},\,\frac{8}{9})) \;=\; (-\frac{1}{3\sqrt{2}},\,-\frac{1}{3\sqrt{2}},\,\frac{4}{3\sqrt{2}}) \\
r &= (-\frac{1}{3\sqrt{2}},\,-\frac{1}{3\sqrt{2}},\,\frac{4}{3\sqrt{2}}) \times (\frac{2}{3},\,\frac{2}{3},\,\frac{1}{3}) \;=\; (-\frac{1}{\sqrt{2}},\,\frac{1}{\sqrt{2}},\,0) \\
f &= (40,\,60,\,40) - 0.50\,(\frac{2}{3},\,\frac{2}{3},\,\frac{1}{3}) \;=\; (\frac{238}{6},\,\frac{358}{6},\,\frac{239}{6})
\end{aligned}
$$

A matriz de projeção completa é, portanto

$$
\mathsf{V}_{f,r,s,d} \;=\; \begin{bmatrix} 6 & -238 & -238 & -238 \\ 0 & 6 & 0 & 0 \\ 0 & 0 & 6 & 0 \\ 0 & 0 & 0 & 6 \end{bmatrix} \begin{bmatrix} 3\sqrt{2} & 0 & 0 & 0 \\ 0 & -3 & -1 & 2\sqrt{2} \\ 0 & 3 & -1 & 2\sqrt{2} \\ 0 & 0 & 4 & 1\sqrt{2} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -2 & 0 & 0 & 1 \end{bmatrix}
$$

6.1.9 Unidades de medida na imagem

Note que as fórmulas vistas até agora pressupõem que a mesma unidade de medida é usada nos dois sistemas de coordenadas. No exemplo dado, como as coordenadas SC são medidas em metros, as coordenadas SI que forem calculadas através da matriz acima também serão medidas em metros, a partir do foco f da tela. Portanto, um ponto só será visível na tela se, depois da transformação acima, suas coordenadas cartesianas X e Y estiverem nos intervalos $[-0.15 \;\_\; +0.5]$ e $[-0.10 \;\_\; +0.10]$, respectivamente.

Em geral, antes que possamos desenhar a imagem na tela ou outro dispositivo gráfico, precisamos converter as coordenadas SI para o sistema de coordenadas "nativo" do dispositivo de saída. Por exemplo, na maioria dos monitores gráficos cada ponto da tela é identificado por dois inteiros (h, v), sendo que o eixo h aponta para a direita, o eixo v aponta para baixo, e a origem (0.0) está no canto superior esquerdo da tela. Em alguns casos, as unidades dos dois eixos são diferentes; isto é, os píxels não são quadrados, mas sim retangulares.

A conversão entre os sistemas SI e nativo em geral consiste numa combinação de translação e mudança de escala, e pode ser calculada separadamente para cada eixo. Basicamente, queremos a função de primeiro grau $X \mapsto aX + b$ que leva uma certa coordenada SI X_0 para uma coordenada h_0 do dispositivo; e um certo X_1 para um certo h_1. Essa transformaçào é

$$X \mapsto \frac{h_1 - h_0}{X_1 - X_0} X - X_0 + h_0 \tag{6.6}$$

Note que esta fórmula funciona também quando os eixos X e h tem direções opostas; isto é, quando $X_0 < X_1$ mas $h_0 > h_1$, ou vice-versa.