

Metody reprezentacji i przetwarzania danych w trójwymiarowej grafice komputerowej czasu rzeczywistego

Przygotował: Grzegorz Oliwa

Specjalność: Inżynieria systemów informatycznych [INS]

Zagadnienia:

1. Reprezentacje danych: siatki wielokątów, woksele, opisy matematyczne. Definicje, zastosowania i terminy powiązane.
2. Przetwarzanie danych: współrzędne jednorodne, podstawowe przekształcenia (translacja, rotacja oraz skalowanie), cieniowanie wielokątów (płaskie oraz Gourouda), algorytmy oświetlenia (metoda śledzenia promieni oraz energetyczna).
3. Sprzętowe wspomaganie przetwarzania danych: na czym polega i do czego jest wykorzystywane; technologie powiązane.

1. Reprezentacje danych

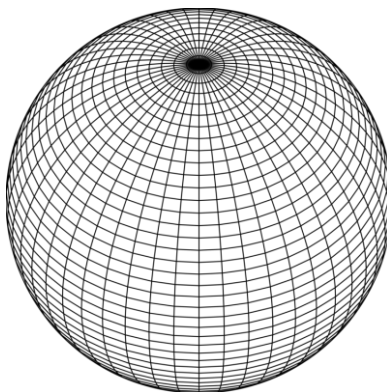
Wyróżniamy trzy podstawowe reprezentacje danych w trójwymiarowej grafice komputerowej; należą do nich:

- siatki wielokątów – najczęściej wykorzystywane do tworzenia symulatorów, gier komputerowych, filmów animowanych, bądź modeli przeznaczonych do wydruku 3D
- woksele – wykorzystywane w diagnostyce medycznej, gdzie uzyskuje się szereg przekrojów ciała pacjenta, a na ich podstawie tworzy modele trójwymiarowe
- opisy matematyczne – ich tworzenie wymaga dużej wiedzy z zakresu zarówno informatyki jak i matematyki, ale pozwalają na uzyskiwanie obiektów o wysokiej precyzji wykonania

1.1. Siatki wielokątów

Ważnymi pojęciami powiązаныmi z samymi siatkami wielokątów są:

- wierzchołek – pojedynczy punkt, zanurzony w przestrzeni trójwymiarowej
- krawędź – odcinek łączący dwa wybrane wierzchołki
- wielokąt – grupa co najmniej trzech wierzchołków, połączonych ze sobą krawędziami (tworząc cykl)



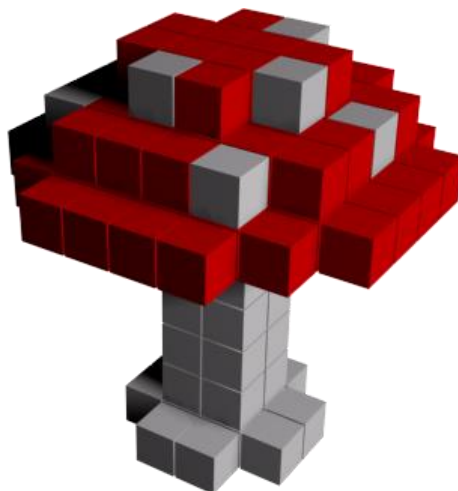
W oparciu o trzy powyższe definicje jesteśmy w stanie powiedzieć, że siatka wielokątów to nic innego jak grupa co najmniej dwóch wielokątów, połączonych ze sobą wspólnymi wierzchołkami lub krawędziami.

Istotna jest również tutaj znajomość pojęcia teselacji siatki wielokątów. Jest to dzielenie wygenerowanych podczas tworzenia obrazu trójwymiarowego wielokątów, na mniejsze wielokąty, dzięki czemu wyświetlany obiekt może być dokładniej odwzorowany. Rozsądnym rozwiązaniem jest często zastosowanie teselacji Delone, który zaproponował, aby obiekty płaskie były reprezentowane dużymi teselacjami, a nierówności małymi teselacjami. Najczęściej wykorzystywane są do tego trójkąty (mówimy wtedy o triangulacji), rzadziej czworokąty, a praktycznie w ogóle, aczkolwiek bywają takie przypadki – sześciokąty.

1.2. Woksele

Woksele to nic innego jak sześciiany (trójwymiarowe piksele) jednakowej wielkości w ramach danego obrazu trójwymiarowego, służące do modelowania określonych obiektów. W porównaniu z siatkami wielokątów:

- zachowują informację o wewnętrznej strukturze tworzonych brył – jest to ważne podczas analiz danych medycznych (informacje uzyskanych w wyniku działań różnego rodzaju urządzeń tomograficznych)
- wymagają dłuższego czasu renderowania
- nieprzydatne, gdy interesuje nas głównie wygląd, zamiast stanu obiektu

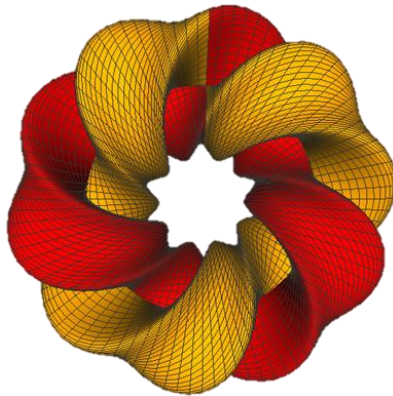


Często do grupowania wokseli w większe skupiska (celem chociażby przyspieszenia analizy czy edycji) wykorzystuje się programowanie dynamiczne.

1.3. Modele matematyczne

Modele matematyczne czyli obiekty reprezentowane przy pomocy równań. Powtórzę to, ale kluczowym elementem jest wykorzystanie ich do projektowania trójwymiarowych modeli w których szczególnie istotna jest precyzja wykonania. Przykładem z seminarium było tutaj zaprojektowanie karoserii samochodu w taki sposób, aby minimalizować wpływ oporu powietrza na jego jazdę, a co za tym idzie zminimalizowanie kosztu zużycia paliwa. W porównaniu z siatkami wielokątów oraz woksłami, najważniejszymi cechami modeli matematycznych są:

- najmniejszy koszt przechowania informacji o modelu
- bardziej skomplikowane obiekty wymagają bardzo dużej wiedzy
- bardzo wysoka precyzja produktu finalnego



Przykładowym środowiskiem służącym do manipulowania matematycznymi modelami, którego nazwy zapomniałem podać na seminarium jest K3DSurf.

2. Przetwarzanie danych

2.1. Współrzędne jednorodne i przekształcenia podstawowe

Pojęcie wprowadzone do geometrii przez Augusta Möbiusa w 1827 roku. Zauważył on, że liczbę operacji powiązanych z przekształceniami (rotacją, translacją, skalowaniem, pochyleniem) na obiektach w przestrzeni dwuwymiarowej, możemy przedstawić w inny sposób.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = A \cdot X + T = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} = \begin{bmatrix} a_{11}x + a_{12}y + t_x \\ a_{21}x + a_{22}y + t_y \end{bmatrix}$$

Macierz zawierająca skumulowane informacje na temat obrotu, skalowania oraz pochylenia

Współrzędne początkowe

Wektor translacji

Zaproponował on zanurzenie reprezentacji dwuwymiarowej w trójwymiarowej, dzięki czemu mogliśmy zredukować liczbę niezbędnych operacji do jednej – mnożenia macierzy.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = A \cdot X + T = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} = \begin{bmatrix} a_{11}x + a_{12}y + t_x \\ a_{21}x + a_{22}y + t_y \end{bmatrix}$$
$$\begin{bmatrix} x' \\ y' \\ W' \end{bmatrix} = A \cdot X = \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ W \end{bmatrix} = \begin{bmatrix} a_{11}x + a_{12}y + t_x W \\ a_{21}x + a_{22}y + t_y W \\ W \end{bmatrix}$$

To samo działanie, ale dla odwzorowania macierzy 3x3 w macierz 4x4, około 150 lat później znalazło zastosowanie w grafice komputerowej.

Macierz translacji (przesunięcia):

$$T_v \vec{p} = \begin{bmatrix} 1 & 0 & 0 & v_x \\ 0 & 1 & 0 & v_y \\ 0 & 0 & 1 & v_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_x \\ p_y \\ p_z \\ 1 \end{bmatrix} = \begin{bmatrix} p_x + v_x \\ p_y + v_y \\ p_z + v_z \\ 1 \end{bmatrix} = \vec{p} + \vec{v}$$

Macierz skalowania (dla przypadku ogólnego):

$$S_v \vec{p} = \begin{bmatrix} v_x & 0 & 0 & 0 \\ 0 & v_y & 0 & 0 \\ 0 & 0 & v_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_x \\ p_y \\ p_z \\ 1 \end{bmatrix} = \begin{bmatrix} p_x v_x \\ p_y v_y \\ p_z v_z \\ 1 \end{bmatrix}$$

Macierz rotacji wokół osi X:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \alpha & -\sin \alpha & 0 \\ 0 & \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Macierz rotacji wokół osi Y:

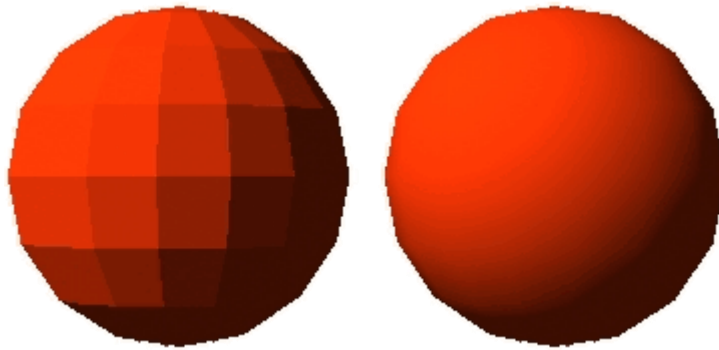
$$\begin{bmatrix} \cos \beta & 0 & \sin \beta & 0 \\ 0 & 1 & 0 & 0 \\ \sin \beta & 0 & \cos \beta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Macierz rotacji wokół osi Z:

$$\begin{bmatrix} \cos \gamma & -\sin \gamma & 0 & 0 \\ \sin \gamma & \cos \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

2.2. Cieniowanie

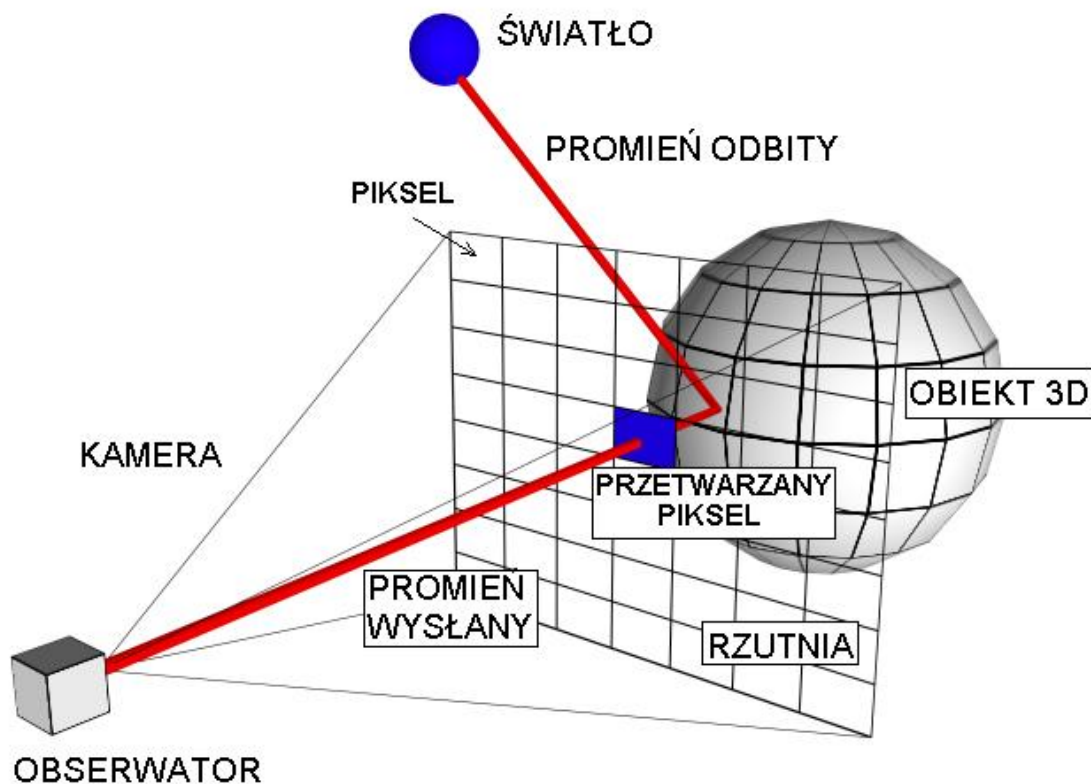
Cieniowanie to sposób obliczania barwy pikseli należących do wielokąta w oparciu o informację związaną z jego wierzchołkami. Wyróżniamy chociażby takie metody jak cieniowanie płaskie i Gourauda (czyt. /Guro/, bo ten Pan jest Francuzem jakby ktoś był ciekawy). Jest ich oczywiście więcej, np. Phonga.



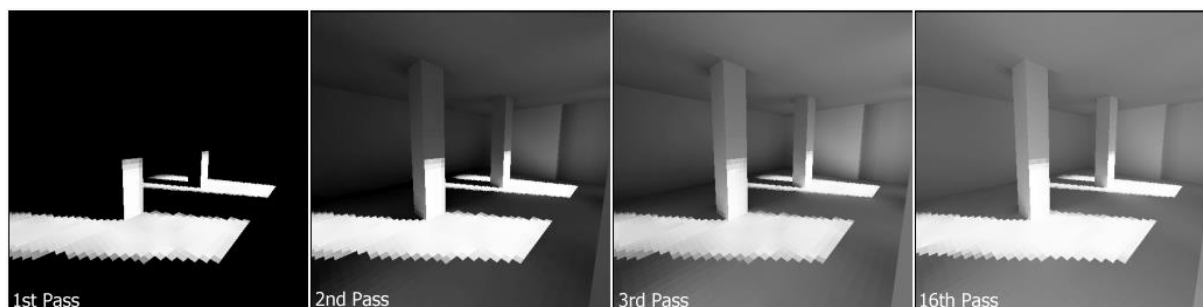
Do lewej kuli zastosowano metodę płaską – z każdego wierzchołka tworzącego wielokąt uzyskano wartość średnią, którą wypełniono samą figurę. Do prawej kuli użyto wspomnianej metody cieniowania Gourauda. Jest ona stosowana wyłącznie dla trójkątów, a więc obiekt został wcześniej poddany triangulacji. Samo wypełnienie wielokątów nastąpiło w wyniku interpolacji liniowej (istnieją alternatywne warianty, np. z interpolacją kubiczną, ale to zależy od implementacji konkretnego przypadku i nie ma na to reguły).

2.3. Metody oświetlenia

Dwa najpopularniejsze w tej chwili pojęcia, związane z generowaniem oświetlenia sceny to metoda śledzenia promieni (ang. ray-tracing) oraz metoda energetyczna (ang. radiosity). Pierwsza z nich zakłada, że przez rzutnię z kierunku obserwatora, poprowadzone zostaną promienie reprezentujące światło. Promienie te w pierwotnym założeniu przechodzą przez każdy punkt (piksel) ekranu, tak więc w przypadku rozdzielczości 1920x1080 mówimy o ponad 2 milionach promieni. Następnie wyszukiwane są punkty przecięcia tychże promieni z obiektami sceny, a na samym końcu wyznaczona zostaje jasność dla każdego takiego punktu. Metoda ta jest niesamowicie kosztowna obliczeniowa, dlatego też powstały jej wersje alternatywne. Wśród nich możemy wyróżnić chociażby forward ray-tracing, polegający na doprowadzeniu promieni świetlnych do obserwatora, a więc zakłada on odwrócenie koncepcji algorytmu.

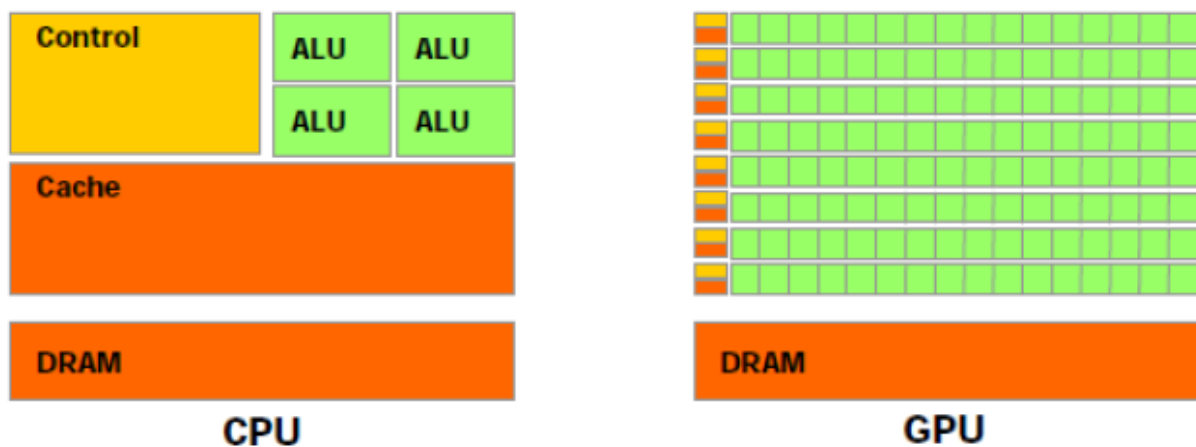


Metoda energetyczna polega na wyliczeniu współczynnika pochłaniania światła dla każdego obiektu. Te jednak muszą zostać poddane wcześniejszej teselacji trójkątnej bądź czworokątnej (częściej jednak stosuje się tą pierwszą, czyli triangulację). W celu wyznaczenia wspomnianego wcześniej współczynnika, należy przyporządkować wielokątom – dalej nazywanymi dla tej konkretnej metody płacami – ich początkowe wartości liczbowe (z pewnym przybliżeniem; niedokładnie). Dodatkowa wartość oznacza pochłanianie światła, przy czym im jest ona większa tym więcej światła mu podlega. Ujemna wartość natomiast oznacza jego emisję. Następnym i ostatnim krokiem algorytmu jest wyznaczenie współczynnika sprzężenia optycznego dla każdej pary płatów. Jest to nic innego jak iteracyjne obliczanie jasności tychże płatów przy uwzględnieniu sąsiadów. Oczywiście jest tutaj, że im więcej płatów powstało na etapie teselacji, tym więcej obliczeń będziemy musieli algorytm wykonać.



3. Sprzętowe wspomaganie przetwarzania danych

GPU przeznacza więcej tranzystorów na przetwarzanie danych niż na sterowanie i pamięć cache, dzięki czemu znacznie lepiej nadaje się do zrównoleglania zadań, a także wykonywania wielu operacji macierzowych. Dzięki wprowadzeniu omówionych wcześniej współrzędnych jednorodnych, zadanie to zostało dodatkowo uproszczone, ponieważ karty graficzne zobligowane są w gruncie rzeczy do wykonywania prostych operacji mnożenia. Trzeba jednak podkreślić, że jest to prawdą dla większości, ale nie wszystkich operacji elementarnych.



Do technologii umożliwiających wykonywanie różnorodnych algorytmów przy pomocy kart graficznych, należą chociażby CUDA (ang. Compute Unified Device Architecture) czy Vulkan. Jedno i drugie to w zasadzie uniwersalna architektura procesorów wielordzeniowych, umożliwiająca wykorzystanie ich mocy obliczeniowej do problemów takich jak chociażby znany wszystkim TSP (ang. Travelling Salesman Problem). W zależności od implementacji algorytmu, a także samej karty graficznej, możemy mówić o niewielkim lub ogromnym zwiększeniu wydajności.