

ПРОГРАММНЫЕ СРЕДСТВА ПОСТРОЕНИЯ ПРОСТРАНСТВЕННОЙ МОДЕЛИ ДЛЯ СИСТЕМЫ МАШИННОГО ЗРЕНИЯ

Иодо М. А.

Кафедра интеллектуальных информационных технологий, Белорусский государственный университет
информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: mikhailiodo@gmail.com

В работе предлагается подход к реализации построения пространственной модели окружения на основе параллакса, то есть изменения видимого положения объектов относительно друг друга в зависимости от положения камеры. Для применения данного подхода необходимо решить две проблемы – это отслеживание изменения положения камеры и отслеживание перемещения объектов на различных кадрах.

ВВЕДЕНИЕ

В настоящее время машинное зрение получило широкое распространение. Самый простой пример – станки с числовым программным управлением, в которых машинное зрение предполагает получение визуальной информации об объекте и её преобразование для формирования команд, управляющих исполнительными механизмами. Первые опыты по внедрению машинного зрения относятся к 70-м годам XX века, однако сложность заключалась в отсутствии оборудования с достаточными вычислительными мощностями. Кроме того, видеооборудование было такими громоздкими, что это сводило на нет всю целесообразность машинного зрения. В настоящее время практически любой мобильный телефон может использоваться как устройство для систем машинного зрения, что и происходит, постоянно появляются системы распознавания различных образов и дополненной реальности. Несмотря на активное развитие в данной области, построение трехмерных моделей с помощью входного видео сигнала применяются редко, вместо этого моделирование происходит вручную в специализированных программных пакетах, таких как 3ds Max, Blender и прочие.

I. ПОСТАНОВКА ЗАДАЧИ

Исходными данными являются видео, т.е. последовательность изображений, и последовательность положений камеры в пространстве. Положение камеры представлено в виде матрицы преобразования, которая аналогична той, что используется в компьютерной графике, и включает в себя изменение местоположения камеры относительно начала системы координат, в которой будет представлена модель окружения, изменение ориентации в пространстве и параметры камеры, такие как угол обзора. Данная матрица получается в результате перемножения матриц перемещения, вращения и перспективной проекции. Необходимо определить трехмерную модель окружения в векторном или воксельном виде.

II. ПРЕДСТАВЛЕНИЕ ДАННЫХ

Видео есть последовательность кадров, т.е. последовательность изображений. Цвет не имеет значения, поэтому все кадры черно-белые. Одно изображение представлено в виде матрицы пикселей, где один пиксель представлен в виде дробного числа в диапазоне от нуля до единицы, где ноль – это черный цвет, а единица – белый.

Пусть имеется L входных кадров, причем n – количество пикселей в одном кадре. Тогда совокупность входных изображений можно представить в виде следующей матрицы:

$$Y = \begin{pmatrix} y_1^1 & y_2^1 & \dots & y_n^1 \\ y_1^2 & y_2^2 & \dots & y_n^2 \\ \dots & \dots & \dots & \dots \\ y_1^L & y_2^L & \dots & y_n^L \end{pmatrix},$$

где $Y^L = (y_1^L y_2^L \dots y_n^L)$ соответствует L -му входному образу.

Положение камеры, как отмечалось ранее, можно представить в виде матрицы преобразования, которые используются в компьютерной графике. Такие матрицы представляются в двумерном виде размерностью 4×4 , т.к. предназначены для работы с однородной системой координат, используемой в проективной геометрии. Однородные координаты обладают тем свойством, что определяемый ими объект не меняется при умножении всех координат на одно и то же ненулевое число. Из-за этого количество координат, необходимое для представления точек, всегда на одну больше, чем размерность пространства, в котором эти координаты используются, т.е. для представления точки в трехмерном пространстве используется четырехмерный вектор. Для простоты подобную матрицу можно представить в одномерном виде размерностью 1×16 , тогда всю последовательность положений камеры можно представить в виде следующей матрицы:

$$X = \begin{pmatrix} x_1^1 & x_2^1 & \dots & x_{16}^1 \\ x_1^2 & x_2^2 & \dots & x_{16}^2 \\ \dots & \dots & \dots & \dots \\ x_1^L & x_2^L & \dots & x_{16}^L \end{pmatrix},$$

Количество положений камеры равно количеству кадров, т.е. каждому кадру соответствует одно положение камеры в пространстве.

III. АППАРАТНАЯ РЕАЛИЗАЦИЯ

Первоначальным этапом построения модели окружения является получение входных данных, т.е. видео сигнала и последовательности положения камеры. Для этого необходимо устройство оснащенное камерой, акселерометром и гироскопом, как практически любой современный мобильный телефон. Показания гироскопа и акселерометра используются для вычисления изменения положения камеры и генерации последовательности положений камеры, данная задача уже имеет множество решений в виде готовых библиотек из-за развития направлений виртуальной реальности и дополненной реальности.

IV. ПРИНЦИП ПОСТРОЕНИЯ ТРЕХМЕРНОЙ МОДЕЛИ ОБЪЕКТОВ

Предлагаемый принцип построения трехмерной модели объектов окружения основан на предположениях о модели окружения и проверках предположений. Данный подход в некоторой степени схож с принципом обучения нейронной сети с учителем, т.е. имеется множество входных данных и множество соответствующих им выходных данных, процесс обучения заключается в уменьшении разности между выходными значениями нейронной сети и известными правильными значениями. Входными данными в этом случае являются матрицы преобразований, т.е. матрицы камеры, а выходными изображениями которые сравниваются с кадрами видео. Отличие заключается в том, что в процессе обучения нейронной сети подбирается матрица весов и пороговых значений, а в процессе построения модели окружения подбирается множество вершин объектов. При правильно построенной трехмерной модели выходное изображение рендера будет верно для любого значения матрицы преобразований. Общую схему можно представить в следующем виде:



Рис. 1 – Схема рендеризации

Таким образом построение модели окружения заключается в предположении о 3D-объектах окружения, результатом которого является 3D-сцена, проверке этой сцены на соответствие действительности, внесении изменений в сцену до тех пор пока выходные значения рендеринга не будут схожи с кадрами видео.

Предположения об объектах сцены строятся на выявлении контуров и фигур на кадрах видео, например бинаризацией изображения по порогу, меняя порог можно выявлять различные объекты на одном изображении:



Рис. 2 – Выявление фигур на изображении

Сравнивать выходные изображения рендера и кадры видео также стоит в рамках выявленных фигур и контуров например методами Робертса, Превитта и Собеля [1–3].

1. Р. Гонсалес, Р. Вудс Цифровая обработка изображений – М: Техносфера, 2005 – 1007с.
2. Кудрявцев Л.В. Краткий курс математического анализа – М.: Наука, 1989 – 736с.
3. Анисимов Б.В. Распознавание и цифровая обработка изображений – М.: Вышп. школа, 1983 – 295с.
4. Kolmogorov V. and Zabih R. Computing visual correspondence with occlusions using graph cuts / V. Kolmogorov, R. Zabih // Proc. Int. Conf. Computer Vision, Vancouver, BC, Canada, – 2001. – P. 508–515.
5. Greig D., Porteous B., and Seheult A. Exact maximum a posteriorestimation for binary images / D. Greig, B. Porteous, J. Roy // Statist. Soc. – 1989. – Vol. 51, № 2. – P. 271–279.