

УДК 004

Отслеживание положения камеры системы компьютерного зрения в трёхмерном пространстве в режиме реального ВРЕМЕНИ

09, сентябрь 2012

Афиногенов Е.И.

*Научный руководитель: к.т.н, доцент Волосатова Т.М.
Кафедра «САПР» (РК6) МГТУ им. Н.Э.Баумана, Москва, Россия*

МГТУ им. Н.Э. Баумана
bauman@bmstu.ru

Введение

Компьютерное зрение – одна из самых востребованных областей на современном этапе развития информационных технологий. Оно требуется на производстве, при управлении роботами, при автоматизации процессов, в медицинских и военных приложениях [1].

Одной из важнейших задач в системе компьютерного зрения является задача распознавания объектов. Особенно остро данный вопрос стоит в роботизированных системах, осуществляющих манипуляцию телами в пространстве, когда требуется определить не просто принадлежность объекта к какому-либо классу, а установить его позицию в трехмерной сцене по отношению к роботу и другим телам.

В связи с развитием доступных вычислительных мощностей актуальным становится вопрос создания систем распознавания трехмерных объектов, функционирующих в режиме реального времени. В данной работе рассматривается задача 3D трекинга или, другими словами, отслеживание положения камеры в трехмерном пространстве в режиме реального времени как важнейшая составляющая системы распознавания 3D объектов в динамике.

Постановка задачи

Пусть имеется некоторая мобильная роботизированная установка, снабженная видеокамерой, обеспечивающей зрение системы. Если не учитывать ограничения, связанные с конкретной системой, то камера обладает 6-ю степенями свободы. Необходимо определить положение камеры при её движении в пространстве относительно стационарных объектов сцены (в системе координат, связанной с неподвижными объектами сцены). Система не предполагает использование каких-либо маркеров (зафиксированных датчиков положения), опираясь главным образом на пассивное зрение [1]. В данной работе не рассматривается случай наличия в сцене отдельных движущихся объектов.

На следующем рисунке представлена упрощенная схема работы предлагаемого 3D трекера, отслеживающего положение камеры в пространстве на основе видеопоследовательности, поступающей на вход системы.

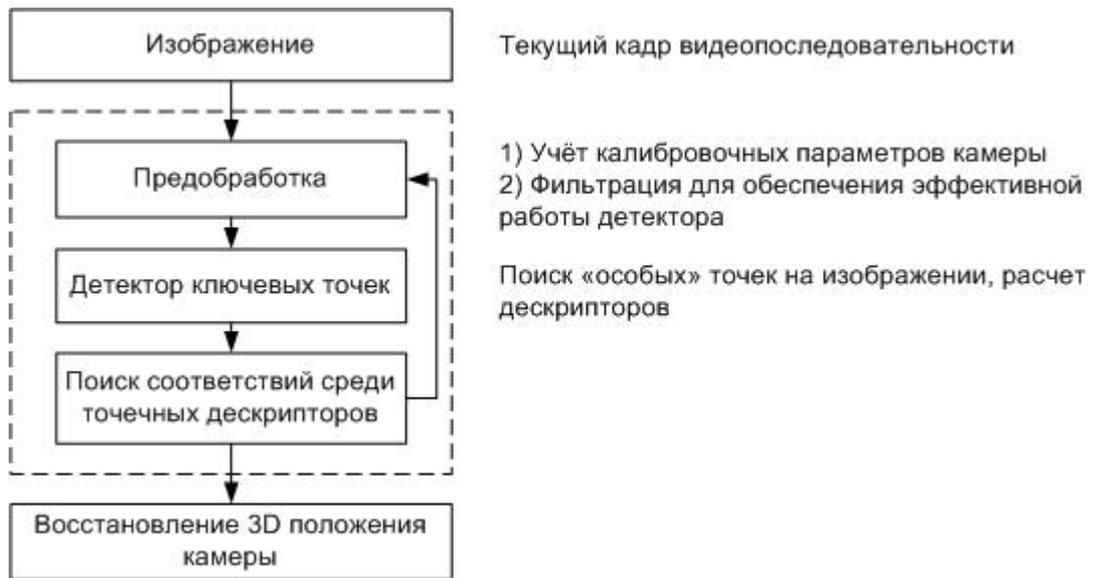


Рис. 1 Обобщенный алгоритм «компьютерного зрения»

Поступающее на вход системы изображение предварительно обрабатывается: переводится в градацию серого, нормализуются интенсивности пикселей. Это необходимо для того, чтобы полученные далее дескрипторы соответствующих точек имели высокую степень схожести независимо от изменения освещенности сцены.

Для обеспечения устойчивой работы системы при наличии смаза, появляющегося в результате движения камеры, используются пирамиды изображений с масштабным коэффициентом 2. Поиск особенностей осуществляется на каждом уровне пирамиды.

Далее изображение поступает на вход детектора ключевых точек. В настоящее время разработано огромное количество методов поиска особенностей. Наиболее известными являются уголкового детектор Харриса [4], SIFT (Scale-Invariant Feature Transform) [5], SURF (Speed-Up Robust Features) [6], Ferns [7]. Однако данные детекторы имеют высокую вычислительную сложность и не могут быть эффективно использованы для решения поставленной задачи.

Наиболее применимым в данном случае видится детектор FAST (Feature-Accelerated Segment Test) [8], осуществляющий поиск ключевых точек изображения. Он классифицирует точку p (рис. 2) как «уголок», если существует ряд смежных пикселей на окружности, интенсивности которых больше $I_p + t$ или интенсивности всех меньше $I_p - t$, где I_p – интенсивность точки p , а t – пороговая величина.

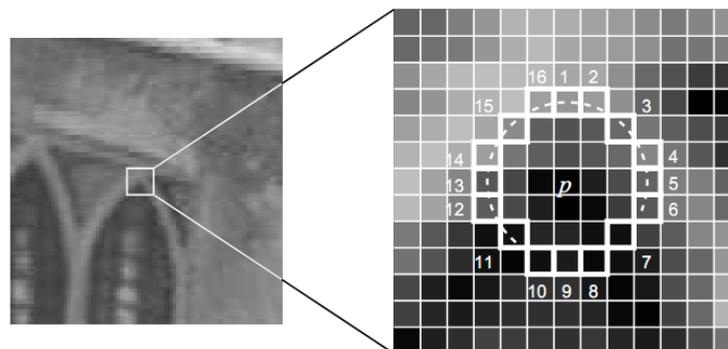


Рис. 2 Поиск ключевых точек

Далее для найденных на текущем кадре видеопоследовательности «особых» точек осуществляется поиск соответствующих точек на предыдущем кадре путем сравнения их дескрипторов. FAST-дескриптор представляет собой нормализованную гистограмму интенсивностей для 64 пикселей, взятых разреженно вокруг целевой точки (рис. 3а).

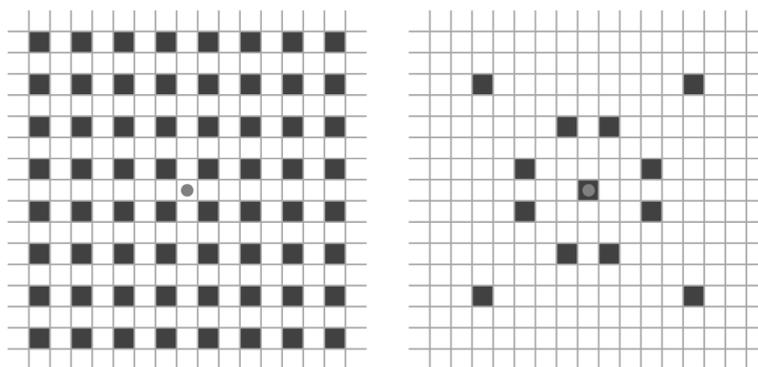


Рис. 3 Пиксели, используемые в дескрипторе FAST (а);
Пиксели, используемые для формирования индекса поиска (б).

Дескриптор ключевой точки можно записать следующим образом:

$$\begin{array}{ccccc}
 D_{0,0} & D_{0,1} & D_{0,2} & D_{0,3} & D_{0,4} \\
 D_{1,0} & D_{1,1} & D_{1,2} & D_{1,3} & D_{1,4} \\
 \vdots & \vdots & \vdots & \vdots & \vdots \\
 D_{63,0} & D_{63,1} & D_{63,2} & D_{63,3} & D_{63,4}
 \end{array}$$

Здесь D_i – нормализованная гистограмма для точки i ($i = 0 \dots 63$), занимающая всего 5 бит ($D_{i,0} \dots D_{i,4}$). Таким образом, чтобы определить степень схожести двух ключевых точек, необходимо вычислить количество выставленных бит в 64-разрядном числе, получающемся путем побитового умножения соответствующих столбцов дескрипторов с последующим побитовым сложением.

Для того, чтобы ускорить поиск соответствий, применяется механизм индексации, при котором для каждой точки рассчитывается индекс поиска на основе интенсивностей 13 пикселей (рис. 3б), причем значение каждого пикселя кодируется одним битом. Таким образом, таблица индексов будет иметь размер 8192.

После определения достаточного количества точек в сцене, повторяющихся в нескольких последовательных кадрах, осуществляется восстановление трёхмерного положения камеры. Для решения данной задачи используется устойчивая схема факторизации на основе разложения по сингулярным значениям (SVD – Singular Value Decomposition) [2, 3].

Заключение

В настоящее время существует множество систем, позволяющих по видеоряду восстанавливать трехмерное положение камеры [9]. Данные системы получили широкое распространение в кинематографии для создания различных спецэффектов. Однако они не предусматривают использования в режиме реального времени, часто требуют наличия в сцене определенных маркеров.

В данной работе рассматривается 3D трекер, вычисляющий положение камеры в пространстве в режиме реального времени. Предлагаемая схема определения ключевых точек на изображении с ускоренным поиском соответствий занимает незначительное процессорное время, позволяя использовать совместно с системой распознавания трёхмерных объектов в динамике.

Литература

1. Форсайт Д., Понс Ж. Компьютерное зрение. Современный подход. : Пер. с англ. – М.: Издательский дом «Вильямс», 2004. – 928 с.
2. Tomasi C., Kanade T. Shape and Motion from Image Streams: a Factorization Method Full Report on the Orthographic Case, 1992
3. Афиногенов Е.И. Восстановление аффинной структуры сцены по движению.-М.: МГТУ им. Н.Э.Баумана. Тезисы доклада 13-й ММ НТК «Наукоемкие технологии и интеллектуальные системы», 2011, с.125-128.
4. C. Harris and M.J. Stephens. A combined corner and edge detector. In Alvey Vision Conference, pages 147-152, 1988.
5. David G. Lowe. Object recognition from local scale-invariant features. International Conference on Computer Vision, pages 1150-1157, 1999.
6. H. Bay, A. Ess, T. Tuytelaars, Luc Van Gool “SURF: Speeded Up Robust Features”, Computer Vision and Image Understanding, Vol. 110, No. 3, pp. 346-359, 2008.
7. V. Lepetit and P. Fua. Keypoint recognition using randomized trees. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(9): 1465-1479, Sept. 2006.
8. E. Rosten, R. Porter, T. Drummond. Faster and better: a machine learning approach to corner detection. 2008.
9. S. Mirpour. A Comparison of 3D Camera Tracking Software. 2008.