

# Реконструкция траектории движения и формы объекта с использованием инфракрасных маркеров и стереопары

Раимиз Зейналов, Антон Конушин

Факультет вычислительной математики и кибернетики

Московский государственный университет имени М.В.Ломоносова, Москва, Россия

{rzeynalov, ktosh}@graphics.cs.msu.ru

## Аннотация

В предыдущей работе был описан метод отслеживания объекта в пространстве с использованием инфракрасных маркеров. В этой работе описываются улучшения, сделанные в алгоритмах отслеживания объекта, а также описываются алгоритмы реконструкции формы движущегося объекта. Кроме того, в предыдущей работе было отмечено, что существует проблема наложения маркеров (“marker swapping”). В этой работе приводится обоснование отсутствия этой проблемы в рамках решаемой задачи.

**Keywords:** *отслеживание, сопоставление проекций, реконструкция геометрии, инфракрасные маркеры, эпиполярные ограничения.*

## 1. ВВЕДЕНИЕ

Иногда возникают практические задачи, связанные с необходимостью отслеживания перемещения объекта в пространстве. При этом не всегда имеется возможность ориентироваться на текстуру объекта – например, объект может иметь однородную окраску или его материал может быть глянцевым, что приводит к тому, что методы, основанные на текстуре объекта (например, методы оптического потока [5, 7, 9] и использование дескрипторов особых точек [1]), становятся неприменимыми. В таких случаях нередко используют контрастные маркеры, которые устанавливаются на объект исследований. В данной работе речь идёт об использовании инфракрасных маркеров (диодов, которые излучают в инфракрасном диапазоне) совместно с использованием стереопары, которая снимает в инфракрасном диапазоне. Таким образом, в качестве исходных данных (после распознавания маркеров, которое тривиально) выступают наборы проекций точек для каждой камеры во все моменты времени.

Для построения траекторий отдельных точек необходимо знать межкадровые соответствия проекций точек. Изначально этой информации нет, т.к. в исходных данных на каждом кадре имеется множество проекций маркеров без уникальных идентификаторов, которые позволили бы различать проекции точек между собой. Подробнее о решении задач, связанных с отслеживанием проекций точек, написано в секции 4.

Проекция точек, по сути, задают точечные особенности объекта. Однако здесь имеется принципиальная проблема – так как маркеры между собой не различимы, не различимы и их проекции – их изображения, которые снимаются камерами. А значит, неизвестны и межвидовые соответствия маркеров. Поэтому перед выполнением триангуляции [2] эти соответствия нужно каким-либо образом восстановить. В этой работе в секции 5

рассказывается об алгоритмах, которые позволяют восстановить эти соответствия.

Когда уже получены межвидовые соответствия точек, можно выполнить триангуляцию. Это позволит восстановить взаимное расположение точек объекта во все моменты времени. Однако ни в один момент времени обе камеры не наблюдают сразу все маркеры объекта. Поэтому на каждом кадре можно восстановить лишь отдельные фрагменты всего объекта. В этой работе рассказывается об алгоритмах, которые позволяют восстановить геометрию всего объекта. Этому посвящена секция 6.

Для апробации построенных алгоритмов был произведён ряд измерительных экспериментов с целью доказательства состоятельности предложенных алгоритмов, а также оценки точности производимых измерений. Проводились как статические эксперименты, основанные на измерении расстояний, так и динамические, основанные на измерении динамических характеристик движения объектов. Описания экспериментов и их результатов представлены в секции 7.

Заключение и будущие планы развития алгоритмов описаны в секции 8.

## 2. ОБЗОР СУЩЕСТВУЮЩИХ РЕШЕНИЙ

Для отслеживания движущихся объектов на практике часто используются дескрипторы особых точек, такие как SIFT и SURF [1], которые позволяют найти особые точки на объекте [2, 9]. Для отслеживания небольшого количества объектов на статической сцене иногда используется оптический поток для поиска движущихся объектов [5, 7, 9]. Для этого берутся соседние кадры и вычисляется их разность. Те регионы, в которых эта разность больше определённого порога, считаются регионами движения, то есть регионами, в которых движется объект интереса. Просматривая попарно все кадры последовательности, можно получить проекции траекторий движения объекта на камеры. В нашей задаче оптический поток бесполезен, так как при его использовании не будут найдены межвидовые соответствия маркеров.

В некоторых задачах трекинга требуется отслеживать перемещение объекта заранее известной формы [6]. В таких случаях нередко этот объект помечен маркерами, что позволяет эффективнее отслеживать его на изображениях. Когда форма отслеживаемого объекта известна, можно использовать свойства, связанные с расположением маркеров, и легче отличать один маркер от другого. Это т.н. техника Motion Capture [8, 10].

Такие подходы не пригодны для решения поставленной задачи, потому что форма отслеживаемого объекта заранее неизвестна и должна быть получена в процессе решения задачи.

### 3. ПРЕДЛОЖЕННОЕ РЕШЕНИЕ

Предложенное решение задачи состоит из набора алгоритмов, каждый из которых решает отдельную подзадачу. Для получения результата рассматриваемые алгоритмы последовательно применяются к исходным данным.

Рассматриваемая система состоит из стереопары (двух синхронизированных камер, которые снимают в инфракрасном диапазоне) и отслеживаемого объекта, помеченного инфракрасными маркерами (рис. 1).

Итак, в качестве исходных данных всей задачи выступают:

1. Калибровка стереопары
2. Стереопоследовательность кадров

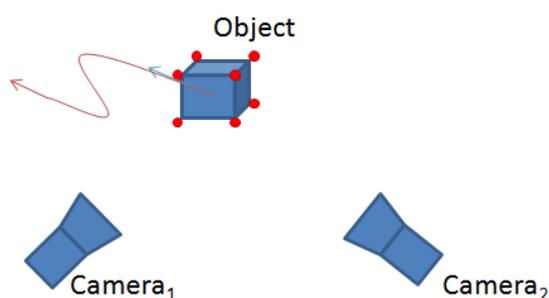


Рисунок 1: Объект и стереопара.

Под калибровкой стереопары понимаются внутренние параметры камер (матрицы внутренних калибровок и наборы коэффициентов дисторсий линз), положения и ориентации в пространстве в некоторой зафиксированной системе координат. Следует отметить, что камеры снимают кадры синхронно, то есть моменты времени съёмки очередного кадра  $t_i$  можно считать одинаковыми, что обеспечивается оборудованием, управляющим камерами.

Под стереопоследовательностью кадров понимаются две синхронизированные последовательности кадров с камер. Каждый кадр содержит набор изображений маркеров и, возможно, шум. Так как маркеры представляют собой инфракрасные диоды, а камеры снимают в инфракрасном диапазоне, кадр, как правило, представляет собой изображение с белыми пятнами на чёрном фоне (рис 2).



Рисунок 2: Пример кадра.

Сформулируем решаемую задачу. Итак, дано:

- Калибровка стереопары (внутренние параметры камер  $K_c$ , дисторсии линз  $D_c$ , их положения  $T_c$  и ориентации  $R_c$ ):

$$C_c = \langle K_c; D_c; R_c; T_c \rangle, \quad c = 1, 2 \quad (1)$$

- Две последовательности кадров (по последовательности на камеру), состоящих из множества проекций точек ( $c$  – номер камеры,  $n_i^c$  – количество проекций  $q^{c,i}$  на кадре  $i$  камеры  $c$ ,  $l_{i,j}$  – номер  $j$ -ой точки на  $i$ -ом кадре,  $F_i^c$  – кадр (2),  $S^c$  – последовательность кадров (3),  $N$  – количество кадров):

$$F_i^c = \{p_{l_{i,j}}^{c,i}, j = 1..n_i^c\} \quad (2)$$

$$S^c = \{F_i^c\}, \quad i = 1..N, \quad c = 1..2 \quad (3)$$

Требуется:

- Определить межкадровые соответствия точек (4) ( $M^c$  – количество проекций точек, соответствующих разным точкам в пространстве,  $p^{c,i,j}$  – проекции точек, упорядоченные согласно межкадровым соответствиям,  $l(i,j,c)$  – номер проекции  $q$ , соответствующей точке  $j$  на кадре  $i$  для камеры  $c$ )

$$C_F^c : q_{l(i,j,c)}^{c,i} \rightarrow p_j^{c,i} \quad (4)$$

- Определить межвидовые соответствия проекций точек (5) (множество пар индексов  $j_1$  и  $j_2$ , для которых отношение эквивалентности, определяющее соответствие всех проекций одной камеры для одной точки всем проекциям одной точки на другой камере):

$$C_V^{1,2} : p_{j_1}^{1,i} \times p_{j_2}^{2,i} \rightarrow \{True, False\} \quad (5)$$

Задача поиска и распознавания маркеров на подобных изображениях достаточно тривиальна, поэтому не будем на ней подробно останавливаться.

### 4. ОТСЛЕЖИВАНИЕ ПРОЕКЦИЙ

#### 4.1 Модель маркера

Прежде чем описывать алгоритм получения межкадровых соответствий для проекций инфракрасных маркеров рассмотрим модель маркера (рис. 3).

Маркер представляет собой диод шириной  $d$ , высотой  $h$  и с углом рассеивания  $\alpha$  (угол между противоположными боковыми гранями пирамиды, описывающей поток излучаемой диодом энергии).

Проекции маркеров, расположенных на отслеживаемом объекте, теоретически могут накладываться друг на друга на изображениях камер, что может привести к проблемам при отслеживании отдельных маркеров и к ошибкам на этапе поиска межкадровых соответствий маркеров (т.н. проблема marker swapping). На самом деле, можно убедиться в том, что такие ситуации практически невероятны при соблюдении некоторых условий. Для этого необходимо рассмотреть модель маркера (рис. 3), которая адекватно описывает используемые инфракрасные маркеры – светоизлучающие диоды, которые излучают значительную часть энергии в инфракрасную область электромагнитного спектра. Итак, диод излучает свет не во всех направлениях, а на ограниченный угол. Кроме того, если располагать маркеры на значительном расстоянии друг от друга (намного большем, чем размер диода), тогда, в силу ограниченности угла излучения, наложение пятен от маркеров будет невозможно.

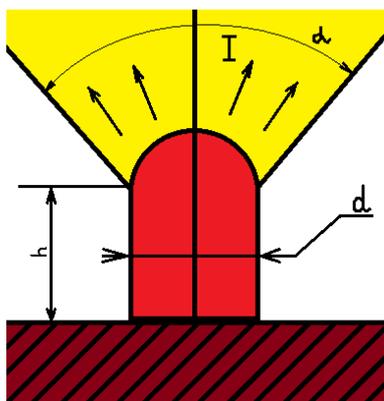


Рисунок 3: Модель маркера.

## 4.2 Алгоритм

На входе дана последовательность кадров (3), которые содержат наборы проекций точек – координат на изображении. Для одной и той же точки в пространстве нет соответствия между её проекциями на разных кадрах, в то время как для оценки траектории эти соответствия необходимы. Таким образом, требуется восстановить эти соответствия.

Проведённые эксперименты показывают, что для получения межкадровых соответствий достаточно использования простого алгоритма, основанного на межкадровом смещении. Ситуации, в которых алгоритм может не сработать, на практике встречаются крайне редко, особенно при малом количестве маркеров.

Кроме того, в силу ограниченности угла излучения маркера и большого расстояния между ними ситуация, когда один маркер накладывается на другой, практически невозможна. В противном случае простой алгоритм, основанный на межкадровом смещении, мог бы давать неверные результаты и нуждался бы в некотором усложнении.

Введём порог, характеризующий величину допустимого смещения проекции точки на изображении между соседними кадрами. Рассмотрим первый кадр последовательности. Каждой проекции, встречающейся на этом кадре, ставим в соответствие уникальный идентификатор – натуральное число (рис. 4а, 4в). Точки  $P_1$ ,  $P_2$  и  $P_3$  получают идентификаторы 1, 2 и 3 соответственно. Затем обрабатываем каждый следующий кадр одним и тем же способом: для каждой проекции на предыдущем кадре ищем ближайшую непомяченную проекцию на текущем кадре. Если расстояние между ними (в плоскости изображения) меньше порога – помечаем проекцию на новом кадре тем же идентификатором, что и на предыдущем кадре.

Каждую непомяченную проекцию на новом кадре помечаем новым уникальным идентификатором (рис. 4б). Если в какой-то момент для проекции на предыдущем кадре не нашлось проекции на текущем кадре, значит, эта точка только что появилась (рис. 4б) – точка  $P_4$ . Если для какой-то точки на предыдущем кадре не удалось найти точку на текущем кадре, значит, эта точка исчезла (рис. 4в, 4г) – точка  $P_2$ . Как было отмечено выше, такой алгоритм работает очень быстро, а ситуации, в которых результат работы может быть неверным, крайне маловероятны.

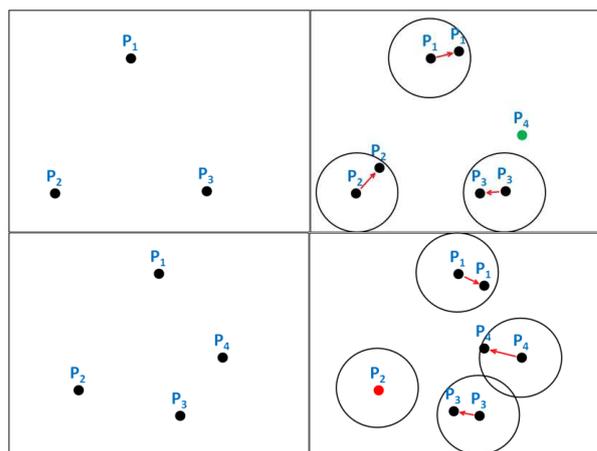


Рисунок 4: Межкадровое сопоставление (а, б, в, г).

## 5. МЕЖВИДОВОЕ СОПОСТАВЛЕНИЕ

Алгоритм делится на два этапа – вычисление возможных межвидовых соответствий для одной пары кадров и вычисление межвидовых соответствий для всей последовательности пар кадров.

### 5.1 Локальное сопоставление

Возможные межвидовые соответствия для кадра вычисляются следующим образом: для каждой точки  $p_j$  на каждой камере выбираются все точки  $\{q_i\}$  на другой камере в окрестности соответствующей эпиполярной линии. Все пары вида  $\langle p_j, q_i \rangle$  добавляются в список возможных соответствий для данной пары кадров. Указанные действия выполняются для всех пар кадров.

$$q^T F p = 0 \quad (6)$$

$$\begin{cases} l_q = F p \\ l_p = q^T F \end{cases} \quad (7)$$

Для вычисления межвидовых соответствий используются эпиполярные ограничения (6) [3], которые заключаются в том, что для точки в пространстве её проекция  $p$  на одной камере ограничивает положение проекции  $q$  той же самой точки на другой камере эпиполярной линией  $l_q$ , которая задаётся фундаментальной матрицей  $F$ , и наоборот (7).

### 5.2 Глобальное сопоставление

Из-за ошибок на отдельных кадрах могут получиться противоречащие между собой множества возможных соответствий. Для того чтобы их избежать, используется схема голосования. Пусть всего разных точек в последовательности кадров одной камеры  $M_1$ , второй камеры –  $M_2$ . Введём матрицу  $MC$ , заполненную нулями. Для каждого соответствия  $\langle p_j^1, p_i^2 \rangle$  каждой пары кадров увеличим на единицу элемент  $MC_{j,i}$  матрицы  $MC$ . В результате в каждом элементе матрицы будет количество голосов за соответствие  $\langle p_j^1, p_i^2 \rangle$ . Чтобы получить множество соответствий для всей последовательности пар кадров будем искать такой набор элементов матрицы  $MC$ , чтобы в каждом столбце и каждой строке было выбрано не более одного элемента, причём сумма этих элементов должна быть максимальной. При этом элементы, значения которых меньше некоторого порога, будем отсекают. Индексы выбранных элементов будут задавать искомые соответствия, сумма элементов – суммарный голос. Для поиска таких индексов воспользуемся алгоритмом

Хопкрофта-Карпа для поиска максимального назначения (паросочетания в двудольном графе [4]).

## 6. РЕКОНСТРУКЦИЯ ГЕОМЕТРИИ

На данном этапе требуется по трёхмерным координатам отдельных точек объекта в разные моменты времени восстановить геометрию объекта и траекторию его движения. Формально: в моменты времени  $\{i\}=(1..N)$  камера  $c$  наблюдает точки  $P_i^c=\{p_{i,j}^c(k,c)\}$ . Требуется восстановить геометрию объекта  $O=\{P_k\}$ ,  $k=1..K$  в некоторой системе координат, связанной с объектом.

Идея алгоритма решения этой задачи заключается в том, что множества наблюдаемых точек  $P_i^c$  для соседних кадров часто бывают одинаковыми – точки появляются и пропадают из видимости камер не на каждом кадре. Это позволяет разбить всё множество моментов времени на непрерывные последовательные промежутки  $R_i=(i'_1..i'_2)$ , на которых множества точек  $Q_i$  одинаковы. Такие промежутки будем называть *диапазонами*. Для диапазонов  $R_x$  и  $R_y$ , для которых  $card(Q_x \cap Q_y) \geq 3$ , можно пересчитать координаты точек из  $Q_x \cup Q_y$ , которых нет в одном из диапазонов. Таким образом диапазоны  $R_x$  и  $R_y$  объединяются в диапазон  $R_z=join(R_x, R_y)$ . При этом можно объединить и геометрии фрагментов объектов, которые были реконструированы для этих диапазонов. Продолжая описанные действия для всех диапазонов, геометрия всего объекта может быть восстановлена.

## 7. ТЕСТИРОВАНИЕ

Тестирование производилось на реальных и синтетических данных. Однако эксперименты с реальными данными производить проблематично, поэтому такие эксперименты были проделаны в ограниченном объёме.

Для оценки алгоритмов поиска соответствий и реконструкции использовались метрики – ошибка репроекции и соотношение количества проекций точек, для которых удалось найти соответствия к общему числу проекций. В качестве тестовых данных были выбраны реальные данные, чистые синтетические данные и синтетические данные с шумом. Все данные содержали движение объекта с маркерами в пространстве, объединённое с его вращением. Тестирование показало, что ошибка репроекции не превышает 1 пикселя при наличии шума, а на чистых синтетических данных не превышает 0.2 пикселя. Процент использованных репроекции точек для реконструкции, в зависимости от последовательности и шума, находится в диапазоне примерно 60-70 % для данных с шумом и выше 80 % для чистых синтетических данных. Использовать все найденные проекции не получается потому, что очень часто маркеры видны на одной камере, в то время как на другой камере их не видно.

Для оценки статической погрешности был измерен объект с размером 20 см на расстоянии порядка 5 метров от обеих камер (калибровка системы осуществлялась с шаблоном на том же расстоянии от камер), вычисленное значение оказалось равно 20.09 см, что соответствует относительной погрешности измерений 0.45 % в данном эксперименте.

Для оценки качества алгоритмов вычисления характеристик были произведены комплексные эксперименты, которые заключались в вычислении величины ускорения свободного падения (с учётом сопротивления воздуха и без него). Для этого использовалась физическая модель, для которой было известно аналитическое решение. Полученные значения

координат подставлялись в аналитическое уравнение для выбранной модели, что позволяло получить значения коэффициентов уравнений. Эти коэффициенты выражаются через ускорение свободного падения и коэффициент линейной части сопротивления воздуха. Например, для процесса свободного падения уравнения представлены формулами (8) и (9) – без учёта сопротивления воздуха и с учётом сопротивления воздуха.

$$m \frac{\partial^2 x}{\partial t^2} = mg \quad (8)$$

$$m \frac{\partial^2 x}{\partial t^2} = mg - m\alpha \frac{\partial x}{\partial t} \quad (9)$$

## 8. ЗАКЛЮЧЕНИЕ

В данной работе был предложен новый алгоритм отслеживания движения объекта в пространстве и вычисление характеристик его движения бесконтактным способом. При этом в качестве источника информации предложен объект с инфракрасными маркерами, который наблюдается двумя камерами, которые снимают в инфракрасном диапазоне.

Для решения проблемы восстановления соответствий проекций точек предложен новый метод, основанный на использовании эпиполярных ограничений и схемы голосования для повышения устойчивости результатов.

Были проведены эксперименты, позволяющие оценить качество предложенных алгоритмов и точность измерений.

## 9. ЛИТЕРАТУРА

- [1] H. Bay, T. Tuytelaars, L. V. Gool, SURF: Speeded Up Robust Features, 2008, CVIU'08, Vol. 110, No. 3, pp. 346-359
- [2] R. Hartley, P. Sturm, Triangulation, 1994
- [3] R. Hartley, A. Zisserman. Multiple view geometry in computer vision. Cambridge, UK, 2000. Cambridge University Press
- [4] Hopcroft, John E.; Karp, Richard M. (1973), "An  $n^5/2$  algorithm for maximum matchings in bipartite graphs", SIAM Journal on Computing 2 (4): 225–231
- [5] Z. Khan, R. Herman, K. Wallen, T. Balch, An outdoor 3-D visual tracking system for the study of spatial navigation and memory in rhesus monkeys, 2005, Behavior Research Methods, vol. 37
- [6] M. Loaiza, A. Raposo, M. Gattass, A Novel Optical Tracking Algorithm for Point-Based Projective Invariant Marker Patterns, 2007
- [7] S. Smith, Real-time motion segmentation and shape tracking, 1995, In Proc. 5th Int. Conf. on Computer Vision
- [8] M. Weber, H. B. Amor, T. Alexander, Identifying Motion Capture Tracking Markers with Self-Organizing Maps, 2008, Virtual Reality Conference, VR'08. IEEE, pp. 297-298
- [9] A. Yilmaz, M. Shah, Contour-Based Object Tracking with Occlusion Handling in Video Acquired Using Mobile Cameras, 2004, IEEE TPAMI, vol. 26, pp. 1531-1536
- [10] Y. Zhao, J. Westhues, P. Dietz, J. Barnwell, S. Nayar, M. Inami, M. Nol, V. Branzoi, E. Bruns, Lighting Aware Motion Capture using Photosensing Markers and Multiplexed Illuminators, 2007, ACM TOG, Vol. 26, Issue 3, Article 36