

# Отслеживание объектов с использованием инфракрасных маркеров

Рамиз Зейналов<sup>1</sup>, Антон Якубенко<sup>1</sup>, Илья Толкунов<sup>2</sup>, Александр Мачихин<sup>3</sup>,

<sup>1</sup>Московский государственный университет имени М. В. Ломоносова, Москва, Россия,

<sup>2</sup>ООО "ЭВЕРЕСТ ВИТ", Москва, Россия,

<sup>3</sup>Научно-технологический центр уникального приборостроения Российской академии наук, Москва, Россия,  
[rzeynalov@graphics.cs.msu.ru](mailto:rzeynalov@graphics.cs.msu.ru), [toh@graphics.cs.msu.ru](mailto:toh@graphics.cs.msu.ru), [itolkunov@pamega.ru](mailto:itolkunov@pamega.ru), [aalexanderr@mail.ru](mailto:aalexanderr@mail.ru)

## Аннотация

В данной работе описывается метод отслеживания (трекинга) объектов с помощью стереопары, которая представляет собой две стационарные высокоскоростные камеры. Видимый диапазон камер отсекается с помощью диафрагмы, поэтому изображения обрабатываются в инфракрасном диапазоне. Объект обклеен инфракрасными маркерами, которые представляют собой диоды, излучающие свет в инфракрасном диапазоне. Для каждого момента времени требуется вычислить трёхмерные положения маркеров в пространстве и рассчитать по ним численные характеристики их движения. Особенность поставленной задачи состоит в том, что используется исключительно инфракрасный диапазон, на котором видны только инфракрасные маркеры, причём они неотличимы друг от друга. Как правило, в задачах, где применяется триангуляция точек, на входе есть видимый диапазон, который позволяет использовать дескрипторы особых точек для сопоставления проекций на разных видах, или оптический поток, позволяющий находить движущийся объект. Также при наличии видимого диапазона используются маркеры, которые несут в себе информацию о своём идентификаторе. В случае ИК-диапазона форма отслеживаемого объекта зачастую известна заранее. В имеющейся ситуации такие возможности отсутствуют, поэтому стоит задача получения межвидовых соответствий для маркеров.

**Ключевые слова:** инфракрасные маркеры, трекинг, сопоставление точек, высокоскоростные камеры.

## 1. ВВЕДЕНИЕ

Задача заключается в том, чтобы отследить движение объекта неизвестной формы, помеченного инфракрасными маркерами, в пространстве. Для отслеживания объекта используются именно ИК-маркеры, т.к. объект движется достаточно быстро и его изображение в видимом диапазоне очень сильно смазывается, в то время как ИК-маркеры выглядят достаточно контрастно, что делает их отслеживание более лёгким. После этого нужно определить центр масс объекта и характеристики его движения, такие как линейные и угловые скорости и ускорения объекта относительно его центра масс.

Задача трекинга объектов в настоящее время весьма востребована, поэтому уже существует множество различных её решений. Однако большинство из этих решений опирается на то, что форма объекта известна заранее [6], либо работа осуществляется в видимом диапазоне [9, 5], что позволяет использовать дескрипторы особых точек [1] или оптический поток [4] для поиска межвидовых соответствий.

## 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 [3, 12]. Для отслеживания небольшого количества объектов на статической сцене иногда используется оптический поток для поиска движущихся объектов [4, 12, 8]. Для этого берутся соседние кадры, 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 ([10, 13]).

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

## 3. ПРЕДЛОЖЕННЫЙ МЕТОД

Предложенное решение задачи состоит из трёх фаз: калибровка системы, вычисление трёхмерных координат точек, вычисление характеристик движения и формы объекта.

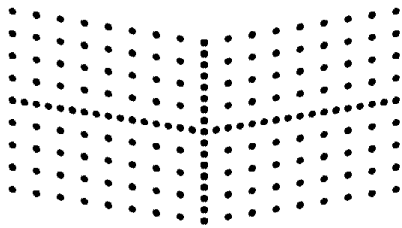
### 3.1 Калибровка

Как и в других случаях, когда стереопара используется в качестве измерительной системы, требуется процесс калибровки для получения внутренних параметров камер ( $K$  – матрица камеры, содержащая принципиальную точку и фокусные расстояния,  $D$  – коэффициенты дисторсии) и их взаимного расположения ( $R$  – матрица поворота,  $T$  – вектор переноса) в общей мировой системе координат. Для калибровки используется трёхмерный шаблон, составленный из инфракрасных маркеров, расположенных на двух плоских стендах под углом 120 градусов друг к другу. Расположение маркеров в разных плоскостях позволяет получить более высокую точность калибровки при меньшем числе кадров. Каждый стенд составлен из маркеров, образующих сетку 9 на 9, причём между соседними маркерами расстояния одинаковые. На каждой сетке на центральной линии

(перпендикулярной к общему ребру сеток) частота маркеров удвоена, частота маркеров удвоена также и на общем ребре сеток (рис. 1). Это необходимо для обеспечения эффективного распознавания шаблона в тех случаях, когда он виден не полностью хотя бы на одной из камер. В качестве критерия качества восстановленных трёхмерных координат  $Q_i$  будем использовать ошибку репроекции точки, то есть расстояние от проекции  $p_j$  до проекции полученной трёхмерной точки  $Q_j$  с матрицей проекции камеры  $P$ :

$$E_{reproj}^{k,j} = \left\| p_j - un(P_k \cdot Q_j) \right\|, P_k = K_k \cdot (R_k | T_k)$$

$$un(P) = \frac{P}{P(3)} \quad (1)$$



**Рисунок 1:** Калибровочный шаблон. Схематично.

Для распознавания маркеров сначала используется предобработка, а именно медианная фильтрация для удаления импульсного шума, гауссово размытие для удаления прочих шумов и для удаления видимого диапазона, если он не был подавлен физически оптическими фильтрами (или диафрагмой) на объективе. Для повышения контраста используется гамма-коррекция.

Для нахождения маркеров используется пороговая бинаризация и поиск связанных компонент. Компоненты фильтруются по площади в пикселях – отсеивается шум и отдельные пиксели. Координаты пикселей ищутся как математическое ожидание яркости пикселей в связанной компоненте.

Для распознавания шаблона сначала вычисляется пересечение основных линий шаблона (с удвоенной частотой маркеров) – две горизонтальные и одна вертикальная, которые вместе составляют крест в центре шаблона. Это пересечение ищется как крест, состоящий из пяти точек, наименее удалённых друг от друга. Пусть это будут точки  $P_c$ ,  $P_{right}$ ,  $P_{left}$ ,  $P_{top}$ ,  $P_{bottom}$ . После этого находятся линии креста – последовательно добавляются ближайшие точки. Затем ищутся точки, которые находятся в тех же строках/столбцах шаблона, что и точки, следующие за  $P_{right}$ ,  $P_{left}$ ,  $P_{top}$ ,  $P_{bottom}$ . Назовём эти точки  $P_1$ ,  $P_2$ ,  $P_3$  и  $P_4$ , начиная с правого верхнего квадранта шаблона по часовой стрелке. Затем берутся вторые точки на этих линиях и вычисляются гомографии из прямоугольника с соотношением сторон 2 : 1 в четырёхугольники, одна сторона которых лежит на вертикальной основной линии между точками, следующими за  $P_{top}$  и  $P_{bottom}$ , а противоположная ей находится между точками  $P_1$  и  $P_2$  ( $P_3$  и  $P_4$  соответственно для левой части шаблона). Используя гомографии, получаем следующие точки: сначала прогнозируем их положения, затем уточняем, пытаясь найти ближайшую точку на изображении, если она есть. После каждого такого шага гомографии уточняются. В результате получаем соответствия точек шаблона и точек на

изображении. Если хотя бы один из них не выполнен, операция распознавания считается неуспешной.

### 3.2 Вычисление движения

Стадия вычисления движения является наиболее трудоёмкой и состоит из нескольких последовательных частей:

1. Предобработка изображений
2. Распознавание маркеров
3. Двухмерный трекинг проекций маркеров между кадрами
4. Вычисление межвидовых соответствий точек
5. Вычисление трёхмерных положений точек
6. Трёхмерное сопоставление точек
7. Вычисления характеристик движения объекта.

Предобработка изображений, распознавание маркеров выполняется так же, как и при калибровке системы.

Сначала производится трекинг проекций точек на каждой камере отдельно. Для этого на первом кадре все маркеры получают свои уникальные идентификаторы  $P_{1,1}..P_{1,M}$ , а на каждом следующем кадре для каждой точки  $P_{i-1,j}$  на предыдущем кадре ищется ближайшая точка  $P_{i,k}$  на данном кадре. Если расстояние между точками  $P_{i-1,j}$  и  $P_{i,k}$  меньше порога  $T_{track2d}$ , точка на новом кадре получает идентификатор точки  $P_{i-1,j}$  на предыдущем кадре. Если на новом кадре остались точки, которые не получили идентификаторы, им назначаются новые идентификаторы. Таким образом, каждая точка на каждом кадре каждой камеры получает некоторый идентификатор, то есть получаются связи положений маркеров в различные моменты времени.

После этого необходимо восстановить трёхмерные координаты точек. Для этого нужно установить межвидовые соответствия. Об этом подробно написано в п. 4. Имея межвидовые соответствия точек, можно восстановить координаты трёхмерных точек путём триангуляции проекций. Для триангуляции используется алгоритм итерационных наименьших квадратов (Iterative-LS, [3]).

Когда становятся известны трёхмерные координаты отдельных точек в разные моменты времени, некоторые маркеры могут встречаться в последовательности несколько раз с разными идентификаторами. Это происходит, например, тогда, когда маркер в какой-то момент исчез из вида, затем через некоторое время появился (и, соответственно, мог успеть переместиться на значительное расстояние по неизвестному закону) и стал снова виден. В таком случае он будет иметь уже другой идентификатор. Для решения этой проблемы нужно восстановить межкадровые соответствия, так как один и тот же маркер может быть представлен в разные моменты времени разными идентификаторами. Эти соответствия восстанавливаются с помощью сопоставления отдельных фрагментов объекта, то есть, если какая-то часть объекта (состоящая из трёх и более точек) была видна некоторое время, затем одна точка пропала из вида, а другая появилась, можно вычислить относительное расположение исчезнувшей и появившейся точек, используя для этого систему координат трёх видимых точек.

### 3.3 Вычисление центра масс

Когда известны положения объекта в каждый момент времени, можно осуществлять анализ движения. Для этого сначала градиентным спуском находится положение центра

масс объекта. При этом оптимизируется суммарный момент вращения всех видимых маркеров объекта:

$$\sum_{i=1}^N \sum_{j=1}^M \left\| \vec{r}_i \cdot \left( \vec{V}_i - \frac{(\vec{r}_i, \vec{V}_i)}{\|\vec{r}_i\|} \cdot \vec{r}_i \right) \right\|, \vec{r}_i = \vec{P}_i - \vec{P}_0, \quad (2)$$

где  $N$  – количество кадров,  $M$  – количество точек,  $V_i$  – скорость точки в системе координат объекта,  $P_i$  – координаты точки,  $P_0$  – координаты центра масс,  $r_i$  – радиус-вектор точки в системе координат объекта. В качестве начального приближения берётся средняя точка объекта.

Затем вычисляются такие характеристики движения, как линейные и угловые скорости и ускорения центра масс объекта.

#### 4. УСТАНОВКА МЕЖВИДОВЫХ СООТВЕТСТВИЙ

Для того чтобы произвести триангуляцию, необходимо узнать межвидовые соответствия точек, то есть соответствие проекций точек на разных камерах. Для этого используются эпиполярные ограничения, которые позволяют ограничить поиск соответствующей точки на другом виде до одной прямой. Однако одних только эпиполярных ограничений не хватает: точки могут лежать вблизи одной эпиполярной линии. Чтобы решить эту проблему, осуществляется перебор возможных вариантов. Это делается таким образом, чтобы минимизировать ошибку репроекции (1) и с учётом непрерывности движения проекций точек на каждом виде.

Рассмотрим пару кадров с разных камер в один момент времени. На первом кадре пусть будут точки  $P_1..P_N$ , на втором –  $Q_1..Q_M$ , причём может быть  $M \neq N$ , так как количество видимых маркеров в один момент времени может быть разным для двух камер. Для каждой точки каждого вида будем искать возможные соответствия на другом виде. Делать это будем путём построения эпиполярных линий для каждой рассматриваемой точки. Возможными соответствиями для данной точки будем считать точки на другой камере, находящиеся от соответствующей эпиполярной линии на расстоянии, не превышающем порог  $T_{EC}$ . Когда находим все возможные соответствия для каждой точки, нужно будет выбрать из них наилучшие, причём если точке  $P_i$  соответствует точка  $Q_j$ , то этой точке не могут соответствовать никакие другие точки, и наоборот. Для этого вычислим ошибки репроекции (1), то есть расстояния от исходных проекций точек на изображения до проекций полученных трёхмерных точек, для всех возможных пар соответствий  $(i; j)$ . Затем будем брать соответствие с минимальной ошибкой репроекции из построенного списка возможных соответствий. Если возможных соответствий больше нет или минимальная ошибка репроекции (1) в этом списке превышает значение порога  $T_{reproj}$ , то заканчиваем искать соответствия для этого кадра, иначе мы нашли соответствие  $(i_0; j_0)$ . После этого удаляем все пары, конфликтующие с найденной, то есть пары вида  $(i_0; j)$  и  $(i; j_0)$ , где  $i \neq i_0$  и  $j \neq j_0$ , и выбираем из списка новую пару.

Таким образом, найдены соответствия для точек на одной паре кадров. Чтобы получить соответствия для всей последовательности, используется стратегия голосования. Это нужно для минимизации ошибок ложных соответствий и для повышения общей устойчивости алгоритма. Пусть всего на всех кадрах появилось  $M_1$  и  $M_2$  разных маркеров для

первой и второй камеры соответственно. Введём матрицу голосования  $M$  ( $M_1 \times M_2$ ) и для каждой пары кадров пройдём по всем полученным соответствиям  $(i; j)$ , увеличивая на единицу значение  $M_{i,j}$ . Таким образом получим матрицу голосов за каждое встретившееся соответствие. После этого можно искать глобальные для всей последовательности соответствия, извлекая их из полученной матрицы. Для этого будем выполнять следующие действия: пока матрица имеет ненулевые размерности и включает в себя ненулевые элементы, будем искать в ней максимальный элемент с индексами  $(i_0; j_0)$ . Такие элементы будем записывать в список глобальных соответствий, при этом удаляя из неё строку с индексом  $i_0$  и столбец с индексом  $j_0$ . В результате получаем глобальный список межвидовых соответствий маркеров. Имея список соответствий, можно производить триангуляцию, то есть вычисление координат трёхмерной точки по известным её проекциям на изображения и положениям камер в пространстве. В процессе триангуляции также можно предсказывать положения исчезнувших точек, то есть таких точек, которые были некоторое время видны на обеих камерах, после чего на одном из видов; аналогично для появившихся точек. Для этого вычисляются средние по нескольким ближайшим кадрам значения скорости в пространстве и из положения точки в пространстве строится луч с направлением, которое характеризует скорость, строится луч для проекции из камеры, на которой данную точку видно, затем вычисляется псевдопересечение этих лучей. Если ошибка репроекции меньше порога, точка оставляется для дальнейших вычислений. Пусть  $R_{ij}$  – точка с номером  $j$  в момент времени  $i$  в пространстве,  $P_{ij}$  и  $Q_{ij}$  – её проекции на изображения первой и второй камеры соответственно. В случае исчезновения из вида первой камеры точки  $j$  на кадре  $i$  имеем: проекции  $P_{i-k,j}, P_{i-k+1,j}, \dots, P_{i-1,j}$  известны,  $P_{ij}$  неизвестна, проекции  $Q_{i-k,j}, Q_{i-k+1,j}, \dots, Q_{ij}$  известны, точки в пространстве  $R_{i-k,j}, R_{i-k+1,j}, \dots, R_{i-1,j}$  известны,  $R_{ij}$  неизвестна. Кроме того, известны калибровки камер и их взаимное расположение, то есть известны матрицы полных калибровок камер  $C_1$  и  $C_2$ , которые включают в себя внутренние и внешние калибровки. По точкам  $R_{i-k,j}, R_{i-k+1,j}, \dots, R_{i-1,j}$  можно вычислить среднюю скорость точки  $R_{ij}$   $V_{ij}$ . К тому же, зная проекцию  $Q_{ij}$  и калибровку второй камеры, можно построить луч, на котором должна лежать точка  $R_{ij}$ . Координаты точки  $R_{i-1,j}$  и её скорость  $V_{ij}$  задают луч в пространстве, на котором должна находиться точка  $R_{ij}$ . Таким образом, точку  $R_{ij}$  можно определить как псевдопересечение этих лучей.

#### 5. ТЕСТИРОВАНИЕ

Для тестирования алгоритма предлагаются такие метрики, как средняя ошибка репроекции (3) и процент использованных проекций точек при триангуляции.

$$\frac{1}{M} \sum_{j=1}^M E_{reproj}^{k,j} \quad (3)$$

Ошибка репроекции является самым распространённым критерием оценки качества результата триангуляции, которая позволяет судить о том, насколько верно вычислены трёхмерные координаты точек и насколько их проекции отличаются от исходных. Кроме того, этот критерий свидетельствует о верности полученных межвидовых соответствий. Процент использованных проекций точек

позволяет судить, насколько эффективно использованы исходные данные, и делать выводы, какую часть трёхмерных точек удалось восстановить в процессе решения.

Тестирование производилось как на реальных, так и на синтетических данных. Реальных и синтетических данных было по 5 наборов. В качестве реальных данных брались результаты съёмки движения объекта с инфракрасными маркерами, при котором объект произвольным образом перемещался в пространстве. Синтетические данные представляли собой наборы изображений, такие же, как и реальные данные. Данные были с шумом и без шума, в движении объекта присутствовали перемещения и значительные вращения. Для создания изображений генерировался объект в движении. В качестве шума был выбран такой шум, при котором проекции смещались на случайный вектор, распределённый по закону нормального распределения с дисперсией в 2 пикселя. Во всех данных часто возникали ситуации, когда маркеры на изображениях камер при движении совмещались.

Однако даже в данных без шума нет 100-процентной точности и эффективности, так как нередко бывают ситуации, когда маркеры на изображении накладываются друг на друга (т.н. "marker swapping", [11, 13]), и при этом не всегда удаётся верно восстановить соответствия. Во всех данных имелось небольшое количество точек (семь маркеров для реальных, восемь - для синтетических), что затрудняло распознавание объекта на некоторых кадрах, и, соответственно, вычисление его движения. В таблице 1 представлены обобщённые данные о результатах тестирования системы.

Данные	Средняя ошибка репроекции, пикс.	Процент использованных проекций, %
Реальные	0.49-0.80	66.73-79.69
Синтетические	0.02-0.278	81.30-84.63
Синтетические с высоким уровнем шума	0.91-1.09	59.84-75.13

**Таблица 2.** Результаты тестирования.

Анализ результатов тестирования подтвердил предположение о важности для алгоритма наличия достаточного числа маркеров на объекте, чтобы в каждый момент времени обе камеры могли видеть хотя бы три общих маркера, что позволило бы восстанавливать форму всего объекта из отдельных фрагментов. Однако число маркеров не должно быть слишком большим, чтобы не возникало проблем с частыми их наложениями друг на друга [11, 13], приводящими к падению качества результата в получении межвидовых соответствий.

## 6. ЗАКЛЮЧЕНИЕ

Предложенный метод эффективно решает задачу трекинга объектов. От существующих методов он отличается тем, что не требуется знать геометрию отслеживаемых объектов. Кроме того, на эти объекты не накладывается требование к неоднородности текстуры для возможности использования дескрипторов особых точек или оптического потока. Более того, во входных данных вообще не требуется наличие видимого диапазона, т.к. обработка целиком происходит в

ИК-диапазоне, что значительно облегчает поиск маркеров. Научная новизна предложенного подхода заключается в том, что здесь не требуется заранее знать межвидовые соответствия для точек. Эти соответствия вычисляются предложенным алгоритмом поиска соответствий. Кроме алгоритма определения соответствий, предлагается новый шаблон калибровки и надёжный алгоритм его распознавания.

В дальнейшем планируется развитие проекта на случай использования более двух камер, что позволит значительно повысить точность и эффективность трекинга.

## 7. 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] C. Beall, B. Lawrence, V. Ila, F. Dellaert, 3D Reconstruction of Underwater Structures, 2010, IROS
- [3] R. Hartley, P. Sturm, Triangulation, 1994
- [4] 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
- [5] Z. Kim, Realtime Obstacle Detection and Tracking Based on Constrained Delaunay Triangulation, 2006, ITSC '06. IEEE, pp. 548-553
- [6] M. Loaiza, A. Raposo, M. Gattass, A Novel Optical Tracking Algorithm for Point-Based Projective Invariant Marker Patterns, 2007
- [7] D. P. Noonan, P. Mountney, D. S. Elson, A. Darzi, G.-Zh. Yang, A Stereoscopic Fibroscope for Camera Motion and 3D Depth Recovery during Minimally Invasive Surgery, 2009, ICRA '09, pp. 4463-4468
- [8] S. Smith, Real-time motion segmentation and shape tracking, 1995, In Proc. 5th Int. Conf. on Computer Vision
- [9] R. Subbarao, P. Meer, Y. Genc, A Balanced Approach to 3D Tracking from Image Streams, 2005, in Proc. IEEE and ACM ISMAR, pp. 70-78
- [10] 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
- [11] G. Welch, E. Foxlin, Motion Tracking: No Silver Bullet, but a Respectable Arsenal, 2002, IEEE Comput. Graph. Appl. 22, 6, 24-38.
- [12] 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
- [13] 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