

А.О. Кутергин

## АЛГОРИТМ РЕШЕНИЯ ЗАДАЧИ РАСПОЗНАВАНИЯ ДЕЙСТВИЙ ОБЪЕКТА НА ВИДЕОПОТОКЕ

*Рассмотрен алгоритм решения актуальной задачи распознавания действий объекта на видеопотоке с использованием неокогнитрона как системы принятия решений. Входные данные представляются в виде векторного поля, полученного путем вычисления оптических потоков последовательных кадров видеопотока.*

**Компьютерное зрение, метод Лукаса — Канада, неокогнитрон, распознавание действий.**

В настоящее время системы компьютерного зрения активно развиваются и применяются в прикладных задачах. Существующие методы и алгоритмы позволяют решать задачи распознавания образов, сегментации изображений, оценки параметров объектов на изображениях, слежения за объектами. Исследования же в области распознавания действий объектов на современном этапе развития компьютерного зрения остаются актуальным направлением. Представленный в настоящей работе алгоритм классификации действий объекта по данным видеопотока при помощи неокогнитрона позволяет решать более широкий класс подобных задач.

Рассматриваемую задачу можно разделить на три логических этапа:

- 1) получение пространственно-временных характеристик перемещений наблюдаемого объекта;
- 2) приведение полученных характеристик к виду, удовлетворяющему требованиям к входным данным неокогнитрона;
- 3) классификация преобразованных входных данных при помощи неокогнитрона.

Пусть существует некоторая область пространства и в данном пространстве — объекты, совершающие перемещения. Некоторую конечную последовательность этих перемещений будем называть действием объекта.

Область пространства, в которой находятся анализируемые объекты, будем называть сценой. Сцену условно можно разделить на набор объектов и фон. К объектам относятся те части сцены, поведение которых требуется проанализировать, к фону — все остальные части.

Кадром будем называть проекцию анализируемой сцены на фокальную плоскость регистрирующего прибора (видеокамеры) в момент времени  $t$ . Видеопотоком будем называть последовательность кадров, зарегистрированных через равные промежутки времени  $\Delta t$ .

Кадр является проекцией светового потока на фокальную плоскость (далее — плоскость). На данной плоскости определена непрерывная функция интенсивности для каждой точки с координатами  $x, y$ , которая определяет значение интенсивности светового потока в данной точке:

$$I = I(x, y), \quad (1)$$

Для анализа кадра при помощи вычислительных машин требуется дискретизация функции интенсивности. Плоскость разбивается на  $M \times N$  равных квадратных ячеек, в каждой ячейке проводится квантование функции интенсивности. Следовательно, кадр можно представить в виде матрицы размерно-

сти  $M \times N$ , элементы которой являются значением квантованной функции интенсивности в соответствующей ячейке. Элемент полученной матрицы будем называть пикселем.

Видеопоследовательность, являющаяся последовательностью кадров, зарегистрированных через равные промежутки времени, может быть записана следующим образом:

$$S = \{I_0, I_1, \dots, I_n\} \quad (2)$$

где  $I_0$  — функция интенсивности, определенная на кадре, зарегистрированном в момент времени  $t_0$ ;  $I_1$  — функция интенсивности, определенная на кадре, зарегистрированном в момент времени  $t_0 + \Delta t$ ;  $I_n$  — функция интенсивности, определенная на кадре, зарегистрированном в момент времени  $t_0 + \Delta t \cdot n$ .

Отсюда видно, что значение интенсивности пикселя зависит от его координат и момента времени, в который был зарегистрирован кадр; следовательно, можно записать функцию интенсивности, определенную на всем видеопотоке, следующим образом:

$$I = I(x, y, t) \quad (3)$$

Так как объект является частью сцены, а сцена представлена матрицей пикселей, то он описывается совокупностью некоторых пикселей данной матрицы. Изменение положения объекта в пространстве, т.е. движение объекта, вызывает изменение координат пикселей объекта в последовательности кадров.

Для вычисления пространственно-временных характеристик перемещений объекта будем использовать оптический поток [1]. Оптический поток есть векторное поле двумерных векторов, соответствующее смещению пикселей изображений объектов, вызванное сменой кадра, т.е. характеризует траекторию и скорость движения пикселей при смене кадров.

Существующие методы вычисления оптического потока используются для определения «схожих» точек на двух кадрах, в решении же данной задачи он используется для описания перемещений наблюдаемого объекта. На текущий момент разработано большое количество методов вычисления оптического потока, но исходя из [2] наиболее подходящим для решения поставленной задачи считаем метод Лукаса — Канада [3], обладающий рядом преимуществ относительно других:

- малая вычислительная сложность;
- устойчивость к шумам на входном изображении.

Уравнение оптического потока по Лукасу — Канада имеет следующий вид:

$$\begin{bmatrix} V_x \\ V_y \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n w_i I_x(q_i)^2 & \sum_{i=1}^n w_i I_x(q_i) I_y(q_i) \\ \sum_{i=1}^n w_i I_x(q_i) I_y(q_i) & \sum_{i=1}^n w_i I_y(q_i)^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum_{i=1}^n w_i I_x(q_i) I_t(q_i) \\ -\sum_{i=1}^n w_i I_y(q_i) I_t(q_i) \end{bmatrix}, \quad (4)$$

где  $(V_x, V_y)$  — вектор оптического потока в точке  $p$ ;  $q_1, q_2, \dots, q_n$  — пикселя окрестности точки  $p$ ;  $I_x(q_i), I_y(q_i), I_t(q_i)$  — частные производные изображения  $I$  по координатам  $x, y$  и времени  $t$ , вычисленные в точке  $q_i$ ;  $w_i$  — веса, которые будут присвоены пикселям  $q_i$  окрестности точки  $p$ , в качестве весов  $w_i$  будем использовать нормальное гауссовское распределение расстояния между  $q_i$  и  $p$ .

Решением данного уравнения будет множество векторов смещений пикселей объекта, которые и образуют векторное поле смещений, описываемое вектор-функцией  $V(x, y, t)$ . Метод Лукаса — Канада имеет проблему апертюры на малотекстурированных областях изображения. Но для определения на-

правления и характера перемещения объекта достаточно векторов смещения границ объекта. В точках границы объекта функция интенсивности  $I(x, y)$  будет иметь локальные экстремумы, так как грани на изображении обусловлены резким перепадом яркости, а следовательно, для областей данных точек уравнение оптического потока разрешимо методом Лукаса — Канада.

Оптический поток для данных двух кадров графически представлен на рис. 1. Для данного примера решением уравнения оптического потока будет являться совокупность векторов смещения наблюдаемого объекта  $V_1, V_2, V_3, V_4$ .

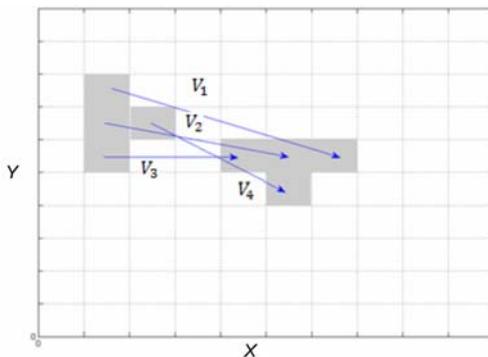


Рис. 1. Графическое представление оптического потока

Движение имеет следующие характеристики, которые однозначно позволяют классифицировать данное движение:

- скорость;
- траектория;
- ускорение;
- угловая скорость.

Полученные поля векторов есть множество векторов смещения точек наблюдаемого объекта в моменты времени  $t_0 \dots t_n$ ,  $\Delta t$  для стандартного видеопотока есть величина постоянная, таким образом, полученное векторное поле характеризует скорость и ускорение перемещения объекта. Усредненная интегральная кривая совокупного поля векторов смещения  $V(x, y, t)$  описывает траекторию движения объекта. Ротор полученного поля  $rot V$  есть вихревая характеристика совокупного поля векторов смещения, т.е., по сути, является угловой скоростью объекта. Как известно из кинематики твердого тела, для задания движения твердого тела нужно определить положение только трех его точек, не лежащих на одной прямой, а кинематической моделью твердого тела будет треугольник, построенный по трем этим точкам. Следовательно, для точного описания движения тела для каждого момента времени  $t_n$  достаточно трех векторов из поля  $V(x, y, t)$ , где

$$|V(x, y, t_n)| > 0 \quad (5)$$

и начала векторов не лежат на одной прямой. То есть для кадров, полученных в момент времени  $t_n > t_1$ , требуется решать уравнение оптического потока не для всех пикселей кадра, а только для выбранных на предыдущей итерации трех, что существенно понижает вычислительную сложность представленного алгоритма.

Классификацию действий объекта по полученным характеристикам будем производить при помощи нейронной сети типа «неокогнитрон» [4]. Использование нейронной сети данного типа обусловлено тем, что, согласно [4], при

распознавании образов она обладает следующими преимуществами: инвариантностью к позиционным сдвигам, аффинным преобразованиям, масштабированию, частичному перекрытию и потерям изображения.

Так как входом неокогнитрона является двумерная плоскость нейронов, также называемая сетчаткой, то требуется преобразовать полученное векторное поле смещений пикселей объекта в вид функции, определенной на конечном дискретном пространстве размерности 2 для подачи на вход неокогнитрона, причем при данном преобразовании не должны происходить потери информации о характере движения объекта.

Как уже было отмечено, для полной характеристики движения твердого тела достаточно всего трех его точек. Выберем три точки  $P_0 = (x_1, y_1)$ ,  $P_1 = (x_2, y_2)$ ,  $P_2 = (x_3, y_3)$  такие, что для момента времени  $t_n$ :

$$\begin{cases} |P_0P_1| = \max(\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}) \\ |P_1P_2| = \max(\sqrt{(x_2 - x_3)^2 + (y_2 - y_3)^2}) \\ |P_2P_0| = \max(\sqrt{(x_3 - x_1)^2 + (y_3 - y_1)^2}), \\ |V(x_1, y_1, t_n)| > 0 \\ |V(x_2, y_2, t_n)| > 0 \\ |V(x_3, y_3, t_n)| > 0 \end{cases} \quad (6)$$

т.е. выбранные точки являются максимально удаленными друг от друга (рис. 2).

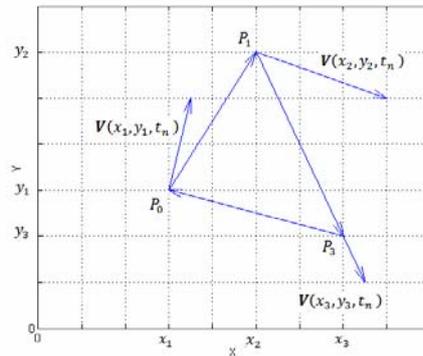


Рис. 2. Характеристические точки

Из кинематики твердого тела известно, что точкой, характеризующей движение системы частиц как целого, является центр масс. Найдем центр масс  $P_c$  системы  $P_0P_1P_2$ :

$$x_c = \frac{x_1 + x_2 + x_3}{3}, \quad (7)$$

$$y_c = \frac{y_1 + y_2 + y_3}{3}, \quad (8)$$

$$P_c = (x_c, y_c). \quad (9)$$

Вычислим вектор скорости  $V_c$  для найденного центра масс  $P_c$ :

$$V_{P_c} = \left( \frac{V_{x1} + V_{x2} + V_{x3}}{3}, \frac{V_{y1} + V_{y2} + V_{y3}}{3} \right), \quad (10)$$

Найденный вектор скорости является вектором средней скорости объекта и описывает поступательное движение центра масс. Так как в общем случае движение объекта характеризуется двумя составляющими, а именно поступа-

тельным и вращательным движением, требуется учитывать вращательную характеристику.

Введем вращательную характеристику. Рассмотрим движение центра масс как вращение относительно какой-либо из выбранных вначале точек. В качестве примера возьмем точку  $P_0$ . Воспользуемся уравнением Эйлера для связи скоростей двух точек:

$$\mathbf{V}_c = \mathbf{V}_{P_0} + \vec{\omega} \times \overline{P_0 P_c}, \quad (11)$$

где  $\vec{\omega}$  — вектор угловой скорости объекта. Скорость вращения центра масс относительно точки  $P_0$  находится следующим образом:

$$\mathbf{V}_{P_c P_0} = \vec{\omega} \times \overline{P_0 P_c}, \quad (12)$$

То есть

$$\mathbf{V}_{P_c P_0} = \mathbf{V}_{P_c} - \mathbf{V}_{P_0}, \quad (13)$$

$\mathbf{V}_{P_0}$  известно, так как входит в найденное векторное поле смещений точек объекта,  $\mathbf{V}_{P_c}$  вычисляется по соотношению (10). Из соотношений (11)–(13) очевидно, что вектор угловой скорости  $\vec{\omega}$  равен для всех точек объекта. Отсюда следует, что, зная значения вектора скорости центра масс  $\mathbf{V}_{P_c}$  и вектора скорости вращения центра масс относительно какой-либо точки объекта  $\mathbf{V}_{P_c P_0}$ , можно найти вектора скоростей для любой точки объекта. То есть движение объекта однозначно характеризуется двумя вектор-функциями:

$$\begin{cases} \mathbf{V}_{P_c}(t) = \frac{V_{x_1}(t) + V_{x_2}(t) + V_{x_3}(t)}{3} \cdot \mathbf{x} + \frac{V_{y_1}(t) + V_{y_2}(t) + V_{y_3}(t)}{3} \cdot \mathbf{y}, \\ \mathbf{V}_{P_c P_0}(t) = \mathbf{V}_{P_c}(t) - \mathbf{V}(x_1, x_2, t) \end{cases} \quad (14)$$

где  $\mathbf{x}$  и  $\mathbf{y}$  — координатные орты.

Как было отмечено выше, неоконитрон принимает на вход матрицу значений, размерность которой соответствует размерности входной плоскости нейронов. Пусть  $\mathcal{X}$  и  $\mathcal{Y}$  — количество нейронов по соответствующему направлению входной плоскости. Тогда функция  $\mathbf{F} = \mathbf{F}(x, y)$  будет дискретной функцией входных значений, а областью определения данной функции является матрица  $\mathbf{M}$  входных значений размерности  $\mathcal{X} \times \mathcal{Y}$ .

Функция должна однозначно характеризовать движения объекта в целом. Как мы уже знаем, система уравнений (14) однозначно характеризует движение объекта, следовательно, значения функции  $\mathbf{F}$  должны зависеть от положения точки центра масс  $P_c$ , вектора скорости точки центра масс  $\mathbf{V}_{P_c}$  и вектора скорости вращения точки центра масс относительно любой другой точки  $\mathbf{V}_{P_c P_0}$ .

Возьмем точки  $K_0, K_1, K_2$  такие, что:

$$\begin{aligned} K_0 &= (x_c, y_c), \\ K_1 &= (x_c + V_{P_c P_0 x}, y_c + V_{P_c P_0 y}), \\ K_2 &= (x_c + V_{P_c x}, y_c + V_{P_c y}), \end{aligned} \quad (15)$$

т.е. координатами точки  $K_0$  являются координаты точки центра масс объекта, координатами точки  $K_1$  — координаты конца вектора скорости вращения центра масс относительно точки  $P_0$ , а координатами точки  $K_2$  — координаты конца вектора скорости точки центра масс.

Найдем кривую Безье для данных трех точек:

$$\mathbf{B}(z) = (1 - z)^2 K_0 + 2z(1 - z)K_1 + z^2 K_2, z \in [0, 1]. \quad (16)$$

На рис. 3 графически представлено данное преобразование.

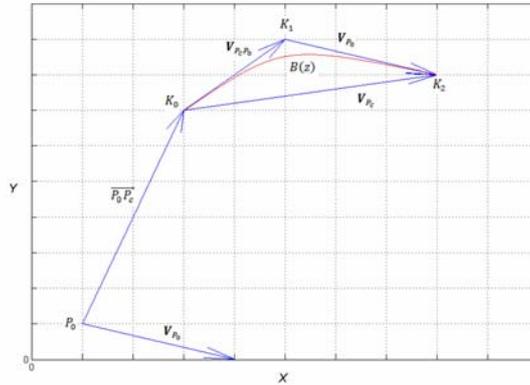


Рис. 3. Кривая Безье, построенная по трем характеристическим точкам

Как видно из графика, вектора  $V_{P_0P_0}$  и  $V_{P_c} - V_{P_0P_0}$  являются касательными к кривой Безье. Из соотношений (15) и (16) очевидно, что по полученной кривой можно восстановить значения векторов скоростей  $V_{P_c}$  и  $V_{P_0P_0}$ , в то же время исходя из (11)–(13) можно восстановить вектор скорости для любой точки данного тела. То есть полученное уравнение кривой однозначно характеризует движение данного тела, но не учитывает направление движения.

Для каждой пары кадров  $t_n$  и  $t_{n+1}$  находится кривая Безье  $B_n(z)$ , множество найденных точек всех кривых Безье  $B_1(z) \dots B_n(z)$  за промежуток времени  $[t_0, t_{n+1}]$  будет являться областью определения функции  $F$  входных данных неоконитрона. Для учета направления движения требуется ввести временную составляющую. Пусть значение  $F$  растет с течением времени, т.е. производная функции  $F$  по  $x$  и  $y$  будет являться характеристикой направления движения. Тогда функция  $F$  принимает следующий вид:

$$F(B_n(z)) = n + z, \quad z \in [0,1]. \quad (17)$$

Следовательно, область определения функции  $F$ , матрица  $M$  входных значений для подачи на сетчатку неоконитрона, будет такой, что:

$$\begin{cases} m_{ij} = n + z \\ i = B_{n_x}(z) \\ j = B_{n_y}(z) \\ z \in [0,1] \end{cases} \quad (18)$$

Для рассмотренного выше примера функция  $F$  графически представлена на рис. 4.

Далее матрица значений функции  $F$  подается на вход неоконитрона. В случае успешной классификации действия объекта область определения функции обнуляется, в обратном случае — расширяется на  $B_n(z)$ .

Ввиду использования для классификации нейронной сети типа «неоконитрон» данный алгоритм обладает высокой устойчивостью к шумам на входном видеопотоке, ориентации объекта в пространстве, частичному перекрытию наблюдаемого объекта другими объектами сцены в процессе совершения им действий. Также данный алгоритм обладает небольшой вычислительной сложностью, так как для вычисления пространственно-временных характери-

стик движения наблюдаемого объекта используются только три точки данного объекта, т.е. скорость выполнения алгоритма не зависит от разрешения входного видеопотока и может выполняться в реальном времени.

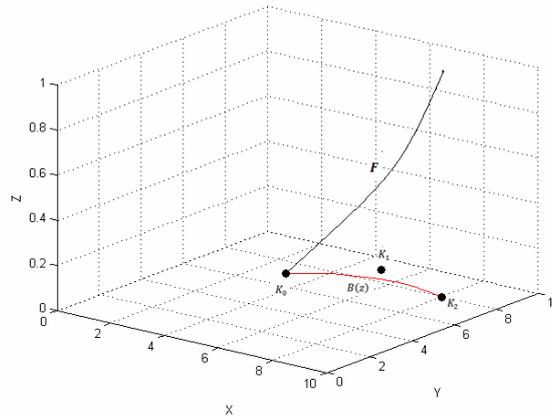


Рис. 4. Графическое представление функции  $F$

Данный алгоритм применим для классификации действий любых объектов, так как использует только ключевые точки наблюдаемого объекта. Может использоваться во многих прикладных приложениях, таких как охранные системы, робототехника, системы мониторинга, автоматизированные системы судейства спортивных мероприятий, благодаря низкой вычислительной сложности и устойчивости к качеству входных данных.

#### ЛИТЕРАТУРА

1. Prazdny K. The information in optical flows. Computer Science Department. Univ. of Essex, Colcheste, 1980.
2. Barron J. L., Beauchemin S. S. The computation of optical flow // ACM Computing Surveys. 1995. № 27. С. 433–466.
3. Lucas B. D., Kanade T. An iterative image registration technique with an application to stereo vision // Proceedings of Imaging Understanding Workshop. 1981. С. 121–130.
4. Fukushima K., Miyake S., Takayuki I. Neocognitron: A neural network model for a mechanism of visual pattern recognition // IEEE Transaction on Systems, Man and Cybernetics SMC. 1983. 13 (5). 826–34.

A.O. Kutergin

*Algorithm of solving a problem to recognize actions of object at the video stream*

*The article considers algorithm of solving a topical problem to recognize actions of object at the video stream, using a neocognitron as a system of taking decisions, while the input data being presented as a vector field obtained by calculating optical flows of the video stream sequence shots.*

**Computer vision, Lukas method-Canada, neocognitron, recognition of actions.**