

УДК 004.93

РАСПОЗНАВАНИЕ ЖЕСТОВ РУКИ В ЗАДАЧЕ БЕСКОНТАКТНОГО УПРАВЛЕНИЯ БЕСПИЛОТНЫМ ЛЕТАТЕЛЬНЫМ АППАРАТОМ*

В. Э. Нагапетян¹, В. М. Хачумов²

¹Российский университет дружбы народов,
117198, Москва, ул. Миклухо-Маклая, 6

²Институт системного анализа РАН,
117312, Москва, просп. 60-летия Октября, 9

E-mail: vahagnahapetyan@gmail.com
vmh48@mail.ru

Рассматривается задача бесконтактного управления беспилотным летательным аппаратом посредством жестов руки. Предложена система команд для четырёхвинтового вертолёта AR.Drone 2.0, оснащённого встроенным компьютером, двумя цветными видеокамерами, операционной системой Linux и датчиками для измерения высоты, скорости и стабильности полёта. В качестве устройства ввода информации в систему управления используется трёхмерный сенсор "Asus Xtion Pro Live", работающий на принципах триангуляции и структурированного света. Распознавание жестов осуществляется посредством покадровой обработки видеоряда, состоящего из дальностных изображений. Метод не требует начального обучения, не чувствителен к изменениям освещения, инвариантен к размерам ладони и тела человека.

Ключевые слова: управление БПЛА, распознавание жестов, обработка изображений.

Введение. В последнее время беспилотные летательные аппараты (БПЛА) приобретают большую популярность не только в военной сфере, но и в сельском хозяйстве, в исследованиях ледовой обстановки, зондировании Земли и т. д. В зависимости от конкретной задачи управление БПЛА может осуществляться тремя основными способами: автономное управление, самонаведение и телеуправление. Телеуправление обычно выполняется специально обученным оператором посредством специальных контроллеров и компьютерных систем. Как правило, управление такими устройствами требует высокого мастерства и долгого времени обучения. Пользовательский интерфейс оператора ограничен двумерной плоскостью, что делает управление неестественным и трудным. Для преодоления этих недостатков проводятся многочисленные исследования. Например, компанией "Parrot" ведутся работы по созданию системы управления БПЛА посредством жестов руки с помощью надеваемой на руку специальной сенсорной повязки [1]. В [2] исследуется система управления БПЛА бесконтактными жестами тела человека с использованием трёхмерного сенсора "Kinect".

В данной работе предлагается метод управления БПЛА посредством бесконтактных трёхмерных жестов руки человека, которые в последние годы стали часто применять в системах человеко-машинного общения как естественный и удобный метод взаимодействия между человеком и компьютером [3]. В качестве базовых команд управления выбираются

*Работа выполнена в рамках Программы фундаментальных исследований ОНИТ 1 РАН «Интеллектуальные информационные технологии, системный анализ и автоматизация».

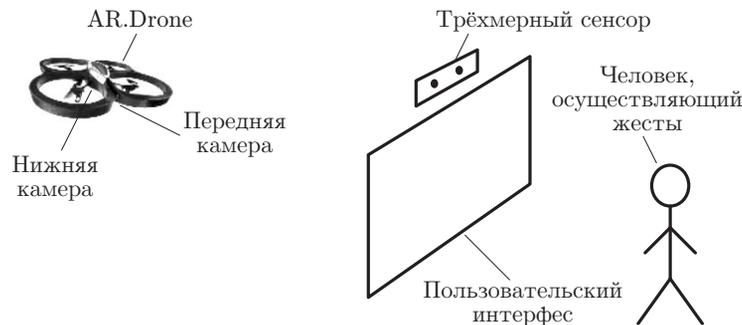


Рис. 1. Структура системы бесконтактного управления БПЛА посредством жестов человека

движения летательного аппарата по всем направлениям пространства, включая поворот вокруг вертикальной оси, взлёт и посадку. На рис. 1 приведена структура предложенной системы управления. Компонентами системы являются:

- 1) вертолёт AR.Drone 2.0 [4] в качестве модели БПЛА;
- 2) трёхмерный сенсор "Asus Xtion Pro Live" [5] для захвата видеоряда и распознавания жестов руки человека;

- 3) пользовательский интерфейс, в котором видны изображения, получаемые от двух камер БПЛА и трёхмерного сенсора, а также распознанные команды управления.

Распознавание жестов человека осуществляется посредством анализа каждого кадра видеоряда трёхмерного сенсора. Распознанные жесты передаются в качестве базовых команд в блок управления AR.Drone 2.0 через связь Wi-Fi.

Жесты управления. Для естественности управления жесты человека подобраны таким образом, чтобы все движения руки соответствовали движению полёта БПЛА. В табл. 1 приведены выбранные жесты и базовые команды управления AR.Drone 2.0.

Управление полётом БПЛА происходит следующим образом.

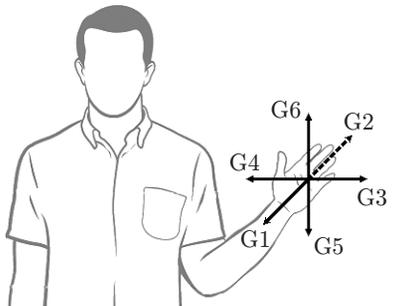
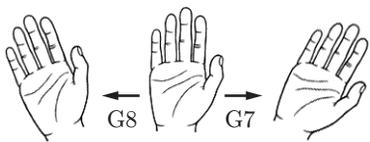
1. Человек согнутую в кулак руку устанавливает в некоторое положение и открывает ладонь руки (жест G9), что приводит к старту двигателей БПЛА.
2. Для управления направлением полёта человек тянет руку в желаемом направлении (жесты G1–G6).
3. Поворот БПЛА вокруг своей оси и посадка осуществляются посредством жестов G7, G8 и G10.

Распознавание жестов руки проводится посредством покадровой обработки видеоряда, формируемого трёхмерным сенсором. В настоящее время известны алгоритмы [6] и свободно доступные программные библиотеки [7], которые позволяют различать ключевые точки тела человека. Для определения позиции ладони руки используются библиотеки OpenNI и NITE, с помощью которых можно в реальном времени распознавать позицию произвольной точки ладони руки посредством обработки дальностного изображения. Пусть заданы пространственная точка M с координатами (x, y, z) и её ортогональная проекция m с координатами (x, y) на плоскость наблюдения $z = 0$. Под дальностным изображением будем понимать цифровое изображение $d(x, y)$, каждая точка (x, y) которого принимает неотрицательные целые значения, равные величине расстояния от точки M до точки m . Дальностное изображение всегда можно преобразовать во множество $\{(x, y, d(x, y))\}$, являющееся моделью наблюдаемого объекта.

Пусть имеется дальностное изображение человека и известна произвольная точка ладони (x_r, y_r, z_r) . Обозначим через (x_0, y_0, z_0) позицию центра ладони руки в момент открытия ладони, где z_0 — расстояние от ладони до плоскости наблюдения трёхмерного сенсора,

Т а б л и ц а 1

Жесты руки, соответствующие базовым командам управления

Жест	Обозначение жеста	Базовая команда
	G1	Вперёд
	G2	Назад
	G3	Влево
	G4	Вправо
	G5	Вниз
	G6	Вверх
	G7	Поворот вокруг своей оси против часовой стрелки
	G8	Поворот вокруг своей оси по часовой стрелке
	G9	Старт
	G10	Посадка

(x_0, y_0) — позиция ладони в системе координат OXY , началом которой является левый верхний угол изображения, полученного от трёхмерного сенсора.

Позиция центра ладони руки вычисляется алгоритмом извлечения изображения руки.

1. Создаётся сфера вокруг точки (x_r, y_r, z_r) радиуса $\lambda = \lceil C/z_r \rceil$, где C — некоторая константа, зависящая от характеристик конкретного сенсора (в экспериментах $C = 80000$).

2. Помечаются все точки дальностного изображения, которые не входят в построенную сферу.

3. Помечаются все точки дальностного изображения, которые не связаны с распознанной точкой ладони (x_r, y_r, z_r) во множестве непомеченных точек.

4. На базе исходного изображения создаётся новое дальностное изображение размером $(2\lambda + 1) \times (2\lambda + 1)$, в которое входят только непомеченные точки (рис. 2, *a, b*).

После извлечения изображения руки оно преобразовывается в бинарное (рис. 2, *c*). За центр ладони принимается центр окружности, вписанной в фигуру руки и имеющей максимальный радиус. Вычисляется эта позиция в один проход посредством карт расстояний [8] (рис. 2, *d*).

Зная позицию центра ладони (x, y, z) в каждом кадре видеоряда и начальную позицию центра ладони (x_0, y_0, z_0) , жесты G1–G6 можно распознать путём анализа изменений позиции центра ладони по трём осям пространства. Так, например, если $z - z_0 < 0$, что соответствует приближению ладони руки к сенсору, то выполняется команда «вперёд».

Одним из недостатков дальностных изображений является присутствие шума в контурах наблюдаемых фигур, что затрудняет распознавание позиции центра ладони. Для преодоления данной проблемы производим сглаживание. Обозначим через (x_k, y_k, z_k) рас-

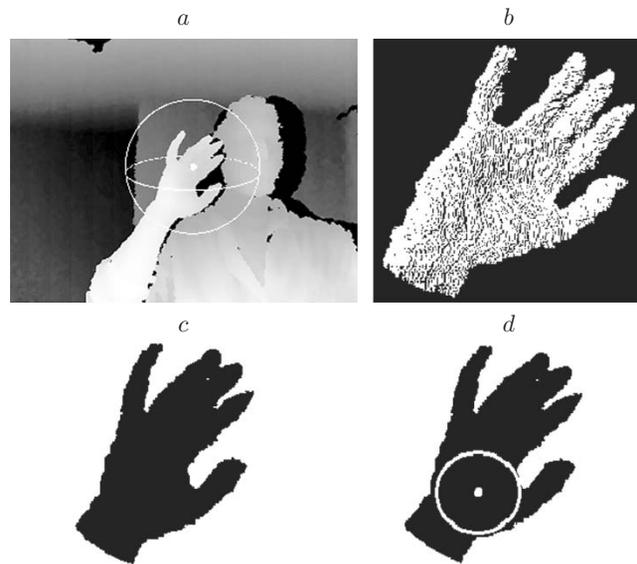


Рис. 2. Извлечение и обработка изображения руки: *a* — дальностное изображение человека, распознанная точка ладони и сфера вокруг найденной точки; *b* — извлечённое изображение руки; *c* — бинарное изображение; *d* — центр ладони и вписанная окружность с максимальным радиусом

познанную в кадре k позицию центра ладони руки и через (x'_k, y'_k, z'_k) — сглаженную позицию центра ладони, которая вычисляется следующим образом:

$$x'_k = \begin{cases} x'_{k-1}, & \text{если } |x_k - x'_{k-1}| < \tau, \\ x_k - \text{sg}(x_k - x'_{k-1})\tau, & \text{если } |x_k - x'_{k-1}| \geq \tau, \end{cases}$$

$$y'_k = \begin{cases} y'_{k-1}, & \text{если } |y_k - y'_{k-1}| < \tau, \\ y_k - \text{sg}(y_k - y'_{k-1})\tau, & \text{если } |y_k - y'_{k-1}| \geq \tau, \end{cases}$$

$$z'_k = \begin{cases} z'_{k-1}, & \text{если } |z_k - z'_{k-1}| < \tau, \\ z_k - \text{sg}(z_k - z'_{k-1})\tau, & \text{если } |z_k - z'_{k-1}| \geq \tau, \end{cases}$$

$$\text{sg}(a) = \begin{cases} -1, & \text{если } a < 0, \\ 0, & \text{если } a = 0, \\ 1, & \text{если } a > 0, \end{cases}$$

где τ — целочисленная константа, устанавливаемая пользователем и определяющая величину допустимого шума (в проведённых экспериментах $\tau = 12$). В первом кадре видеоряда $(x'_1, y'_1, z'_1) = (x_1, y_1, z_1)$.

Жесты G7 и G8 распознаются посредством вычисления моментов бинарного изображения руки [9]. Обозначим через $\{M_{0,0}\}$, $\{M_{1,0}, M_{0,1}\}$ и $\{M_{2,0}, M_{1,1}, M_{0,2}\}$ моменты нулевого, первого и второго порядка соответственно, причём

$$M_{0,0} = \sum_x \sum_y f(x, y),$$

$$M_{1,0} = \sum_x \sum_y x f(x, y), \quad M_{0,1} = \sum_x \sum_y y f(x, y),$$

$$M_{2,0} = \sum_x \sum_y x^2 f(x, y), \quad M_{1,1} = \sum_x \sum_y xy f(x, y), \quad M_{0,2} = \sum_x \sum_y y^2 f(x, y),$$

где

$$f(x, y) = \begin{cases} 1, & \text{если точка } (x, y) \text{ имеет чёрный цвет,} \\ 0 & \text{в противном случае.} \end{cases}$$

Введём следующее обозначение [9]:

$$\arctan^*(i, j) = \begin{cases} \arctan(i/j), & j > 0, \\ \arctan(i/j) + \pi, & i \geq 0, j < 0, \\ \arctan(i/j) - \pi, & i < 0, j < 0, \\ +\pi/2, & i > 0, j = 0, \\ -\pi/2, & i < 0, j = 0, \\ \text{не определён,} & i = 0, j = 0. \end{cases}$$

Функция \arctan^* удовлетворяет условию $\arctan(i, j) \in (-\pi, \pi]$. Угол наклона ладони руки определяется величиной

$$\Theta = \frac{\arctan^*(b, (a - c))}{2},$$

где

$$a = \frac{M_{2,0}}{M_{0,0}} - x_c^2, \quad c = \frac{M_{0,2}}{M_{0,0}} - y_c^2, \quad b = 2 \left(\frac{M_{1,1}}{M_{0,0}} - x_c y_c \right), \quad x_c = \frac{M_{1,0}}{M_{0,0}}, \quad y_c = \frac{M_{0,1}}{M_{0,0}}.$$

Поворот ладони руки по часовой стрелке (G8) вызывает увеличение угла Θ , поворот против часовой стрелки (G7) — уменьшение. Таким образом, обнаружение характера изменений угла Θ в каждом кадре видеоряда приводит к распознаванию указанных жестов.

Для распознавания жестов G9 и G10 вычисляются площади S_1 и S_2 описанного и вписанного многоугольников бинарного изображения руки соответственно. Если значение $|S_1 - S_2|$ меньше порогового значения \bar{S} , то ладонь руки считается согнутой в кулак, в противном случае — открытой (в экспериментах принято $\bar{S} = \left\lceil \frac{62 \cdot 10^7}{z} \right\rceil$, где z — расстояние от центра ладони до плоскости наблюдения).

Результаты распознавания жестов. Для оценки качества распознавания приведённых жестов предложенный метод запрограммирован и апробирован на базе тестовых выборок из 1000 элементов, где каждый элемент состоял из двух изображений, соответствующих кадрам начала и конца жеста. Результаты распознавания приведены в табл. 2. Точность распознавания вычисляется как доля жестов действительно принадлежащих к данному классу, относительно всех жестов, которые были отнесены к этому классу. Полнота распознавания определяется как доля найденных классификатором жестов, принадлежащих к классу, относительно всех жестов этого класса в тестовой выборке. Из таблицы

Таблица 2

Характеристики качества распознавания жестов руки

Характеристики качества распознавания	Значения									
	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
Точность	0,99	1	1	1	0,98	0,99	0,98	0,98	0,9	0,92
Полнота	0,98	0,99	1	1	0,99	1	0,96	0,95	0,92	0,95

видно, что средняя точность и полнота распознавания составляют 97 %. Было выявлено, что в основном ошибки классификатора связаны с ошибками оператора, осуществлявшего жесты. Так, например, при показе жеста G1 оператор иногда опускает руку вниз, тем самым производя жест G5, а жест G6 в некоторых случаях может осуществляться при показе жеста G2. Программа была протестирована на персональном компьютере с двухъядерным процессором без использования видекарты. Обработка кадра занимает не более 15 мс, что применительно к решаемой задаче соответствует режиму реального времени.

Заключение. Представленные в данной работе набор команд и метод распознавания жестов предназначены для бесконтактного управления полётом БПЛА. Преимуществом подхода по сравнению с [3] является отсутствие необходимости надевания на руку или тело специальных устройств и повязок. В отличие от подхода [4] предлагаемый метод не требует движения всего тела человека. В качестве недостатка отметим, что метод не пригоден в задачах высокоточного управления полётами. Базовые команды управления могут быть легко дополнены новыми жестами человека применительно к другим моделям БПЛА. В дальнейшем предполагается усовершенствование методов обработки и распознавания дальностных изображений, инвариантных к аффинным преобразованиям [10], и проведение экспериментов по отработке траекторных движений квадрокоптера на основе предложенного метода управления [11].

СПИСОК ЛИТЕРАТУРЫ

1. **Parrot AR.Drone 2.0:** Even more piloting possibilities! URL: <http://blog.parrot.com/2014/01/06/parrot-ar-drone-2-0-even-more-piloting-possibilities/> (дата обращения: 01.02.2014).
2. **AR.Drone + Kinect** [Video]. URL: <http://ardrone.parrot.com/best-of-user-videos/2012/02/17/ardrone-kinect/> (дата обращения: 01.02.2014).
3. **Devyatkov V., Alfimtsev A.** Human-computer interaction in games using computer vision techniques // Image Processing: Concepts, Methodologies, Tools, and Applications. Hershey, USA: IGI Global, 2011. Ch. 10. P. 146–167.
4. **AR.Drone 2.0.** URL: <http://ardrone2.parrot.com/> (дата обращения: 01.02.2014).
5. **Asus Xtion Pro Live.** URL: http://www.asus.com/Multimedia/Xtion_PRO_LIVE/ (дата обращения: 01.02.2014).
6. **Shotton J., Fitzgibbon A., Cook M. et al.** Real-time human pose recognition in parts from single depth images // Proc. CVPR'11. IEEE, 2011. P. 1297–1304.
7. **OpenNI:** The standard framework for 3D sensing. URL: <http://www.openni.ru/openni-sdk/index.html> (дата обращения: 12.02.2015).
8. **Palagyi K.** Skeletonization and its applications // K. Palagyi's home page. URL: <http://www.inf.u-szeged.hu/~palagyi/skel/skel.html> (дата обращения: 01.02.2014).

9. **Freeman W. T., Anderson D. B., Beardsley P. et al.** Computer vision for interactive computer graphics // IEEE. Comput. Graph. Appl. 1998. **18**, N 3. P. 42–53.
10. **Роженцов А. А., Морозовский К. В., Баев А. А.** Модифицированное обобщённое преобразование Хафа для обработки трёхмерных изображений с неизвестными параметрами вращения и масштабирования // Автометрия. 2013. **49**, № 2. С. 30–41.
11. **Белоконь С. А., Золотухин Ю. Н., Котов К. Ю. и др.** Использование фильтра Калмана в системе управления траекторным движением квадрокоптера // Автометрия. 2013. **49**, № 6. С. 14–24.

Поступила в редакцию 26 февраля 2014 г.
