

УДК 004.65

## СИСТЕМА РАСПОЗНАВАНИЯ ЖЕСТОВ НА ОСНОВЕ НЕЙРОСЕТЕВЫХ ТЕХНОЛОГИЙ

А. В. Чуйков<sup>1</sup>, А. М. Вульфин<sup>2</sup>

<sup>1</sup>artem.guy@gmail.com, <sup>2</sup>vulfin.alexey@yandex.ru

ФГБОУ ВО «Уфимский государственный авиационный технический университет» (УГАТУ)

Поступила в редакцию 16.05.2017

**Аннотация.** Целью является повышение эффективности систем распознавания жестов на основе нейросетевых алгоритмов анализа изображений. Выполнен аналитический обзор существующих на сегодняшний день программных решений и подходов к реализации систем распознавания жестов, включающих в себя стадии обнаружения, отслеживания и распознавания. Разработана структурная схема системы распознавания жестов, разработан алгоритм распознавания на основе нейросетевых технологий, выбраны подходящие дескрипторы образцов и структура классификаторов для достижения наилучших результатов. Для снижения размерности вектора признаков использована ассоциативная сеть с узким горлом и многослойным персептроном с вероятностным методом главных компонент. Также произведена оценка эффективности программного модуля распознавания жестов.

**Ключевые слова:** распознавание жестов; нейросетевые технологии; метод опорных векторов; локальные бинарные шаблоны; алгоритм адаптивного усиления классификаторов; гистограмма направленных градиентов; OpenCV; MATLAB.

### ВВЕДЕНИЕ

В области создания эффективного человеко–машинного интерфейса и управления посредством жестов на сегодняшний день существует достаточное количество технологий:

- бесконтактные сенсоры, которые реализуют технологию распознавания жестов различной сложности (Microsoft Kinect, LeapMotion [1], Ractiv Touch+ [2], SoftKinetic [3], Xtion Pro [4]);

- приложения, которые позволяют управлять компьютером, используя простые жесты, выделяемые из видеопотока (Flutter, NPointer [5], EyeSight PC [6], Control Air [7]).

Несмотря на достигнутые успехи в данной области, существует и ряд недостатков [8], ограничивающих использование этих технологий как в массовом сегменте, так и в узкоспециализированном, например в военной технике.

При построении систем распознавания жестов необходимы высококачественные камеры с достаточной светочувствительностью и разрешением матрицы, и таких камер, как правило, требуется несколько. Также приходится использовать высокопроизводительные системы для обработки потокового видео высокого разрешения. Все это существенно влияет на конечной стоимости оборудования и препятствует массовому производству и повсеместному внедрению подобных систем.

Для построения эффективных систем распознавания требуется реализация алгоритмов машинного зрения, обработки и оптимизации информационных потоков.

Таким образом, технология распознавания как динамических, так и статических жестов все еще не достигла достаточного уровня по совокупности качества

распознавания, скорости работы и требуемых вычислительных и аппаратных ресурсов. На сегодняшний день актуальной задачей является разработка и совершенствование алгоритмов и методов распознавания жестов, которые могут быть интегрированы в системы эффективного взаимодействия человека и машины. Цель работы – повысить эффективность алгоритмов распознавания жестов на основе анализа двухмерной модели пространства (геометрической модели плоской проекции трехмерного пространства, которая представлена на фотоизображении или видеокadre) с помощью нейросетевых технологий.

### АНАЛИЗ СУЩЕСТВУЮЩИХ АЛГОРИТМОВ РАСПОЗНАВАНИЯ ЖЕСТОВ

В ходе анализа алгоритмов обнаружения объектов [9–13] было установлено, что они просты в реализации, но большая часть из них обладает низкой эффективностью и подвержена сильному влиянию внешних факторов, таких как: освещение, фон, размер объекта и прочее. Для повышения эффективности применения данных алгоритмов, необходимо их комбинировать, что позволит повысить точность локализации объекта на изображении.

Алгоритмы отслеживания объектов [14–16] обладают набором существенных преимуществ, однако недостатком является высокая вычислительная сложность и необходимость соблюдения определенных условий. Данные алгоритмы целесообразно сочетать с методами обнаружения, для корректного отслеживания объектов в видеопоследовательности.

В процессе исследования методов распознавания объектов [17–26] были установлены предпочитаемые качественные характеристики, такие как точность распознавания и скорость работы.

В предлагаемом подходе для детектирования и распознавания руки в видеопоследовательности используются методы машинного обучения, поскольку данные, получаемые с web-камеры, имеют высокий уровень шумов – это делает указанные выше алгоритмы обнаружения руки на изображении неэффективными без использования особых условий или оборудования.

Оценка эффективности системы распознавания жестов производится на изображениях и видеопоследовательности, в ходе которой с помощью матриц неточностей рассчитано количество допускаемых ошибок. Для оценки эффективности классификаторов использован метод перекрестной проверки.

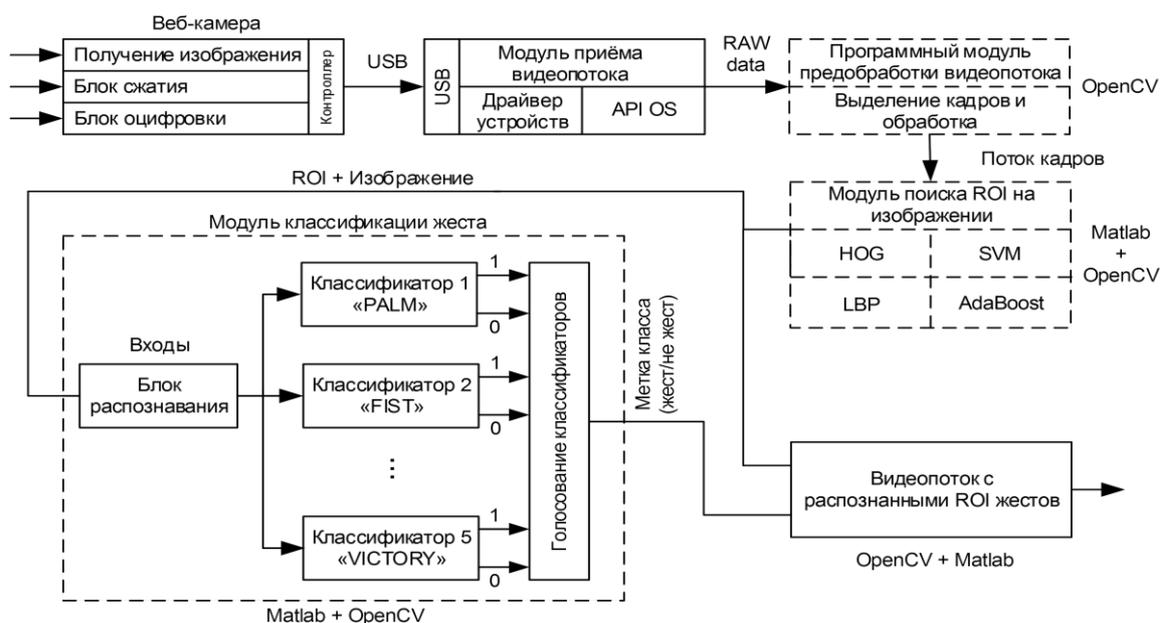


Рис. 1. Структура системы распознавания жестов

## РАЗРАБОТКА СТРУКТУРЫ СИСТЕМЫ РАСПОЗНАВАНИЯ ЖЕСТОВ

Структурная схема системы распознавания жестов с использованием web-камеры представлена на рис. 1.

Система состоит из web-камеры, прикладной программы с нейросетевым модулем распознавания и устройства вывода информации. После снятия изображения данные передаются по интерфейсу USB с последующей загрузкой в модуль приема видеопотока, который далее передает необработанные кадры (raw data) в программный модуль предобработки видеопотока. После предобработки происходит выделение кадров и их обработка для поиска области интереса (Region of Interest, ROI) на изображении. Как только ROI найдена, осуществляется поиск признаков гистограммы направленных градиентов (Histogram of Oriented Gradients, HOG) [18] или локальных бинарных шаблонов (Local Binary Patterns, LBP) [27] для дальнейшей классификации жеста в модуле. Набор признаков должен быть достаточным для идентификации класса объекта, и он представляется в виде n-мерного вектора. Сам классификатор выдаст «1», если жест распознан, «0» если нет. Голосование классификаторов выполняется выбором классификатора с максимальным откликом, затем полученный результат выводится на экран компьютера.

## РАЗРАБОТКА АЛГОРИТМА РАСПОЗНАВАНИЯ ЖЕСТОВ НА ОСНОВЕ НЕЙРОСЕТЕВЫХ ТЕХНОЛОГИЙ

Алгоритм работы системы распознавания жестов представлен на рис. 2.

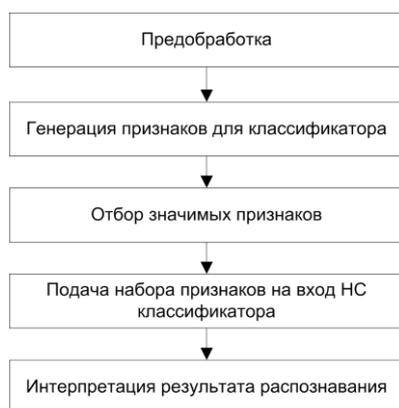


Рис. 2. Алгоритм распознавания жестов

На этапе предобработки осуществляется подготовка двух обучающих выборок: позитивной и негативной. Дополнительно на данном шаге осуществляется перевод цветного изображения в оттенки серого, и выравнивание гистограммы, если жест распознается в видеопоследовательности.

Генерация признаков для классификаторов является важным этапом, на котором формируется вектор признаков. В данном случае в качестве признаков используется гистограмма направленных градиентов, поскольку обладает достаточной точностью, а также локальные бинарные шаблоны, которые выигрывают в скорости обучения у Каскадов Хаара [28].

Главной концепцией алгоритма гистограмм направленных градиентов (histogram of oriented gradients) является возможность представления объекта на изображении описанием распространения интенсивности градиентов или направлением контуров [18].

Локальные бинарные шаблоны (local binary patterns) являются описанием окрестности пикселя на изображении в двоичном представлении. При этом задействованный оператор локального бинарного шаблона использует восемь пикселей, которые окружают центральный, принимая значение интенсивности пикселя в центре в качестве некоторого «порогового значения» [29].

При отборе значимых признаков существуют два подхода: когда нет необходимости в отборе значимых признаков и когда такая необходимость есть. Первый подразумевает, что жест будет найден на изображении как область интереса и идентифицирован одним из нескольких классификаторов. В таком случае целесообразно использовать метод опорных векторов (support vector machine, SVM), поскольку SVM может обработать все признаки HOG дескриптора целиком. Либо алгоритмы усиления композиции классификаторов (AdaBoost), которые справляются с большим количеством признаков локальных бинарных шаблонов.

Во втором случае из большого объема признаков формируется компактный вектор, который будет использоваться в качестве вектора признаков образца. Отбираются

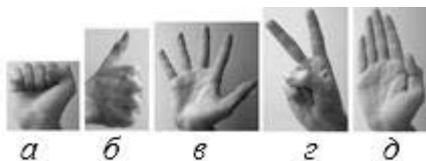
наиболее подходящие из всех признаки, причем неподходящие не удаляются, а используются при дальнейших отборах.

В ходе экспериментов с различными стратегиями формирования вектора признаков и отбора значимых признаков выбраны следующие дескрипторы и классификаторы: HOG+SVM и LBP+AdaBoost. Для снижения размерности вектора признаков решено использовать ассоциативную сеть с узким горлом и многослойного перцептрона (МЛП) с вероятностным методом главных компонент.

### РЕАЛИЗАЦИЯ АЛГОРИТМА РАСПОЗНАВАНИЯ ЖЕСТОВ

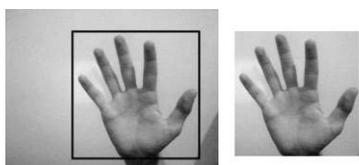
Обучение классификаторов для модуля распознавания можно разделить на следующие этапы: сбор необходимых данных для формирования обучающей, негативной и тестовой выборки, разметка обучающей выборки, обучение классификатора.

Для построения позитивного множества были выбраны жесты, которые представлены на рис. 3.



**Рис. 3.** Типы используемых жестов:  
а – кулак; б – большой палец вверх;  
в – открытая ладонь; г – «Виктория»;  
д – закрытая ладонь

Всего было сделано 2500 фотографий, отобрано по 500 на каждый жест. Изображения были получены под разными углами наклона объекта в меняющихся условиях освещения, а также с однородным фоном. Изначально изображения имели размер  $320 \times 240$ , затем были размечены и обрезаны таким образом, чтобы содержать минимальное количество пикселей за пределами жеста. Пример отображен на рис. 4.



**Рис. 4.** Выделение области интереса (ROI)

Негативная выборка состоит из 1000 фотографий с разрешением  $320 \times 240$ , пример некоторых из них представлен на рис. 5.



**Рис. 5.** Негативные образцы

Очень важно, чтобы в негативном множестве изображений не было жестов. Также желательно учитывать соотношение позитивных изображений к негативным – 1:2.

Первые два этапа просты, но трудоемки, поскольку для каждого изображения необходимо выделить интересующую область. Причем полностью автоматизировать данный процесс нельзя.

После создания обучающих выборок классификаторы обучались со следующими параметрами: минимальная доля верно классифицированных жестов обучающей выборки для каждого уровня каскада: 0,90 для LBP, 0,95 для HOG. Максимальная доля ложноположительных срабатываний на изображениях без жестов – 0,4 для LBP и 0,5 для HOG. Разрешение для каждого жеста было подобрано индивидуально, с учетом пропорций. Изначально количество стадий обучения для всех классификаторов составляло 15, но из-за эффекта переобучения количество этапов для конкретных классификаторов пришлось уменьшить. Время обучения классификаторов на компьютере с процессором Intel Core i5 с частотой 2,4 GHz и 4GB ОЗУ составило 7 часов. Стоит отметить, что обучение классификаторов с использованием каскадов Хаара заняло бы значительно больше времени.

Необходимо учитывать эвристические методы, которые будут влиять на качество распознавания жестов, поскольку алгоритмы искусственного интеллекта и компьютерного зрения зависят от различных эвристических параметров: к примеру, подбор количества нейронов в скрытом слое или выставление уровня ложных тревог при обучении классификатора. Настройка данных параметров является длительным процессом, который требует учета исходных выборок для получения лучшего результата [30, 31].

Входной информацией является видеопоток с web-камеры. В видео может присутствовать частично зашумленный фон и движение на заднем плане, это практически не влияет на работоспособность алгоритма. Разные типы освещения также не воздействуют на эффективность распознавания жестов.

Входное видео имеет разрешение 640×480 пикселей и при обработке 10 кадров в секунду, распознавание одного жеста необходимо выполнять за 0,1 секунду. В результате за минуту обрабатывается 600 кадров, каждый из которых содержит 640×480=307200 пикселей. Учитывая то, что предлагаемый алгоритм анализирует изображение попиксельно, без эффективной аппаратной реализации время работы алгоритма существенно возрастет.

На рис. 6 изображена блок-схема алгоритма работы модуля распознавания жеста в видеопоследовательности

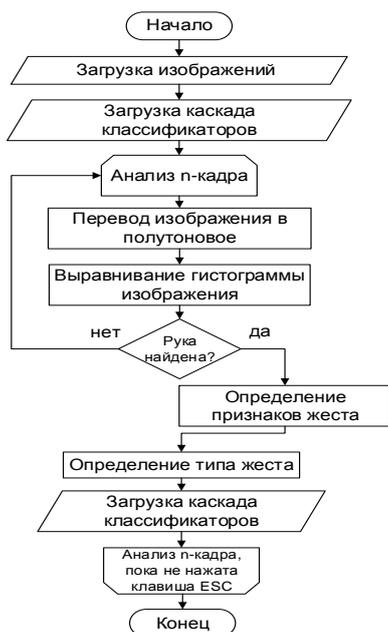


Рис. 6. Блок-схема алгоритма работы модуля распознавания жестов в видеопотоке

На первом шаге происходит загрузка каскада классификаторов из xml файла, затем начинается захват изображения с web-камеры. Видео разделяется на кадры, каждый из которых переводится в полутоновое изображение. Дополнительно производится выравнивание гистограммы изображения для нормализации определенных участков кадров с различными яркостями.

После преобразований происходит поиск руки на изображении с использованием многомасштабного детектирования, затем осуществляется определение признаков жеста для дальнейшего установления его типа.

Алгоритм распознавания жестов в последовательности кадров практически идентичен блок-схеме на рис. 6, единственное различие в алгоритме – это отсутствие шага выравнивания гистограммы изображения, что в данном случае не выполняется для анализа с помощью машины опорных векторов.

### ОЦЕНКА ЭФФЕКТИВНОСТИ ПРОГРАММНО-АППАРАТНОГО МОДУЛЯ РАСПОЗНАВАНИЯ ЖЕСТОВ

Тестирование спроектированной системы распознавания жестов производилось в два этапа: сначала тестировался классификатор на основе комбинации машин опорных векторов на множестве фотографий. Затем тестировался классификатор, использующий локальные бинарные шаблоны на нескольких видеопоследовательностях, причем, в самом видео менялась скорость смены жестов с целью проверки эффективности определения жестов в динамичных сценах. Фото и видео были сделаны в меняющихся условиях освещения и под разными углами наклона объекта.

Далее приведены результаты испытаний пяти классификаторов жестов SVM в три этапа:

- 1) тест на обучающей выборке при перекрестной проверке (табл. 1);
- 2) тест на обучающей выборке (табл. 2);
- 3) тест на тестовой выборке (табл. 3).

Первые два этапа – это тестирование на множестве из 1400 негативных и 350 позитивных фото, а третий – на 150 изображениях позитивной тестовой выборки.

Таблица 1

#### Тест на обучающей выборке при перекрестной проверке

		Предсказанный				
		Класс	1	2	3	4
Исходный	1	348	3	2	0	4
	2	0	344	1	2	1
	3	2	0	346	1	5
	4	0	2	0	347	0
	5	0	1	1	0	340

Элементы на главной диагонали матрицы характеризуют количество верно классифицированных жестов, все числа, отличные от нуля за пределами главной диагонали показывают количество ошибочных классификаций. «Предсказанный» – класс, к которому классификатор отнес образец, а «исходный» – к какому он на самом деле относится.

Процент ошибочно распознанных изображений по матрице неточностей:

$$\frac{348+344+346+347+340}{350*5} \times 100\% = \frac{1725}{1750} \times 100\% = 98,57\% - 100\% = 1,43\%.$$

Матрица неточностей по результатам тестирования на обучающей выборке:

Таблица 2

## Тест на обучающей выборке

		Предсказанный				
		Класс	1	2	3	4
Исходный	1	<b>346</b>	3	1	0	1
	2	0	<b>344</b>	0	2	0
	3	2	2	<b>348</b>	2	1
	4	2	0	1	<b>343</b>	0
	5	0	1	0	3	<b>348</b>

Ошибки классификации по полученной матрице составляют:

$$\frac{346+344+348+343+348}{350*5} \times 100\% = \frac{1729}{1750} \times 100\% = 98,8\% - 100\% = 1,2\%.$$

Матрица неточностей по результатам тестирования на тестовой выборке приведена в табл. 3.

Таблица 3

## Тест на тестовой выборке

		Предсказанный				
		Класс	1	2	3	4
Исходный	1	<b>143</b>	1	0	0	0
	2	0	<b>145</b>	0	2	0
	3	4	2	<b>149</b>	0	1
	4	2	0	1	<b>148</b>	0
	5	1	2	0	0	<b>149</b>

Количество ошибок классификации по полученной матрице:

$$\frac{143+145+149+148+149}{150*5} \times 100\% = \frac{734}{750} \times 100\% = 97,86\% - 100\% = 2,14\%.$$

Количество верных и ошибочных классификаций при распознавании жестов путем классификации гистограмм направленных градиентов с помощью SVM показано в табл. 4.

Таблица 4

## Ошибки при распознавании жестов

		Предсказанный				
		Класс	1	2	3	4
Исходный	1	<b>98</b>	0	0	0	3
	2	0	<b>97</b>	1	2	0
	3	0	0	<b>98</b>	0	0
	4	2	1	0	<b>96</b>	0
	5	0	2	1	2	<b>97</b>

Количество ошибок при распознавании жестов составляет:

$$\frac{98+97+98+96+97}{100*5} \times 100\% = \frac{486}{500} \times 100\% = 97,2\% - 100\% = 2,8\%.$$

Качество распознавания с помощью алгоритма адаптивного усиления и локальных бинарных шаблонов оценивалось в видеопоследовательности. Было записано пять разных видео продолжительностью 30 секунд. При 10 кадрах в секунду получено 300 кадров в каждом из видео.

Для того чтобы посчитать количество ложных срабатываний, необходимо разложить видео на кадры. Для этих целей использовалась бесплатная программа Free Video to JPG Converter. На рис. 7 отображено два ложных срабатывания.

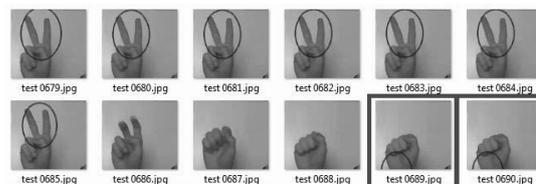


Рис. 7. Пример ложного срабатывания на кадрах видео

Матрица неточностей для классификатора видеопоследовательности приведена в табл. 5.

Таблица 5

**Классификация жестов  
в видеопоследовательности**

		Предсказанный				
		1	2	3	4	5
Исходный	Класс					
	1	<b>292</b>	3	0	4	2
	2	4	<b>293</b>	2	5	0
	3	0	2	<b>295</b>	0	6
	4	2	0	0	<b>291</b>	2
5	2	2	3	0	<b>290</b>	

Количество ошибочно распознанных жестов в процентном соотношении составляет:

$$\frac{292+293+295+291+290}{300 \cdot 5} \times 100\% = \frac{1461}{1500} \times 100\% = 97,4\% - 100\% = 2,6\%.$$

Количество ошибок отдельных классификаторов представлено на рис. 8.

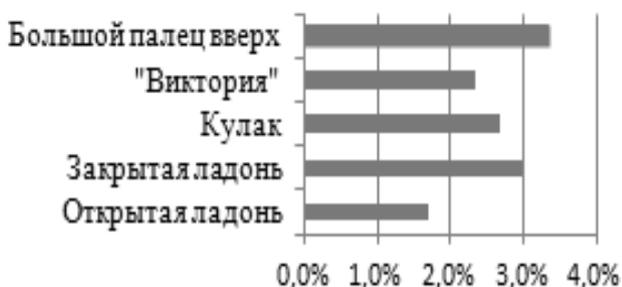


Рис. 8. Ошибки классификаторов

Меньшее количество ошибок получается при распознавании «открытой ладони». Это можно объяснить тем, что для «открытой ладони» число выделенных признаков при обучении больше, чем у остальных.

Для сжатия вектора признаков использовалась ассоциативная нейронная сеть с узким горлом, которая не смогла обработать крупный массив векторов (ввиду слишком большой размерности входных векторов, что приводило к насыщению выходов нейронов первого слоя и бинаризации сжимаемого вектора признаков, что существенно снижало количество сохраняемой информации). Однако с этой задачей хорошо справился вероятностный метод главных компонент, результаты работы которого отображены в табл. 6.

Таблица 6

**Результат работы вероятностного метода  
главных компонент (ВМГК)  
и многослойного персептрона (МЛП)**

№ прохода	Ошибка на обучающей выборке	Ошибка на тестовой выборке
1	0,2%	2,1%
2	0,5%	2,4%
3	1,3%	2,2%
4	1,9%	2,6%
5	2,4%	2,8%
6	2,7%	2,4%
7	3,1%	3,5%
8	1,7%	3,2%
9	2,8%	4,1%
10	2,9%	4,3%
<b>Среднее:</b>	<b>1,95%</b>	<b>2,96%</b>

На рис. 9 показан график «вклада» каждой компоненты в общую дисперсию. Ось X – номер компоненты, Y – вклад каждой компоненты в суммарную дисперсию выборки.

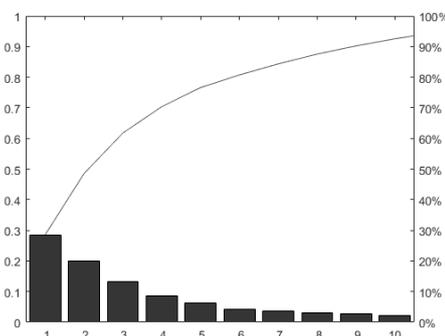


Рис. 9. Вклад компоненты в суммарную дисперсию выборки

Из графика видно, что 10 компонент обеспечивают до 90% дисперсии исходных данных. Было решено использовать 15 компонент и сжать каждое изображение в вектор с таким же количеством элементов. ВМГК использовался вместе с многослойным персептроном, конфигурация которого включает: 12 нейронов в скрытом слое, гиперболический тангенс в качестве функции активации нейронов, 1000 эпох обучения и алгоритм обучения на основе обратного распространения ошибки. Данный подход имеет довольно высокие требования аппаратных ресурсов для вычислений матрицы проецирования с помощью ВМГК. Так, про-

граммная реализация алгоритма ВМГК в течение 20 часов нагружал центральный процессор второй рабочей машины на 90% и потребляла до 9 гигабайт оперативной памяти.

### ЗАКЛЮЧЕНИЕ

В результате численных экспериментов было установлено, что разработанный алгоритм уменьшает количество ошибок во время распознавания жестов. Он является универсальным, поскольку может быть дополнен базой зашумленных изображений для увеличения точности распознавания на сложных изображениях и в видеопоследовательности.

В работе [32] количество ошибок при распознавании в видеопоследовательности составило 6,6%, в [33] – 8,2%, а в [34] – 8%. В данной работе в результате применения HOG+SVM количество ошибок составило 2,8% и 2,6% при использовании LBP + AdaBoost. Предложенный алгоритм снизил уровень ошибок при распознавании жестов в видеопоследовательности web-камеры и может использоваться при создании естественных человеко-машинных интерфейсов, специальных систем для глухонемых людей, а также применяться для управления программным обеспечением на компьютере посредством жестов.

### СПИСОК ЛИТЕРАТУРЫ

1. Leap Motion, Kinect, Oculus [Электронный ресурс]. URL: <https://vc.ru/p/natural-user-interfaces> (дата обращения 16.01.2017). [ (2017, Jan. 16). *Leap Motion, Kinect, Oculus* [Online], (in Russian). Available: <https://vc.ru/p/natural-user-interfaces> ]
2. Touch+ make any surface multitouch and more [Электронный ресурс]. URL: <http://www.ractiv.com/> (дата обращения 16.01.2017). [ (2017, Jan. 16). *Touch+ make any surface multitouch and more* [Online]. Available: <http://www.ractiv.com/> ]
3. Softkinetic, Wikipedia [Электронный ресурс]. URL: <https://en.wikipedia.org/wiki/Softkinetic> (дата обращения 16.01.2017). [ (2017, Jan. 16). *Softkinetic, Wikipedia* [Online]. Available: <https://en.wikipedia.org/wiki/Softkinetic> ]
4. Xtion PRO [Электронный ресурс]. URL: [https://www.asus.com/3D-Sensor/Xtion\\_PRO/](https://www.asus.com/3D-Sensor/Xtion_PRO/) (дата обращения 16.01.2017). [ (2017, Jan. 16). *Xtion PRO* [Online]. Available: [https://www.asus.com/3D-Sensor/Xtion\\_PRO/](https://www.asus.com/3D-Sensor/Xtion_PRO/) ]
5. Как управлять компьютером жестами с помощью веб-камеры [Электронный ресурс]. URL: <http://www.lookatme.ru/mag/how-to/inspiration-howitworks/199949-gesture-control> (дата обращения 16.01.2017). [ (2017, Jan. 16). *How to control a computer with gestures using a webcam* [Online], (in Russian). Available: <http://www.lookatme.ru/mag/how-to/inspiration-howitworks/199949-gesture-control> ]
6. Embedded computer vision [Электронный ресурс]. URL: <http://www.eyesight-tech.com/> (дата обращения 16.01.2017). [ (2017, Jan. 16). *Embedded computer vision* [Online]. Available: <http://www.eyesight-tech.com/> ]
7. How to control your Mac with hand gestures [Электронный ресурс]. URL: <http://www.macworld.co.uk/how-to/mac-software/control-you-mac-with-hand-gestures-3598714/> (дата обращения 16.01.2017). [ (2017, Jan. 16). *How to control your Mac with hand gestures* [Online]. Available: <http://www.macworld.co.uk/how-to/mac-software/control-you-mac-with-hand-gestures-3598714/> ]
8. Алгоритм и методы обнаружения и распознавания жестов руки на видео в режиме реального времени [Электронный ресурс]. URL: <http://sj.kubsau.ru/2014/03/20.pdf> (дата обращения 28.09.2016). [ (2016, Sep. 28). *Algorithm and detection & identification methods for real-time hand gesture detection into the video* [Online], (in Russian). Available: <http://sj.kubsau.ru/2014/03/20.pdf> ]
9. Нахождение контуров и операции с ними [Электронный ресурс]. URL: <http://robocraft.ru/blog/computervision/640.html> (дата обращения 28.09.2016). [ (2016, Sep. 28). *Finding contours and operations with them* [Online], (in Russian). Available: <http://robocraft.ru/blog/computervision/640.html> ]
10. Vision-based Hand Gesture Recognition for Human-Computer Interaction [Электронный ресурс]. URL: <http://www.ics.forth.gr> (дата обращения 28.09.2016). [ (2016, Sep. 28). *Vision-based Hand Gesture Recognition for Human-Computer Interaction* [Online]. Available: <http://www.ics.forth.gr> ]
11. Кузнецов А. И., Сурков А. В., Рудь М. Н. Использование машинного зрения в целях автоматизации и контроля в технических системах [Электронный ресурс]. URL: <http://earchive.tpu.ru/handle/11683/16932> (дата обращения 28.09.2016). [ А. И. Kuznetsov, A. V. Surkov, M. N. Rud (2016, Sep. 28). *Using computer vision to automate and control in technical systems* [Online], (in Russian). Available: <http://earchive.tpu.ru/handle/11683/16932> ]
12. Determining optical flow [Электронный ресурс]. URL: [http://image.diku.dk/imagecanon/material/HornSchunckOptical\\_Flow.pdf](http://image.diku.dk/imagecanon/material/HornSchunckOptical_Flow.pdf) (дата обращения 28.09.2016). [ (2016, Sep. 28). *Determining optical flow* [Online]. Available: [http://image.diku.dk/imagecanon/material/HornSchunckOptical\\_Flow.pdf](http://image.diku.dk/imagecanon/material/HornSchunckOptical_Flow.pdf) ]
13. Шальнов Е., Кононов В., Коношин В. Алгоритм вычитания фона, основанный на поблочных классификаторах // Труды 21-й Международной конференции по компьютерной графике и зрению GraphiCon'2011. GraphiCon. МАКС Пресс Москва, 2011. С. 227–230 [ Е. Shalnov, V. Kononov, V. Konushin, "Background subtraction algorithm based on block-based classifiers", (in Russian), in *21th International Conference on Computer Graphics and Vision GraphiCon'2011*. GraphiCon. MAX Press Moscow, pp. 227-230, 2011. ]
14. Background subtraction techniques: a review [Электронный ресурс]. URL: [http://profs.sci.univr.it/~cristanm/teaching/sar\\_files/lezione4/Piccardi.pdf](http://profs.sci.univr.it/~cristanm/teaching/sar_files/lezione4/Piccardi.pdf) (дата обращения

2.10.2016). [ (2016, Oct. 2). *Background subtraction techniques: a review* [Online]. Available: [http://profs.sci.univr.it/~cristanm/teaching/sar\\_files/lezione4/Piccardi.pdf](http://profs.sci.univr.it/~cristanm/teaching/sar_files/lezione4/Piccardi.pdf) ]

15. Towards Ubiquitous Document Scanning [Электронный ресурс]. URL: [https://robau.files.wordpress.com/2010/06/final\\_report\\_00012.pdf](https://robau.files.wordpress.com/2010/06/final_report_00012.pdf) (дата обращения 2.10.2016). [ (2016, Oct. 2). *Towards Ubiquitous Document Scanning* [Online]. Available: [https://robau.files.wordpress.com/2010/06/final\\_report\\_00012.pdf](https://robau.files.wordpress.com/2010/06/final_report_00012.pdf) ]

16. Анализ изображений и видео Лекция 8: Сегментация изображений [Электронный ресурс]. URL: <https://compscicenter.ru/media/slides/> (дата обращения 2.10.2016). [ (2016, Oct. 2). *Analysis of the images and video Lecture 8: Image segmentation* [Online], (in Russian). Available: <https://compscicenter.ru/media/slides/> ]

17. Перевод книги Learning OpenCV [Электронный ресурс]. URL: <http://roboforum.ru/forum51/topic9153.html> (дата обращения 2.10.2016). [ (2016, Oct. 2). *Translation of the book Learning OpenCV* [Online], (in Russian). Available: <http://roboforum.ru/forum51/topic9153.html> ]

18. Гистограмма направленных градиентов [Электронный ресурс]. URL: [https://ru.wikipedia.org/wiki/Гистограмма\\_направленных\\_градиентов](https://ru.wikipedia.org/wiki/Гистограмма_направленных_градиентов) (дата обращения 2.10.2016). [ (2016, Oct. 2). *Histogram of Oriented Gradients* [Online], (in Russian)..Available: [https://ru.wikipedia.org/wiki/Гистограмма\\_направленных\\_градиентов](https://ru.wikipedia.org/wiki/Гистограмма_направленных_градиентов) ]

19. Алгоритм машинного обучения AdaBoost [Электронный ресурс]. URL: <http://www.machinelearning.ru> (дата обращения 2.10.2016). [ (2016, Oct. 2). *AdaBoost machine learning algorithm* [Online], (in Russian). Available: <http://www.machinelearning.ru> ]

20. Классификация данных методом опорных векторов [Электронный ресурс]. URL: <https://habrahabr.ru/post/105220/> (дата обращения 2.10.2016). [ (2016, Oct. 2). *Support vector machines data classification* [Online], (in Russian). Available: <https://habrahabr.ru/post/105220/> ]

21. Уменьшение размерности в данных [Электронный ресурс]. URL: <http://courses.graphicon.ru/files/courses/smisa/2009/lectures/lecture12.pdf> (дата обращения 15.10.2016). [ (2016, Oct. 15). *Reduction of data dimension* [Online], (in Russian). Available: <http://courses.graphicon.ru/files/courses/smisa/2009/lectures/lecture12.pdf> ]

22. Искусственная нейронная сеть [Электронный ресурс]. URL: [https://ru.wikipedia.org/wiki/Искусственная\\_нейронная\\_сеть](https://ru.wikipedia.org/wiki/Искусственная_нейронная_сеть) (дата обращения 15.10.2016). [ (2016, Oct. 15). *Artificial neural network* [Online], (in Russian). Available: [https://ru.wikipedia.org/wiki/Искусственная\\_нейронная\\_сеть](https://ru.wikipedia.org/wiki/Искусственная_нейронная_сеть) ]

23. **Maraqqa M., Al-Zboun F., Dhyabat M., Zitar R. A.** Recognition of Arabic Sign Language (ArSL) Using Recurrent Neural Networks // *Journal of Intelligent Learning Systems and Applications*, vol. 4, 2012. [ M. Maraqqa, F. Al-Zboun, M. Dhyabat, R.A. Zitar, *Recognition of Arabic Sign Language (ArSL) Using Recurrent Neural Networks* // *Journal of Intelligent Learning Systems and Applications*, vol. 4, 2012. ]

24. **Murakami K., Taguchi H.** Gesture Recognition using Recurrent Neural Networks // *ACM Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology (CHI '91)*, 1999. [ K. Murakami, H. Taguchi, *Gesture Recognition using Recurrent Neural Networks* // *ACM*

*Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology (CHI '91)*, 1999. ]

25. Преимущества нейронных сетей [Электронный ресурс]. URL: <http://www.aiportal.ru/articles/neural-networks/advantages.html> (дата обращения 22.10.2016). [ (2016, Oct. 22). *Advantages of Neural Networks* [Online], (in Russian). Available: <http://www.aiportal.ru/articles/neural-networks/advantages.html> ]

26. Скрытая марковская модель [Электронный ресурс]. URL: [ru.wikipedia.org/wiki/Скрытая\\_марковская\\_модель](http://ru.wikipedia.org/wiki/Скрытая_марковская_модель) (дата обращения 22.10.2016). [ (2016, Oct. 22). *Hidden Markov model* [Online], (in Russian). Available: [ru.wikipedia.org/wiki/Скрытая\\_марковская\\_модель](http://ru.wikipedia.org/wiki/Скрытая_марковская_модель) ]

27. Применение локальных бинарных шаблонов к решению задачи распознавания лиц [Электронный ресурс]. URL: <https://habrahabr.ru/post/193658/> (дата обращения 22.10.2016). [ (2016, Oct. 22). *The use of local binary patterns to solve problems face recognition* [Online], (in Russian). Available: <https://habrahabr.ru/post/193658/> ]

28. Cascade classifier [Электронный ресурс]. URL: [http://docs.opencv.org/trunk/d7/d8b/tutorial\\_py\\_face\\_detection.html](http://docs.opencv.org/trunk/d7/d8b/tutorial_py_face_detection.html) (дата обращения 22.10.2016). [ (2016, Oct. 22). *Cascade classifier* [Online]. Available: [http://docs.opencv.org/trunk/d7/d8b/tutorial\\_py\\_face\\_detection.html](http://docs.opencv.org/trunk/d7/d8b/tutorial_py_face_detection.html) ]

29. Local binary pattern [Электронный ресурс]. URL: <http://docs.opencv.org/2.4/modules/> (дата обращения 22.10.2016). [ (2016, Oct. 22). *Local binary pattern* [Online]. Available: <http://docs.opencv.org/2.4/modules/> ]

30. Tutorial: OpenCV haartraining [Электронный ресурс]. URL: <http://note.sonots.com/SciSoftware/haartraining.html> (дата обращения 22.10.2016). [ (2016, Oct. 22). *Tutorial: OpenCV haartraining* [Online]. Available: <http://note.sonots.com/SciSoftware/haartraining.html> ]

31. Cascade Classifier Training [Электронный ресурс]. URL: [http://docs.opencv.org/2.4/doc/user\\_guide/ug\\_traincascade.html](http://docs.opencv.org/2.4/doc/user_guide/ug_traincascade.html) (дата обращения 22.10.2016). [ (2016, Oct. 22). *Cascade Classifier Training* [Online]..Available: [http://docs.opencv.org/2.4/doc/user\\_guide/ug\\_traincascade.html](http://docs.opencv.org/2.4/doc/user_guide/ug_traincascade.html) ]

32. **Nianjun Liu, Brian C. Lovell**, Gesture Classification Using Hidden Markov Models and Viterbi Path Counting [Электронный ресурс]. URL: <http://staff.itee.uq.edu.au/lovell/aprs/dicta2003/pdf/0273.pdf> (дата обращения 28.10.2016). [ L. Nianjun, C. Brian Lovell (2016, Oct. 28). *Gesture Classification Using Hidden Markov Models and Viterbi Path Counting* [Online]. Available: <http://staff.itee.uq.edu.au/lovell/aprs/dicta2003/pdf/0273.pdf> ]

33. Распознавание жестов на видеопоследовательности в режиме реального времени на основе применения метода Виолы – Джонса, алгоритма CAMShift, вейвлет-преобразования и метода главных компонент [Электронный ресурс]. URL: <http://earchive.tpu.ru/handle/11683/4244> (дата обращения 28.10.2016). [ (2016, Oct. 28). *Real-time hand gesture recognition base on Viola –Jones method, algorithm CAMShift, wavelet transform and principal component analysis* [Online], (in Russian). Available: <http://earchive.tpu.ru/handle/11683/4244> ]

34. **Нгуен Т. Т., Спицын В. Г.**, Алгоритмическое и программное обеспечение для распознавания формы руки в

реальном времени с использованием SURF дескрипторов и нейронной сети [Электронный ресурс]. URL: [http://www.lib.tpu.ru/fulltext/v/Bulletin\\_TPU/2012/v320/i5/09.pdf](http://www.lib.tpu.ru/fulltext/v/Bulletin_TPU/2012/v320/i5/09.pdf) (дата обращения 28.10.2016). [ Т. Т. Nguen, V. G. Spitsin (2016, Oct. 28). *Algorithmic and the software to recognize the shape of hands in real-time using SURF descriptors and neural network* [Online], (in Russian). Available: [http://www.lib.tpu.ru/fulltext/v/Bulletin\\_TPU/2012/v320/i5/09.pdf](http://www.lib.tpu.ru/fulltext/v/Bulletin_TPU/2012/v320/i5/09.pdf) ]

#### ОБ АВТОРАХ

**ЧУЙКОВ Артем Валерьевич**, студент каф. ВТиЗИ УГАТУ. Дипл. бакалавра по направлению: вычислительные машины, комплексы, системы и сети (УГАТУ, 2016). Иссл. в обл. нейросетевых технологий и компьютерного зрения.

**ВУЛЬФИН Алексей Михайлович**, доцент каф. ВТиЗИ УГАТУ. Дипл. инженера-программиста (УГНТУ, 2008). Кандидат технических наук по системному анализу, управлению и обработке информации (УГАТУ, 2012). Иссл. в обл. интеллектуального анализа данных и моделирования сложных технических систем.

#### METADATA

**Title:** Gesture recognition system.

**Authors:** A. V. Chuikov<sup>1</sup>, A. M. Vulfin<sup>2</sup>

**Affiliation:** Ufa State Aviation Technical University (UGATU), Russia.

**Email:** <sup>1</sup>artem.guy@gmail.com, <sup>2</sup>vulfin.alexey@yandex.ru

**Language:** Russian.

**Source:** Vestnik UGATU (scientific journal of Ufa State Aviation Technical University), vol. 21, no. 3 (77), pp. 113–122, 2017. ISSN 2225–2789 (Online), ISSN 1992-6502 (Print).

**Abstract:** The goal of this research is to improve the algorithms for gesture recognition based on the analysis of a two-dimensional space model using neural network technologies. Completed analytical review of existing software solutions and approaches to the implementation the gesture recognition systems, including detection, tracking and recognition stage. Also block diagrams of the gesture recognition module were drawn up, appropriate descriptors and classifiers have been selected for better results. To reduce the dimensionality of the feature vector, it was decided to employ a special associative network with a bottleneck and multilayer perceptron with a probabilistic principal component analysis. Also evaluated the efficiency of gesture recognition module.

**Key words:** gesture recognition; neural networks; support vector machines; local binary patterns; AdaBoost; histogram of oriented gradients; OpenCV; MATLAB.

#### About authors:

**CHUIKOV, Artem Valerievich**, student dept. of computing equipment and information protection (USATU), bachelor's degree of computer systems and networks (USATU, 2016).

**VULFIN, Alexey Mikhailovich**, prof., dept. of computing equipment and information protection, software engineer dipl. (UGNTU, 2008). cand. of tech. sci., systems analyst, councils recommend measures and information processing (USATU, 2012).