УДК 004.392, 004.93'1

Н.Х. Фан, Т.Т. Буй, В.Г. Спицын

РАСПОЗНАВАНИЕ ЖЕСТОВ НА ВИДЕОПОСЛЕДОВАТЕЛЬНОСТИ В РЕЖИМЕ РЕАЛЬНОГО ВРЕМЕНИ НА ОСНОВЕ ПРИМЕНЕНИЯ МЕТОДА ВИОЛЫ – ДЖОНСА, АЛГОРИТМА CAMSHIFT, ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ И МЕТОДА ГЛАВНЫХ КОМПОНЕНТ

Предложен новый алгоритм распознавания жестов на цифровых изображениях, основанный на совместном применении вейвлет-преобразования и метода главных компонент. Представлены результаты тестирования работы предложенного алгоритма. Показано, что использование указанного алгоритма дает возможность эффективного распознавания жестов на цифровых изображениях. Предложен оригинальный комплексный алгоритм, основанный на методе Виолы – Джонса, алгоритме *CAMShift*, вейвлет-преобразовании и методе главных компонент, предназначенный для распознавания жестов на видеопоследовательности. На основе проведенных численных экспериментов установлено, что предложенный алгоритм позволяет распознавать жесты на видеопоследовательности в режиме реального времени.

Ключевые слова: *Распознавание жестов*, *метод Виолы – Джонса*, *алгоритм CAMShift*, *вейвлет-преобразование*, *метод главных компонент.*

Распознавание жестов является одной из наиболее сложных и актуальных задач в области обработки изображений. Системы распознавания жестов предназначены для идентификации определенных человеческих жестов с целью использования их для передачи информации или для управления различными устройствами. В данной работе рассматривается задача распознавания жестов на цифровых изображениях и видеопоследовательности в режиме реального времени.

Для решения задачи распознавания объектов на видеопоследовательности необходимо решить задачу поиска и отслеживания объектов. Метод Виолы – Джонса [1, 2] является самым популярным методом для поиска области объектов на изображении, из-за его высокой скорости и эффективности. Детектор Виолы – Джонса основан на трех главных идеях: интегральном представлении изображения, методе построения классификатора на основе алгоритма адаптивного бустинга (*AdaBoost*) и методе комбинирования классификаторов в каскадную структуру. Эти идеи позволяют построить детектор, способный работать в режиме реального времени. Информация о статистическом распределении цветовой информации изображения также нашла применение в алгоритмах отслеживания объектов. Так, в 1998 г. Гарри Брадски создал алгоритм *CAMShift* (***C****ontinuously* ***A****daptive* ***M****ean****S****hift*) [3], который на основе цветовой информации был способен отслеживать объекты.

Процесс распознавания объектов обычно состоит из двух этапов: первый этап – извлечение и сохранение признаков известных объектов в базу данных, второй этап – сравнение признаков объектов с признаками, находящимися в базе данных. В настоящее время установлено, что вейвлет-преобразование является хорошим способом для получения характеристик изображения. В данной работе используются вейвлет-преобразования Хаара и Добеши для извлечения признаков жестов

на изображениях. В задаче распознавания объектов метод главных компонент успешно применяется в процессе сравнения компонент, характеризующих неизвестное изображение, с компонентами, соответствующими известным изображениям.

Целью данной работы являются создание нового алгоритма, основанного на применении вейвлет-преобразования и метода главных компонент для распознавания жестов на цифровых изображениях, и разработка оригинального комплексного алгоритма, основанного на применении метода Виолы – Джонса, алгоритма *CAMShift*, вейвлет-преобразования и метода главных компонент для распознавания жестов на видеопоследовательности в режиме реального времени.

1. Метод Виолы – Джонса

Метод был разработан и представлен в 2001 г. Полом Виолой и Майклом Джонсом и до сих пор эффективен для поиска объектов на цифровых изображениях и видеопоследовательностях в режиме реального времени [1, 2]. Основной его идеей является использование каскада простых классификаторов – детекторов характеристик вместо одного сложного классификатора. На базе этой идеи возможно построение детектора, способного работать в режиме реального времени. Характеристики используются вместо непосредственных значений пикселей по многим причинам. Основной причиной является то, что характеристики могут описывать те знания о классе объектов, которые трудно выявить на конечном числе обучающих данных. Вторая важная причина использования характеристик: системы, построенные на их основе, работают гораздо быстрее, чем системы, работающие напрямую с пикселями.

1.1. Интегральное представление изображений

Для того чтобы рассчитывать яркость прямоугольного участка изображения, используется интегральное представление [4]. Оно часто используется и в других методах, например в вейвлет-преобразованиях, *Speeded up robust feature* (*SURF*), фильтрах Хаара и многих разработанных алгоритмах. Интегральное представление позволяет быстро рассчитывать суммарную яркость произвольного прямоугольника на данном изображении, причем время расчета не зависит от площади прямоугольника.

Интегральное представление изображения представляет собой матрицу, совпадающую по размерам с исходным изображением. В каждом ее элементе хранится сумма интенсивностей всех пикселей, находящихся левее и выше данного элемента. Элементы матрицы рассчитываются по следующей формуле:

$$I(x, y) = \sum_{x' \le x, y' \le y} i(x', y'),$$

где $I(x,y)$ – значение точки (x,y) интегрального изображения; $i(x,y)$ – значение интенсивности исходного изображения. На основе применения интегрального представления изображения вычисление признаков одинакового вида, но с разными геометрическими параметрами, происходит за одинаковое время.

Каждый элемент матрицы $I(x,y)$ представляет собой сумму пикселей в прямоугольнике от $i(0,0)$ до $i(x,y)$, т.е. значение каждого элемента $I(x,y)$ равно сумме значений всех пикселей левее и выше данного пикселя $i(x,y)$. Расчет матрицы занимает линейное время, пропорциональное числу пикселей в изображении, и его

можно производить по следующей формуле:

$$I(x, y) = i(x, y) - I(x-1, y-1) + I(x, y-1) + I(x-1, y).$$

Интегральное представление имеет интересную особенность. По интегральной матрице можно очень быстро вычислить сумму пикселей произвольного прямоугольника.

1.2. Хаар-подобные характеристики

С точки зрения необходимости использования достаточно простых алгоритмов получения признаков, перспективным является применение хаар-подобных характеристик (*Haar wavelet-like features*), представляющих собой результат сравнения яркостей в двух прямоугольных областях изображения. В частности, как уже отмечалось выше, Виола и Джонс предложили использовать три вида характеристик. Значением характеристики из двух прямоугольников является разница между суммой пикселей в этих прямоугольных областях. Области имеют одинаковый размер и форму и по горизонтали и по вертикали.

Предположим, что задано множество объектов A и множество допустимых ответов B. Пусть $g{:}A{\to}B$ называется решающей функцией. Решающая функция g должна допускать эффективную компьютерную реализацию, по этой причине её также называют алгоритмом. Признак (*feature*) f объекта a – отображение $f{:}A{\to}D_f$, где D_f – множество допустимых значений признака. В частности, любой алгоритм $g{:}A{\to}B$ также можно рассматривать как признак. Если задан набор признаков $f_1,\ldots,f_n$, то вектор $x = (f_1(a),\ldots,f_n(a))$ называется признаковым описанием объекта $a{\in}A$. Признаковые описания допустимо отождествлять с самими объектами. При этом множество $A = D_{f1}{\times}\ldots{\times}D_{fn}$ называют признаковым пространством [5].

Вычисляемым значением такого признака будет

$$F = U - V,$$

где U – сумма значений яркостей точек, закрываемых светлой частью признака; V – сумма значений яркостей точек, закрываемых темной частью признака. Для их вычисления используется понятие интегрального изображения. Хаар-подобные признаки описывают значение перепада яркости по оси X и Y изображения соответственно.

1.3. Метод построения классификатора на основе алгоритма бустинга

Бустинг – комплекс методов, способствующих повышению точности аналитических моделей. Бустинг (*boosting*) означает дословно «усиление» «слабых» моделей – это процедура последовательного построения композиции алгоритмов машинного обучения, когда каждый следующий алгоритм стремится компенсировать недостатки композиции всех предыдущих алгоритмов. Идея бустинга была предложена Робертом Шапиро (*Schapire*) в конце 90-х гг. прошлого века [6], когда надо было найти решение вопроса о том, каким образом, имея множество плохих (незначительно отличающихся от случайных) алгоритмов обучения, получить один хороший.

В результате работы алгоритма бустинга на каждой итерации формируется простой классификатор вида

$$h_j(z) = \begin{cases} 1, & \text{если } p_j f_j(z) < p_j \theta_j, \\ 0, & \text{иначе,} \end{cases}$$

где p_j – показывает направление знака неравенства; θ_j – значение порога; $f_j(z)$ – вычисленное значение признака; z – окно изображения размером 20×20 пикселей. Полученный классификатор имеет минимальную ошибку по отношению к текущим значениям весов, задействованным в процедуре обучения для определения ошибки.

Развитием данного подхода явилась разработка более совершенного семейства алгоритмов бустинга *AdaBoost* (адаптивное улучшение), осуществленная Йоавом Фройндом и Робертом Шапиро в 1999 г. В *AdaBoost* можно использовать произвольное число классификаторов и производить обучение на одном наборе примеров, поочередно применяя их на различных шагах. В методе Виолы – Джонса вариант *AdaBoost* используется как для выбора особенностей, так и для обучения классификатора. В его оригинальной форме обучающий алгоритм *AdaBoost* используется для повышения эффективности классификации простого (иногда называемого слабым) обучающего алгоритма.

Для повышения скорости обнаружения используется каскадная структура, фокусирующая свою работу на наиболее информативных областях изображения. Каскад состоит из слоев, которые представляют собой классификаторы, обученные с помощью процедуры бустинга.

2. Алгоритм отслеживания объекта *CAMShift*

Алгоритм *CAMShift* был создан Гарри Брадски в 1998 г. и способен отслеживать лица [3]. Он комбинирует алгоритм отслеживания объекта *Mean Shift*, основанный на карте вероятности цвета кожи, с адаптивным шагом изменения размера области отслеживания. Вероятность цвета кожи каждого пикселя изображения определяется методом *Histogram Backprojection*, основанным на цвете, представленном в виде цветового тона (*Hue*) модели *HSV*. Так как алгоритм *CAMShift* способен отслеживать лица на основе вероятности цвета кожи, то он может применяться для отслеживания руки.

Преимуществами данного алгоритма являются: низкие требования к вычислительным ресурсам, гибкие настройки точности позиционирования, возможность работы в различных условиях освещенности. Также дополнительным преимуществом алгоритма является возможность работы в условиях частичного перекрытия отслеживаемого объекта. Указанные выше свойства алгоритма обусловлены использованием модели объекта, построенной на основе гистограммы яркости и цвета, а также использованием процедуры *Mean Shift* для точного позиционирования положения объекта.

3. Вейвлет-преобразование

Вейвлет-преобразование широко используется для анализа нестационарных процессов. Оно показало свою эффективность для решения широкого класса задач, связанных с обработкой изображения. Коэффициенты вейвлет-преобразования содержат информацию об анализируемом процессе и используемом вейвлете. Поэтому выбор анализирующего вейвлета определяется тем, какую информацию необходимо извлечь из процесса. Каждый вейвлет имеет характерные особенности во временной и частотной областях, поэтому иногда с помощью разных вейвлетов можно полнее выявить и подчеркнуть те или иные свойства анализируемого процесса.

В работах [7, 8] представлены разложение изображения и извлечение его признаков для классификации изображений самолетов на основе применения вейвлет-преобразования Хаара и многослойной нейронной сети. В данной работе используются вейвлет-преобразования Хаара и Добеши для извлечения признаков изображения жестов. Пример применения вейвлет-преобразования Добеши для извлечения признаков изображения жеста представлен на рис. 1.

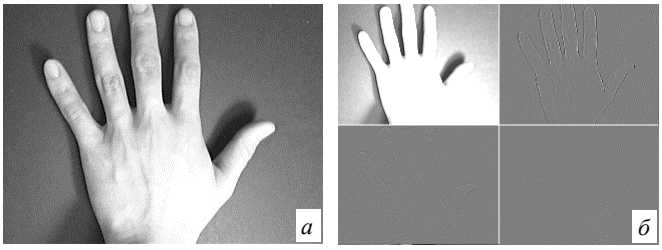

Рис. 1. Пример извлечения признаков жеста: исходное изображение жеста (*а*); результат после применения вейвлет-преобразования Добеши (*б*)

4. Метод главных компонент

Метод главных компонент (*Principal Component Analysis*, *PCA*) – один из наиболее распространенных методов для уменьшения размерности данных, потери наименьшего количества информации. Он заключается в линейном ортогональном преобразовании входного вектора P размерности N в выходной вектор Q размерности M, $M<N$. Компоненты вектора Q являются некоррелированными, и общая дисперсия после преобразования остаётся неизменной.

Вычисление главных компонент сводится к вычислению собственных векторов и собственных значений ковариационной матрицы, которая рассчитывается для изображения. Сумма главных компонент, умноженных на соответствующие собственные векторы, является реконструкцией изображения. Для каждого изображения объекта вычисляются его главные компоненты. Обычно берётся от 5 до 200 главных компонент. Остальные компоненты кодируют мелкие различия между объектами и шум. Процесс распознавания заключается в сравнении главных компонент неизвестного изображения с компонентами всех известных изображений. Из базы данных выбираются изображения-кандидаты, имеющие наименьшее расстояние от входного (неизвестного) изображения [9].

5. Алгоритм распознавания жестов на цифровых изображениях

Целью данной работы является распознавание жестов на цифровых изображениях. Для решения этой задачи предложен новый алгоритм, основанный на применении вейвлет-преобразования и метода главных компонент. Предложенный алгоритм состоит из двух процессов: извлечения и сохранения признаков известных жестов в базе данных и распознавания жестов. Процесс извлечения и сохранения признаков известных жестов происходит следующим образом:

Шаг 1. Преобразование изображения области жеста в полутоновое изображение.

Шаг 2. Изменение размера области жеста до 64×64 пикселей.

Шаг 3. Применение к полученному на шаге 2 изображению вейвлет-преобразования для извлечения признаков жеста (вейвлет-коэффициентов).

Шаг 4. Сохранение извлеченных признаков в базе данных.

В процессе распознавания неизвестного жеста осуществляются шаги 1–3, затем полученные признаки сравниваются с признаками, хранящимися в базе данных, на основе применения метода главных компонент. Функциональная схема предложенного алгоритма представлена на рис 2.

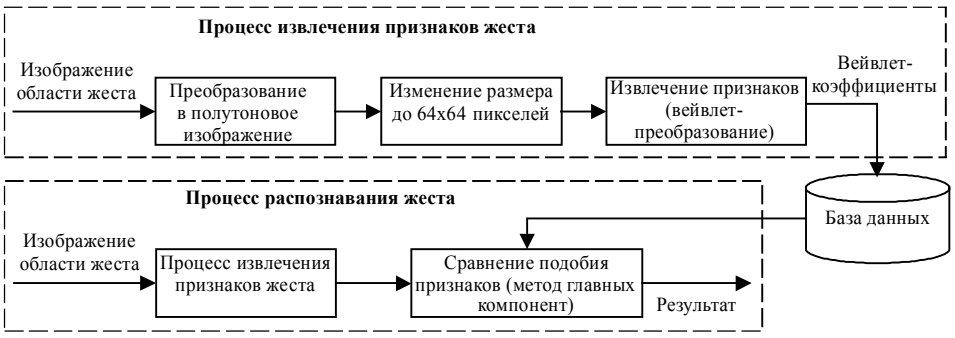

Рис. 2. Функциональная схема предложенного алгоритма распознавания жестов

6. Алгоритм распознавания жестов на видеопоследовательности

В данной работе также рассматривается задача распознавания жестов на видеопоследовательности в режиме реального времени. Для решения этой задачи предложен оригинальный комплексный алгоритм, основанный на применении метода Виолы – Джонса, алгоритма *CAMShift*, вейвлет-преобразования и метода главных компонент. Процесс распознавания жестов на видеопоследовательности происходит следующим образом:

Шаг 1. Запрос очередного видеофрейма. Преобразование видеофрейма в полутоновое изображение. Применение к полутоновому изображению метода Виолы – Джонса для поиска области руки.

Шаг 2. Если область руки обнаружена, то выполняется шаг 3. В обратном случае осуществляется возврат на шаг 1.

Шаг 3. Запрос очередного видеофрейма. Отслеживание области руки на основе применения алгоритма *CAMShift*.

Шаг 4. Если отслеживание осуществлено, то выполняется шаг 5. В обратном случае происходит возврат на шаг 1.

Шаг 5. Выполнение процесса распознавания жеста (рис. 2).

Шаг 6. Возврат на шаг 3.

7. Эксперименты

Для тестирования работы предложенных алгоритмов создано программное обеспечение на языке объектно-ориентированного программирования C# (*Visual studio* 2010) с использованием библиотеки OpenCV. Программа протестирована на ноутбуке с процессором Intel Core™2 Duo 2ГГц, объемом оперативной памяти 2 Гб, видеокамерой 1,3 Мп, передающей 30 кадров в секунду с разрешением 320×240.

При тестировании работы алгоритма распознавания жестов на цифровых изображениях используется часть базы данных *Cambridge Gesture database* [10]. Эта

база изображений жестов состоит из 5 различных частей, изображения в которых получены при различных условиях освещенности (рис. 3). В данной работе, все жесты в базе данных делятся на 12 классов, представленных на рис. 4.

Рис. 3. Примеры изображений жестов каждой части базы данных для тестирования: *а* – часть 1; *б* – 2; *в* – 3; *г* – 4; *д* – 5

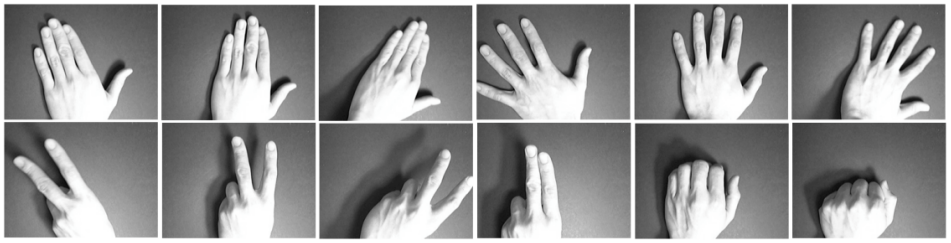

Рис. 4. Примеры 12 классов, использующихся при распознавании жестов на цифровых изображениях

Для каждой части создана база изображений жестов для тестирования, которая содержит 200 изображений каждого класса (всего $12 \times 200 = 2400$ изображений). Смещенная база изображений жестов из 5 частей для тестирования содержит 1000 изображений каждого класса (всего $12 \times 1000 = 12000$ изображений). Для каждой части также создана база изображений жестов для обучения, которая содержит 20 изображений каждого класса (всего $12 \times 20 = 240$ изображений). Смещенная база изображений жестов из 5 частей для обучения содержит 100 изображений каждого класса (всего $12 \times 100 = 1200$ изображений). В процессе тестирования применяются вейвлет-преобразования Хаара и Добеши. В таблице приведены результаты экспериментов по распознаванию жестов на цифровых изображениях.

Результаты распознавания жестов

База данных	Часть 1		Часть 2		Часть 3	
	Хаар	Добеши	Хаар	Добеши	Хаар	Добеши
Достоверность распознаваний жестов, %	94,63	93,67	90,96	90,17	89,46	87,58

База данных	Часть 4		Часть 5		Смещенная часть	
	Хаар	Добеши	Хаар	Добеши	Хаар	Добеши
Достоверность распознаваний жестов, %	92,33	90,79	90,17	87,63	93,30	92,57

При тестировании работы алгоритма распознавания жестов на видеопоследовательности используется 6 классов жестов, представленных на рис. 5. Примеры результатов работы предложенного алгоритма представлены на рис. 6. Скорость работы предложенного алгоритма составляет 30 кадров в секунду.

Рис. 5. Примеры 6 классов, использующихся при распознавании жестов на видеопоследовательности

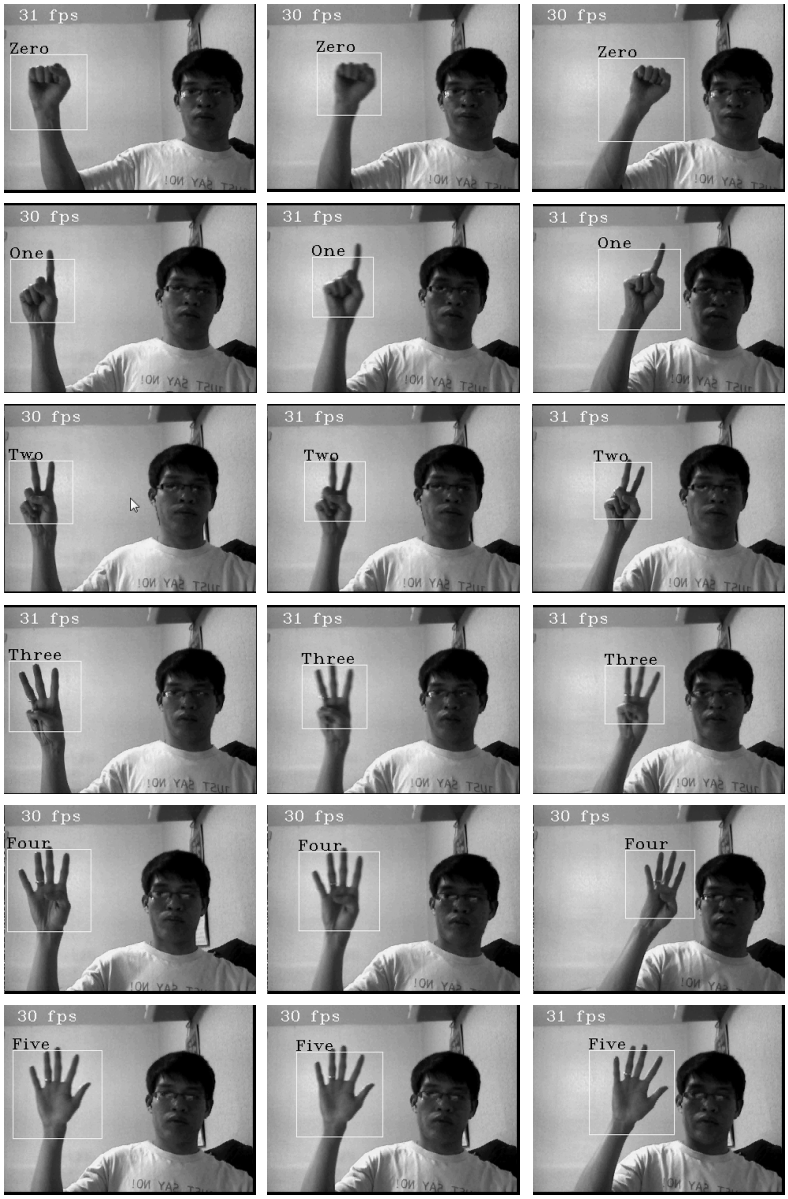

Рис. 6. Результаты работы предложенного алгоритма распознавания жестов на видеопоследовательности

Заключение

Предложен и описан новый алгоритм на основе совместного применения вейвлет-преобразования и метода главных компонент для распознавания жестов на цифровых изображениях. Анализ результатов численных экспериментов позволяет сделать вывод об успешной работе созданного алгоритма, основанного на применении вейвлет-преобразования и метода главных компонент, при распознавании жестов.

Предложен и описан оригинальный комплексный алгоритм, основанный на применении метода Виолы – Джонса, алгоритма CAMShift, вейвлет-преобразования и метода главных компонент для распознавания жестов на видеопоследовательности. Результаты проведенных компьютерных экспериментов показали, что предложенный оригинальный комплексный алгоритм позволяет эффективно распознавать жесты на видеопоследовательности в режиме реального времени.

ЛИТЕРАТУРА

1. *Viola P.*, *Jones M.J.* Rapid object detection using a boosted cascade of simple features // IEEE Conf. on Computer Vision and Pattern Recognition. Kauai, Hawaii, USA, 2001. V. 1. P. 511–518.
2. *Viola P.*, *Jones M.J.* Robust real-time face detection // International Journal of Computer Vision. 2004. V. 57. No. 2. P. 137–154.
3. *Bradski G.R.* Computer vision face tracking for use in a perceptual user interface // Intel Technology Journal. 1998, 2nd Quarter.
4. *Гонсалес Р.*, *Вудс Р.* Цифровая обработка изображений. М.: Техносфера, 2005. 1072 с.
5. *Местецкий Л.М.* Математические методы распознавания образов. М.: МГУ, ВМиК, 2002–2004. С. 42–44.
6. *Freund Y.*, *Schapire R.E.* A Short introduction to boosting // J. Japanese Society for Artifical Intelligence. September 1999. V. 14. No. 5. P. 771–780.
7. *Буй Тхи Тху Чанг*, *Спицын В.Г.* Разложение цифровых изображений с помощью двумерного дискретного вейвлет-преобразования и быстрого преобразования // Известия Томского политехнического университета. 2011. Т. 318. № 5. С. 73–76.
8. *Буй Тхи Тху Чанг*, *Фан Нгок Хоанг*, *Спицын В.Г.* Алгоритмическое и программное обеспечение для классификации цифровых изображений с помощью вейвлет-преобразования Хаара и нейронных сетей // Известия Томского политехнического университета. 2011. Т. 319. № 5. С. 103–106.
9. Pearson K. On lines and planes of closest fit to systems of points in space // Philosophical Magazine. 1901. V. 2. No. 6. P. 559–572.
10. *Kim T.K.*, *Wong S.F.*, *Cipolla R.* Cambrige Hand Gesture Data set. URL: http://www.iis.ee.ic.ac.uk/~tkkim/ges_db.htm (дата обращения 10.02.2012).

Фан Нгок Хоанг, *Буй Тхи Тху Чанг*
Спицын Владимир Григорьевич
Томский политехнический университет,
E-mail: hoangpn285@gmail.com;
trangbt.084@gmail.com; spvg@tpu.ru

Поступила в редакцию 29 апреля 2012 г.

Phan N.H., *Bui T.T.T.*, *Spitsyn Vladimir G.* (Tomsk Polytechnic University). **Real-time hand gesture recognition base on Viola–Jones method, algorithm CAMShift, wavelet transform and principal component analysis.**

Keywords: Hand gesture recognition, method Viola–Jones, algorithm CAMShift, wavelet transform, principal component analysis.

The task of hand gesture recognition on digital images and in video sequence in real-time is considered. A novel algorithm based on wavelet transform and principal component analysis is

proposed for hand gesture recognition on digital images. The experiment results show that the proposed algorithm has the high rate of hand gesture recognition.

A novel complex algorithm using Viola – Jones method, algorithm CAMShift, wavelet transform and principal component analysis is proposed for hand gesture recognition in video sequence. It is shown that use of the proposed algorithm has processing speed about 30 frames per second and allows recognizing hand gesture video sequence in real time.

In this paper, a part of Cambridge Gesture database is used for testing the performance of hand gesture recognition algorithm on digital images. This database consists of 5 parts. The contrast condition of each part is not the same. All the hand gestures are divided into 12 classes using for recognition. For each part one hand gesture database, containing 200 images of each class (summary $12 \times 200 = 2400$ images), is created for testing this algorithm. The combining testing database of all 5 parts contains 1000 images of each class (summary $12 \times 1000 = 12\,000$ images). For each part one training database also is created. This training database contains 20 images of each class (summary $12 \times 20 = 240$ images). The combining testing database of all 5 parts contains 100 images of each class (summary $12 \times 100 = 1200$ images). Testing results of the proposed algorithm show that the number of rightly recognized hand gestures is 94.63 %.