

УДК 004.932

Д.А. Пиманкин, Б.А. Кисельман

ОЦЕНКА ДВИЖЕНИЯ ЧАСТОТНЫМИ МЕТОДАМИ В СИСТЕМАХ КОМПЬЮТЕРНОГО ЗРЕНИЯ

Нижегородский государственный технический университет им. Р.Е.Алексеева

Предлагается метод детектирования и анализа движения в последовательности пространственно совмещенных динамических изображений на основе комплексного ДПФ. Показана эффективность метода в ряде задач по сравнению с традиционными частотными методами.

Ключевые слова: комплексное ДПФ, движущиеся объекты, фоновое изображение, разностное изображение, бинарное изображение, сегментация изображений.

Введение

В прикладных задачах обработки видеопотоков движение возникает в двух случаях. *Во-первых*, при относительном перемещении сенсорной системы (камеры) и наблюдаемой сцены. *Во-вторых*, при изменениях в самой наблюдаемой сцене (движении некоторых объектов) [5, 7]. В данной работе рассматривается задача детектирования и анализа движения объектов при условии неподвижности сенсорной системы, т.е. предполагается, что все кадры последовательности динамических изображений пространственно совмещены.

Использование преобразования Фурье для анализа движения объектов

Частотные методы заключаются в использовании преобразования Фурье для нахождения оценок скоростей подвижных объектов.

Для начала рассмотрим последовательность $f(x, y, t)$, $t = 0, 1, \dots, K-1$, состоящую из K цифровых бинарных изображений размерами $M \times N$. При этом будем полагать, что камера неподвижна. Предположим, что на изображении присутствуют как стационарные компоненты, так и равномерно движущиеся объекты.

Для всех кадров этой последовательности умножим значение каждого пикселя на $\exp(j2\pi a_x x)$, а затем просуммируем взвешенные пиксели. Введем функцию от времени g_x :

$$g_x(t, a_x) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y, t) \exp(j2\pi a_x x), \quad t = 0, 1, \dots, K-1, \quad (1)$$

где a_x – некоторое положительное число.

Аналогично, зависимость суммы взвешенных значений пикселей от времени для оси y :

$$g_y(t, a_y) = \sum_{y=0}^{N-1} \sum_{x=0}^{M-1} f(x, y, t) \exp(j2\pi a_y y), \quad t = 0, 1, \dots, K-1. \quad (2)$$

Если в последовательности изображений не происходит движения, то $g_x = C_1 = \text{const}$ и $g_y = C_2 = \text{const}$.

Предположим, что одиночный единичный пиксель с начальной абсциссой x_0 перемещается вдоль оси x с постоянной скоростью v_x . В функции g_x ему будет соответствовать комплексная экспонента $\exp(j2\pi a_x(x_0 + v_x t))$ [1].

Таким образом, каждому пикселю, движущемуся с постоянной скоростью, в функции g_x соответствует комплексная экспонента. Приведенные рассуждения обобщаются на любую группу пикселей, т.е. если некоторый объект (под объектом подразумевается группа связан-

ных пикселей) поступательно движется вдоль оси x со скоростью v_x пикселей/кадр, то соответствующая ему комплексная экспонента будет иметь частоту $v_x a_x$.

Применяя дискретное преобразование Фурье к реальным частям функций (1) и (2), соответственно получаем:

$$G_x(u_x, a_x) = \sum_{t=0}^{K-1} \operatorname{Re}[g_x(t, a_x)] \exp\left(\frac{-j2\pi u_x t}{K}\right), \quad u_x = 0, 1, \dots, K-1, \quad (3)$$

$$G_y(u_y, a_y) = \sum_{t=0}^{K-1} \operatorname{Re}[g_y(t, a_y)] \exp\left(\frac{-j2\pi u_y t}{K}\right), \quad u_y = 0, 1, \dots, K-1. \quad (4)$$

Стационарным компонентам последовательности изображений соответствует некоторый постоянный член C в функции g . На графике зависимости модуля G от частоты он отобразится в виде пика в нуле. Равномерно движущемуся объекту будет соответствовать пик на частоте $|v_x|a_x$, что определяет отсчет $u_x' = K|v_x|a_x$ и, как следствие, симметрию ДПФ действительной функции, $K(1 - |v_x|a_x)$. Это может не приниматься во внимание [4].

Итак, в спектре Фурье имеется пик на частоте $|v_x|a_x$ ($u_x' = K|v_x|a_x$). При делении этой частоты на a_x получим $|v_x|$, т.е. модуль проекции скорости на ось x в пикселях/кадр. Аналогичными рассуждениями можно получить $|v_y|$ – модуль проекции скорости на ось y :

$$|\hat{v}_x| = \frac{u_x'}{Ka_x}. \quad (5)$$

$$|\hat{v}_y| = \frac{u_y'}{Ka_y}. \quad (6)$$

Поскольку t меняется с целым шагом от 0 до $K-1$, то целесообразно ограничить a_x числами $1/K, 2/K, \dots, 1/2$.

Недостаточная частота дискретизации может стать причиной ошибки, связанной с наложением спектров. Поскольку $u_x' = K|v_x|a_x$, одна из возможностей избежать ее состоит в выборе в качестве a_x ближайшего целочисленного значения к величине $a = u_{\max} / (K|v_{\max}|)$, где u_{\max} – максимальная частота (номер отсчета), определяемая величиной K , а v_{\max} – максимальная ожидаемая скорость объекта [1].

Для случая, когда в последовательности изображений присутствует лишь один объект, знак составляющей скорости в направлении оси x можно определить, рассчитав первую производную от действительной части и вторую производную от мнимой части функции g_x по времени:

$$J_{1x} = \left. \frac{d \operatorname{Re}[g_x(t, a_x)]}{dt} \right|_{t=n}, \quad (7)$$

$$J_{2x} = \left. \frac{d^2 \operatorname{Im}[g_x(t, a_x)]}{dt^2} \right|_{t=n}. \quad (8)$$

Одинаковые знаки J_{1x} и J_{2x} в некоторый момент времени $t = n$ означают, что составляющая скорости v_x положительна. Разные знаки J_{1x} и J_{2x} показывают, что эта компонента отрицательна. В случае нулевого значения одной из величин J_{1x} и J_{2x} рассматривается следующий ближайший момент времени $t = n \pm \Delta t$. Все приведенные рассуждения применимы и для оси y [1].

Анализ недостатков и модификация метода

При непосредственном применении формул (1) и (2) к полутоновому изображению функцию g_x (здесь и далее будем рассматривать только ось x) можно представить в следующем виде:

$$g_x(t, a_x) = \sum_{(x,y) \in P} [Q(x, y, t) - F(x, y)] \exp(j2\pi a_x x) + C, \quad t = 0, 1, \dots, K-1, \quad (9)$$

где $F(x, y)$ – изображение фоновой составляющей; $Q(x, y, t)$ – яркость точек объекта; P – множество точек, принадлежащих движущемуся объекту.

Здесь постоянная составляющая

$$C = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} F(x, y) \exp(j2\pi a_x x). \quad (10)$$

Для случая малых объектов (под малыми подразумеваются объекты, площадь которых достаточно мала для того, чтобы можно было пренебречь неоднородностью участка фонового изображения, заслоняемого объектом):

$$Q(x, y, t) - F(x, y) \approx Q(x_0, y_0, t) - F(x_0, y_0), \quad (x, y) \in P, \quad (11)$$

где x_0 и y_0 – координаты центра объекта P .

С учетом (11) приближенно запишем формулу (9) в виде

$$g_x(t, a_x) \approx [Q(x_0, y_0, t) - F(x_0, y_0)] \sum_{(x,y) \in P} \exp(j2\pi a_x x) + C, \quad t = 0, 1, \dots, K-1. \quad (12)$$

Видно, что множитель, стоящий перед суммой зависит от текущего положения объекта, что вносит дополнительную ошибку в конечный результат оценки скорости.

Теперь вернемся к анализу последовательности бинарных изображений. Распишем функцию g_x для группы пикселей (протяженностью L пикселей), движущихся вдоль оси x :

$$g_x(t, a_x) = \exp(j2\pi a_x (x_0 + v_x t)) \cdot [A_0 + A_1 \cdot \exp(j2\pi a_x) + \dots + A_{L-1} \cdot \exp(j2\pi a_x (L-1))] + C. \quad (13)$$

Введем величину A , зависящую только от геометрии объекта и значения a_x :

$$A = A_0 + A_1 \cdot \exp(j2\pi a_x) + A_2 \cdot \exp(j2\pi a_x \cdot 2) + \dots + A_{L-1} \cdot \exp(j2\pi a_x (L-1)), \quad (14)$$

тогда формула (13) запишется в виде

$$g_x(t, a_x) = |A| \cdot \exp(j \cdot \arg(A)) \cdot \exp(j2\pi a_x (x_0 + v_x t)) + C. \quad (15)$$

Как видим, недостатком классического метода является то, что из значения модуля функции G_x в точке u_x' (из амплитуды комплексной экспоненты) невозможно извлечь никакой информации о размерах объекта. В крайнем случае это может привести к полной нераспознаваемости объекта.

Предположим, что объект имеет размеры 8×8 пикселей и $a = 0,125$. В этом случае, независимо от скорости поступательного движения, он остается совершенно нераспознаваем. Комплексные экспоненты, соответствующие объекту, взаимно уничтожаются. Тот же самый эффект будет наблюдаться при размерах объекта кратных 8 (рис. 2, a).

Рассмотрим 4- либо 8-связанную область (примеры 4- и 8-связанных областей приведены на рис. 1) площадью S и протяженностью L вдоль оси x . Предлагается каждой такой об-

ласти поставить в соответствие ее центральный пиксель и величину площади S . Таким образом, суммирование L комплексных экспонент с различными фазовыми сдвигами заменяется умножением одной экспоненты на площадь области S .

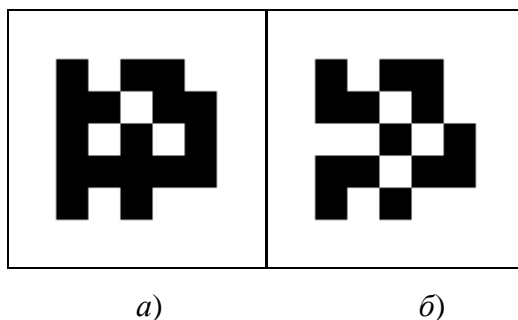


Рис. 1. Пример 4- (а) и 8-связанной (б) областей

Теперь имеется возможность по значениям функции G_x оценить площадь объекта:

$$\hat{S} = \frac{2 \cdot |G_x(u'_x, a_x)|}{K} \quad (16)$$

Графики зависимостей модулей функций G_y (для классического и модифицированного методов) от частоты приведены на рис. 2.

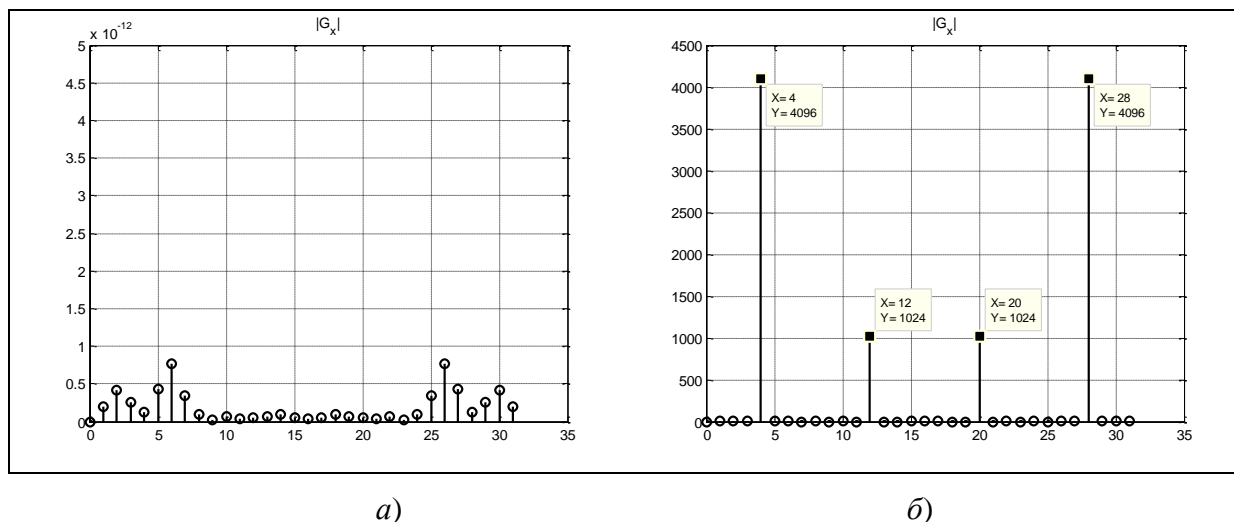


Рис. 2. Графики зависимостей $|G_y|$ от частоты: классический метод (а) и предлагаемый новый метод (б)

Применение формул (7) и (8) существенно затрудняется, если объектов два и больше.

Знак составляющей скорости вдоль осей будем находить при помощи ДПФ. Для этого преобразование Фурье будем искать не от действительной части функции g , а от самой комплексной функции.

Применяя преобразование Фурье к функции (1), получаем:

$$G_x(u_x, a_x) = \sum_{t=0}^{K-1} g_x(t, a_x) \exp\left(\frac{-j2\pi u_x t}{K}\right), \quad u_x = 0, 1, \dots, K-1. \quad (17)$$

При этом симметрия ДПФ пропадает. Пики в левой и правой частях амплитудно-частотного спектра показывают наличие движения объекта по направлению оси x и против нее соответственно.

Теперь формула для оценки значения скорости (со знаком) запишется в виде

$$\epsilon_x = \begin{cases} \frac{u'_x}{K \cdot a_x}, & \text{если } u'_x \in \left[1, \frac{K}{2} - 1\right], \\ \frac{u'_x - K}{K \cdot a_x}, & \text{если } u'_x \in \left[\frac{K}{2} + 1, K\right]. \end{cases} \quad (18)$$

Графики зависимостей модулей функций G_y (для действительного и комплексного ДПФ) от частоты приведены на рис. 3.

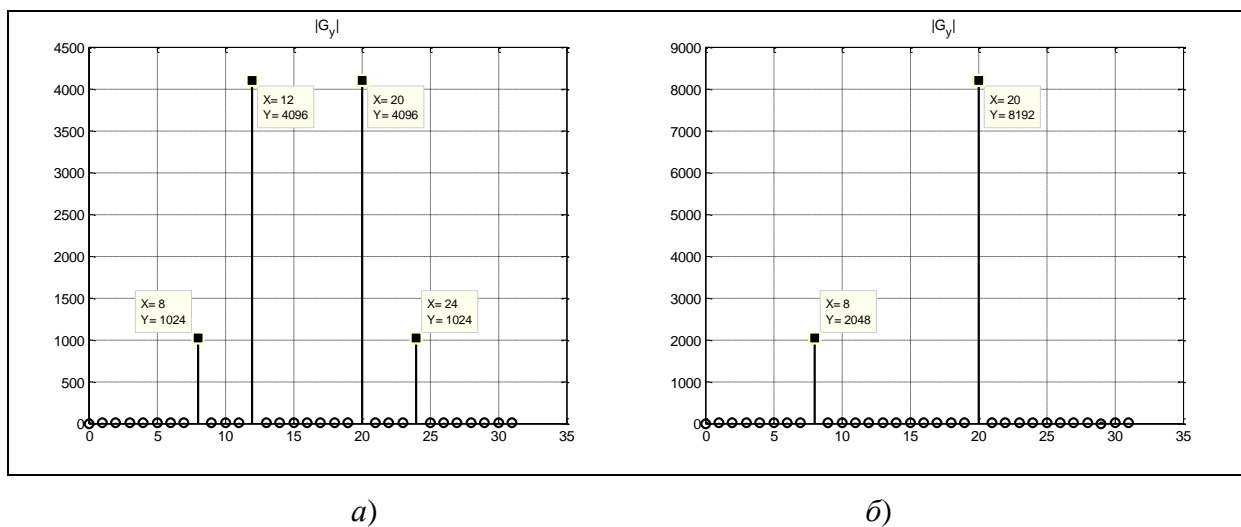


Рис. 3. Графики зависимостей $|G_y|$ от частоты: ДПФ от $\text{Re}(g_y)$ (а) и ДПФ от g_y (б)

Формула для оценки площади объекта (16) также изменится и запишется в виде

$$\hat{S} = \frac{|G_x(u'_x, a_x)|}{K}. \quad (19)$$

Анализ последовательностей полутоновых изображений

Перед применением предлагаемого метода исходные изображения необходимо предварительно сегментировать (бинаризовать с выделением движущихся объектов).

Воспользуемся следующей моделью последовательности динамических изображений:

$$f(x, y, t) = Q(x, y, t) \cdot B(x, y, t) + F(x, y) \cdot [1 - B(x, y, t)] + \xi(x, y, t), \quad (20)$$

где $F(x, y)$ – изображение фоновой составляющей; $Q(x, y, t)$ – яркость точек объекта; ξ – некоррелированный по пространству и между кадрами гауссов шум с нулевым средним и дисперсией σ_ξ^2 ;

$$B(x, y, t) = \begin{cases} 1, & \text{если точка } (x, y) \text{ принадлежит объекту,} \\ 0, & \text{если точка } (x, y) \text{ принадлежит фону.} \end{cases}$$

Задача бинаризации с выделением объектов заключается в нахождении некоторого приближения $B(x, y, t)$.

Эффективная сегментация является отдельной достаточно сложной задачей [3].

При тестировании предлагаемого метода бинаризация осуществлялась при помощи нахождения разностных изображений.

Будем считать, что СКО σ_{ξ} шума достаточно мало по сравнению с диапазоном яркостей наблюдаемого изображения, известна минимальная площадь объектов и имеется возможность хранить некоторое количество кадров, предшествовавших текущему.

Разностное изображение между двумя кадрами, взятыми в различные моменты времени t_i и t_j , определяется следующим образом:

$$d_{ij}(x, y) = \begin{cases} 1, & \text{если } |f(x, y, t_i) - f(x, y, t_j)| > D, \\ 0, & \text{в противном случае,} \end{cases} \quad (21)$$

где D – заранее заданный порог.

Единичные значения изображения $d_{ij}(x, y)$ соответствуют движению. Побочные единичные пиксели, вызванные шумом, устраняются поиском 4- либо 8-связанных компонент, и, если площадь меньше некоторого порога T , удалением последних [1].

Для сглаживания контуров объектов, а также заполнения узких разрывов, углублений и отверстий к изображению применяется операция замыкания по некоторому примитиву b [7].

Результаты эксперимента

При моделировании работы предлагаемого метода использовалась последовательность полутоновых изображений размером 256×256 и глубиной цвета 8 бит/пиксель. Каждый кадр является следствием наложения на фоновое изображение объектов и добавления гауссова шума с нулевым средним и СКО σ_{ξ} .

Каждый из объектов представляет из себя прямоугольную область с равномерным распределением яркости и постоянной скоростью поступательного движения ($v_x = \text{const}$, $v_y = \text{const}$).

Значения параметров, использованных при моделировании, приведены в табл. 1.

Таблица 1

Параметры моделирования

Параметр	Значение	Параметр	Значение
K – количество кадров	32	I_1 – яркость пикселей первого объекта	200
S_1 – площадь первого объекта, пикс	64	I_2 – яркость пикселей второго объекта	200
S_2 – площадь второго объекта, пикс	256	σ_{ξ} – СКО шума	0,1
v_{1x} – составляющая скорости первого объекта вдоль оси x	3	T – пороговое значение площади области	16
v_{1y} – составляющая скорости первого объекта вдоль оси y	4	D – пороговое значение разности изображений	16
v_{2x} – составляющая скорости второго объекта вдоль оси x	1	a_x – коэффициент	4/32
v_{2y} – составляющая скорости второго объекта вдоль оси y	– 6	a_y – коэффициент	2/32

Один из кадров последовательности динамических изображений и результат бинаризации этого кадра приведены на рис. 4.

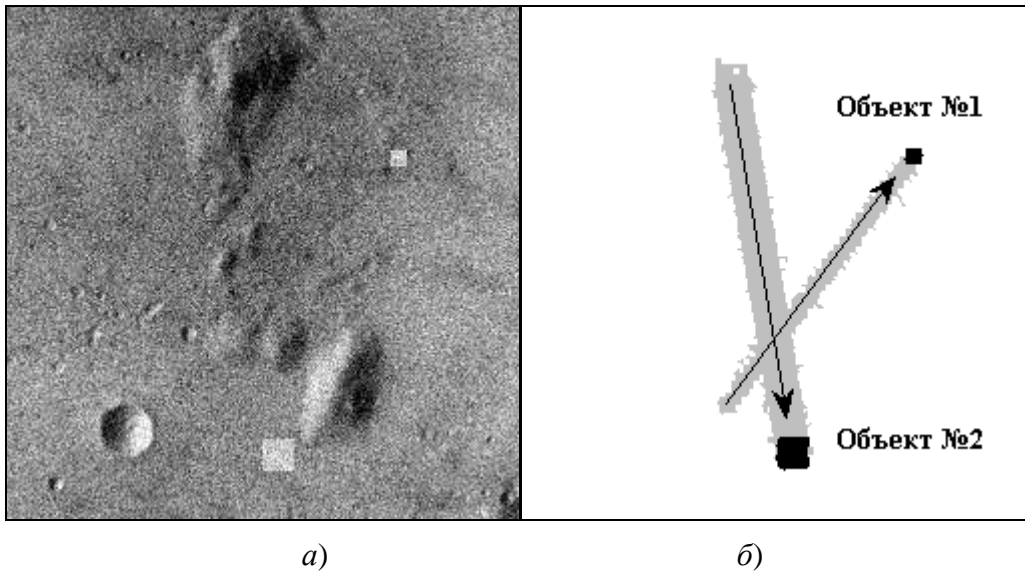


Рис. 4. Кадр из последовательности динамических изображений $f(x, y, t)$ размером 256×256 (а) и результат бинаризации с выделением подвижных объектов (б)

Графики зависимостей модулей функций G_x и G_y от частоты приведены на рис. 5.

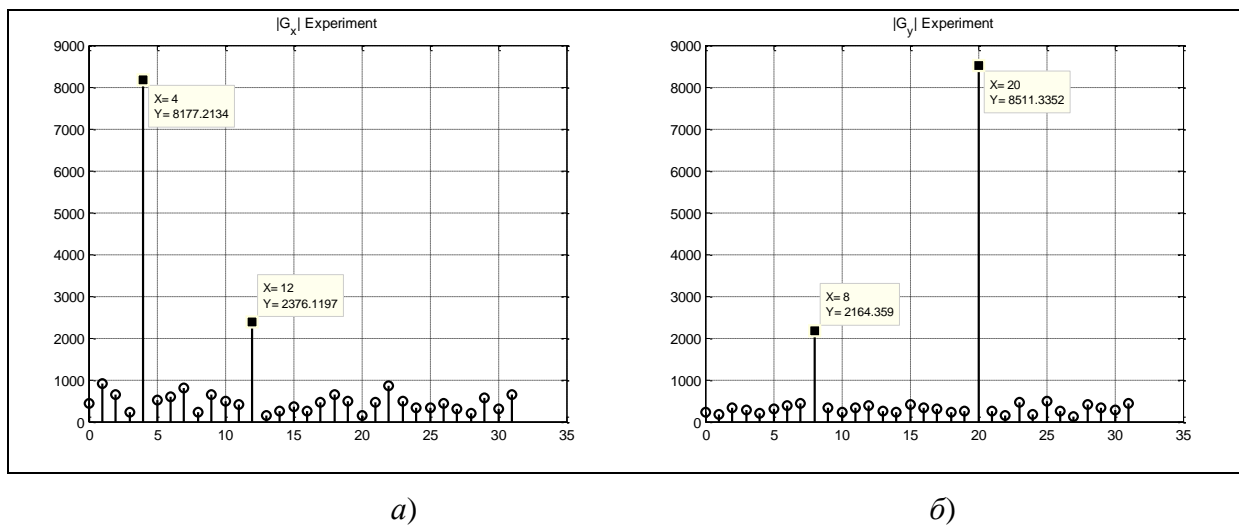


Рис. 5. Графики зависимостей $|G_x|$ от частоты (а) и $|G_y|$ от частоты (б)

Таблица 2

Результаты эксперимента

Номер объекта	1	2
Частота локального максимума u_x	12	4
Частота локального максимума u_y	8	20
Оценка скорости v_x , пикс/кадр	3	1
Оценка скорости v_y , пикс/кадр	4	-6
Оценка площади S_x , пикс	74	256
Оценка площади S_y , пикс	68	266

На практике для выполнения морфологических операций, а также расчета ДПФ, применяются быстрые алгоритмы [2, 6].

Из табл. 2 видно, что найденные оценки скоростей движения объектов соответствуют истинным значениям скоростей. Оценки площадей также найдены с достаточно высокой точностью. При этом следует отметить, что качество работы алгоритма, в частности, точность оценки размеров объектов, во многом зависит от качества выполнения операции бинаризации.

Итак, результаты экспериментальных исследований доказывают эффективность предложенного метода в ряде задач по сравнению с классическим подходом.

Заключение

В данной работе предложен и экспериментально исследован алгоритм нахождения оценок скоростей движения и площадей объектов в последовательности пространственно совмещенных динамических изображений частотными методами. Показана эффективность применения эквивалентной замены областей связанных пикселей и использования комплексного ДПФ в задачах детектирования и анализа движения.

Библиографический список

1. **Гонсалес, Р.** Цифровая обработка изображений / Р. Гонсалес, Р. Вудс. – М.: Техносфера, 2005.
2. **Кормен, Т.** Алгоритмы: построение и анализ / Т. Кормен, Ч. Лейзерсон, Р. Ривест, К. Штайн. – М.: Издательский дом «Вильямс», 2009.
3. **Прэтт, У.** Цифровая обработка изображений / У. Прэтт. – М.: Мир, 1982.
4. **Рабинер, Л.** Теория и применение цифровой обработки сигналов / Л. Рабинер, Б. Гоулд. – М.: Мир, 1978.
5. **Форсайт, Д.** Компьютерное зрение. Современный подход / Д. Форсайт, Ж. Понс. – М.: Издательский дом «Вильямс», 2004.
6. **Хуанг, Т.С.** Быстрые алгоритмы в цифровой обработке изображений: преобразования и медианные фильтры / Т.С. Хуанг. – М.: Радио и Связь, 1984.
7. Шапиро, Л. Компьютерное зрение / Л. Шапиро, Д. Стокман. – М.: Бином, 2006.

*Дата поступления
в редакцию 02.02.2010*

D.A. PIMANKIN, V.A. KISELMAN FREQUENCY-BASED MOTION ESTIMATION IN COMPUTER VISION SYSTEMS

We propose a method of motion detection and analysis in a sequence of spatially correspondent dynamic images, based on complex DFT. In a specific range of tasks the represented method is proved to be more effective in comparison to conventional frequency methods.

Key words: complex DFT, moving objects, background image, difference image, binary image, image segmentation.