

Дескрипторный алгоритм слежения с использованием камеры глубины

Д. Мирамонте-Харамилло*, В.И. Кобер**, В.Х. Диас-Рамирес***, В.Н. Карнаухов**

* Факультет вычислительной техники, Центр научных исследований и высшего образования,
Энсенада, 22860, Мексика

** Институт проблем передачи информации, Российская академия наук, Москва, 127051, Россия

*** Национальный политехнический институт - Центр исследований и развития цифровых
технологий, Тихуана, 22510, Мексика

Поступила в редколлегию 01.02.2016

Аннотация—Появление недорогих и качественных видеокамер и камер глубины привело к созданию большого количества алгоритмов слежения за объектами. В данной статье предлагается новый дескрипторный алгоритм слежения за объектами в реальном масштабе времени, использующий информацию от камеры глубины Microsoft Kinect. В качестве дескрипторного описания объекта слежения используются гистограммы направленных градиентов, вычисленные по круговым скользящим областям изображения сцены. Информация о глубине сцены используется при частичном перекрытии изображения объекта другими предметами сцены. Для ускорения процесса слежения используется модель предсказания движения объекта. Для обеспечения режима реального времени слежения алгоритм реализован на многоядерном графическом процессоре.

КЛЮЧЕВЫЕ СЛОВА: слежение за объектом, гистограммы направленных градиентов, алгоритм реального времени.

ВВЕДЕНИЕ

Слежение за объектами имеет много практических применений в таких областях, как медицина, системы безопасности и наблюдения за порядком в общественных местах и так далее. Для создания надежных алгоритмов слежения нужно учитывать такие факторы, как наличие аддитивного шума датчиков, неравномерность освещения, частичное перекрытие одних объектов другими, геометрические искажения изображений объектов, такие как повороты, масштабирование и т.д. Алгоритмы слежения можно классифицировать по типу дескриптора, используемого для идентификации объекта [1]: алгоритмы на основе эталонов [2]–[4] используют различные статистические характеристики для описания объектов; алгоритмы на основе силуэтов используют формы и границы объектов [5], [6]; алгоритмы на основе признаков [7], [8] выделяют специфические признаки (точки интереса) объектов слежения.

В последнее время алгоритмы слежения вновь обрели популярность в значительной степени благодаря выходу на рынок гибридных камер RGBD, таких как Microsoft Kinect и ASUS Xtion PRO. Это связано с тем, что до появления таких камер получение информации о глубине сцены было вычислительно сложной проблемой. В настоящее время видеосигнал и карта глубины могут быть получены в режиме реального времени и эффективно использованы в системах слежения за объектами. Современные алгоритмы слежения можно разделить на две большие группы: алгоритмы, которые используют адаптивные модели обновления дескрипторов, но не используют информацию о глубине сцены [9], [10]; алгоритмы, которые дополнительно используют информацию о глубине сцены [11], [12]. Алгоритмы слежения обычно состоят из

трех последовательных шагов: сопоставление дескрипторов, классификация и предсказание. На этапе сопоставления осуществляется поиск фрагмента кадра, чей дескриптор является наиболее похожим на дескриптор объекта слежения. Когда соответствующий фрагмент найден, то производится классификация найденного объекта. Для этого часто используется бинарный классификатор. И, наконец, по текущему положению объекта слежения предсказывается наиболее вероятное положение объекта в следующем кадре.

В этой статье предложен алгоритм слежения, использующий в качестве дескрипторного описания объекта гистограммы направленных градиентов [13], которые вычисляются по круговым скользящим окнам изображения сцены [14], [15]. Статистические параметры шума изменяются в начале обработки по изображению сцены, и эта информация используется для вычисления минимального количества дискретных уровней гистограмм направленных градиентов. Кроме того, неравномерное освещение адаптивно корректируется на изображении сцены. Для предсказания положения объекта в следующем кадре по текущему положению объекта используется модель предсказания движения на основе временных рядов [16] и информация о глубине сцены. Информацию о глубине обрабатываемого фрагмента также помогает решить проблему перекрытия объекта другими предметами сцены. Для обеспечения режима реального времени слежения (30 кадров в секунду) алгоритм реализован на многоядерном графическом процессоре с использованием SIMD технологий для обработки нескольких гистограмм одновременно.

Статья организована следующим образом: в разделе 1 описывается алгоритм удаления шума на изображении сцены, а также коррекция неравномерного освещения сцены; в разделе 2 описывается предлагаемый алгоритм слежения; в разделе 3 с помощью компьютерного моделирования на известных базах данных сравнивается качество работы предложенного алгоритма и наиболее успешных современных алгоритмов слежения.

1. УДАЛЕНИЕ ШУМА И КОРРЕКЦИЯ НЕРАВНОМЕРНОГО ОСВЕЩЕНИЯ СЦЕНЫ

Первым шагом большинства алгоритмов обработки изображений является коррекция искажений, вносимых изображающими системами и условиями окружающей среды. Широкополосный аддитивный шум датчиков может быть эффективно удален с помощью оптимальной линейной фильтрации [17]. Для этого необходимо измерить статистические параметры шума, такие как дисперсия шума. Автоковариационная функция белого шума является, по существу, функцией Кронекера. Поэтому ковариационная функция неискаженного сигнала может быть получена из ковариационной функции искаженного сигнала путем линейной экстраполяции значений функции в окрестности начала координат [18]. Как показано на Рис. 1, оценка дисперсии белого шума — разница между этими функциями. Оценка среднеквадратичного отклонения шума (σ_n) используется для квантования гистограммы направленных градиентов с учетом возможных погрешностей, вносимых шумом в вычисление углов. Количество уровней квантования гистограммы Q может быть получено следующим образом:

$$Q = \left\lceil \frac{360}{\sigma_n} \right\rceil, \quad (1)$$

где $\lceil \cdot \rceil$ — оператор целой части вещественного числа. Для коррекции неравномерного освещения в данной работе используется гамма-коррекция изображения сцены [19]. Если динамический диапазон входного сигнала большой, то обычно производится линейное сжатие диапазона для последующего квантования сигнала. Глаз человека производит нелинейное сжатие диапазона автоматически для лучшего восприятия визуальной информации. Гамма-коррекция изображения позволяет улучшить визуальное восприятие при неравномерном освещении, усилить локальный контраст и, как следствие, улучшает качество работы последующего градиентного оператора. Наилучшие результаты получаются при значениях параметра гамма-коррекции

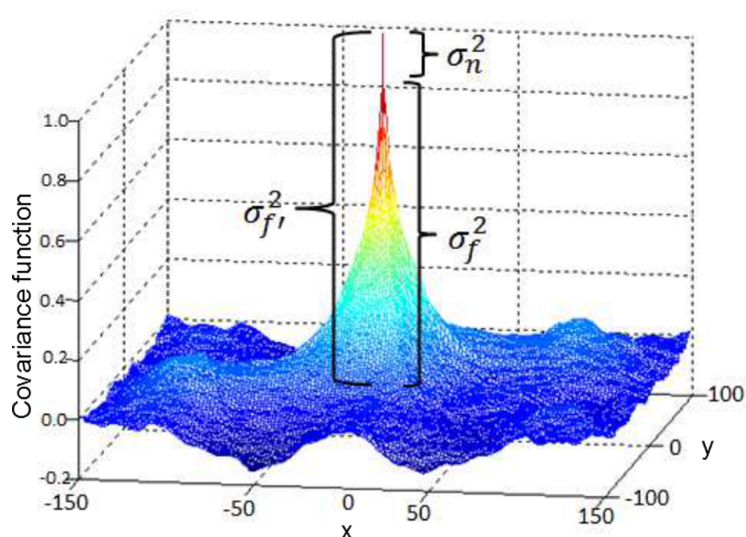


Рис. 1. Вычисление дисперсии белого шума: $\sigma_{f'}^2$ — дисперсия наблюдаемого сигнала, σ_f^2 — дисперсия неискаженного сигнала, σ_n^2 — дисперсия белого шума.

в диапазоне $[1/2.2, 1/2.6]$. Пример гамма-коррекции показан на Рис. 2, и он рассчитывается следующим образом:

$$V_O = V_I^p, \quad (2)$$

где V_O и V_I — выходное и входное изображения, соответственно, p — параметр гамма-коррекции. Для простоты в данной работе, использовалась процедура поэлементного извлечения квадрат-

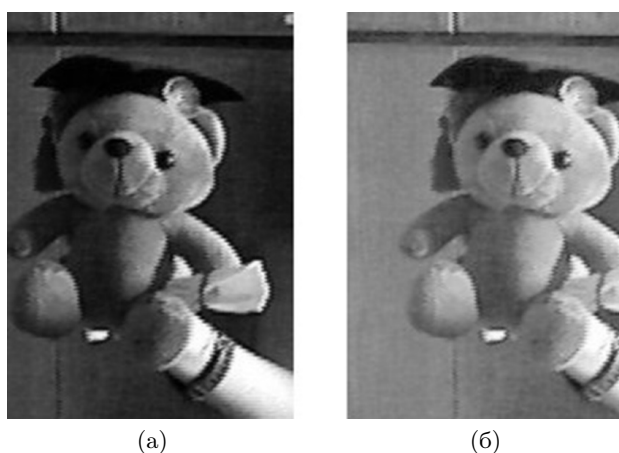


Рис. 2. а) исходное изображение, б) изображение после гамма-коррекции.

ного корня из входного сигнала. Этот метод корректирует неравномерное освещение сцены и уменьшает эффект образования локальных теней. В зонах небольшого размера, таких как фрагмент кадра или зона скользящего окна, неравномерное освещение становится локально-однородным. Таким образом, использование небольших фрагментов, а не всего изображения при решении задачи сопоставления, позволяет существенно улучшить качество сопоставления вычислительно простым способом. Для улучшения качества алгоритма слежения при неравномерном освещении дополнительно используется информация о глубине сцены, полученная от

Kinect камеры. Эта информация инвариантна к освещению. При перекрытии объекта другими предметами сцены эта информация помогает уменьшить число фрагментов, поступающих для сопоставления при поиске объекта слежения.

2. АЛГОРИТМ СЛЕЖЕНИЯ ЗА ОБЪЕКТОМ

В этом разделе описаны основные компоненты предлагаемого алгоритма слежения: геометрическая структура, состоящая из круглых областей; дескрипторы — гистограммы направленных градиентов, которые вычисляются по скользящим круглым областям [14], [15]; сопоставление дескрипторов на основе итерационного алгоритма, уменьшающего пространство признаков; и, наконец, оценка положения объекта в следующем кадре с использованием модели предсказания движения и карты глубины сцены.

2.1. Геометрическая структура

Определим набор круглых окон $\{W_i, i = 1, \dots, M\}$ на эталонном изображении как набор следующих закрытых дисков:

$$W_i = \{(x, y) \in \mathbb{R} : (x - x_i)^2 + (y - y_i)^2 \leq r^2, i = 1, \dots, M\}, \quad (3)$$

$$D_{ij} = \left\{ \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}, i = 1, \dots, M; j = 1, \dots, M \right\}, \quad (4)$$

$$\gamma_i = \left\{ \cos^{-1} \left[\frac{D_{i,j+1}^2 + D_{i,j+2}^2 - D_{i,j+3}^2}{2D_{i,j+1}D_{i,j+2}} \right], i = 1, \dots, M; j = 1, \dots, M \right\}, \quad (5)$$

где (x_i, y_i) — координаты центра окна, r — радиус диска, D_{ij} — расстояние между i -ым и j -ым дисками, γ_i — угол между центром i -го диска и центрами двух соседних дисков. Пример геометрической структуры проиллюстрирован на Рис. 3. Диски формируют некоторую

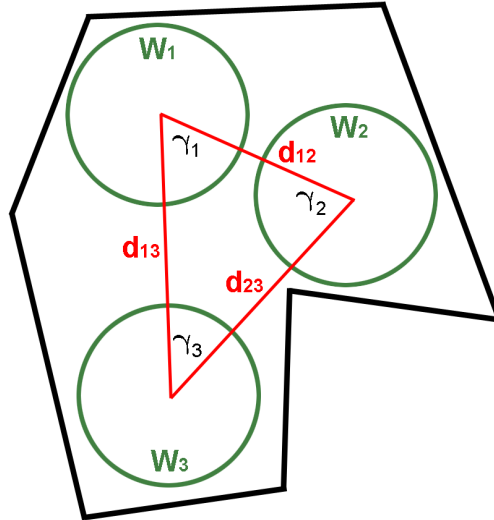


Рис. 3. Пример геометрической структуры, вписанной в объект слежения.

геометрическую структуру с относительными расстояниями и углами между центрами окон.

Гистограммы направленных градиентов вычисляются в круглых областях и используются для сопоставления. Следует отметить, что в любой позиции структуры каждый диск содержит область изображения, которая не изменяется при повороте. Следовательно, гистограммы направленных градиентов, вычисляемые в круглых областях, также инвариантны повороту. Использование только одного окна для сопоставления дает низкие результаты сопоставления. Поэтому рекомендуется выбирать такое минимальное количество дисков, чтобы заполнить как можно большую внутреннюю область эталонного объекта. Гистограммы направленных градиентов являются хорошими признаками для сопоставления [13] потому, что они обладают хорошей дискриминантной способностью и являются устойчивыми к небольшим деформациям, таким как поворот и масштабирование. Кроме того, геометрическая структура обладает еще одной особенностью: каждый диск содержит информацию об объекте, поэтому при частичном перекрытии объекта другими предметами информация какого-то диска может быть недоступна, но структура позволяет определять положение остальных дисков для установления области поиска объекта в следующем кадре.

2.2. Deskriptorное описание

Предлагаемый дескриптор основан на гистограммах направленных градиентов, вычисляемых в закрытых дисках того же размера, что и диски геометрической структуры. Для вычисления градиентов на сглаженном (после удаления шума) изображении сцены обычно используются операторы градиента, такие как Собеля, Робинсона и Кирша [20]. Эти операторы дополнительно сглаживают изображения для уменьшения сильного воздействия градиентного оператора на границах деталей изображения. Как показывают многочисленные эксперименты на реальных изображениях достаточно использовать простые операторы градиента без эффекта сглаживания важных деталей, такие как,

$$gx = [-1 \quad 0 \quad 1], \quad gy = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}. \quad (6)$$

Модуль и угол градиента в каждой точке изображения (x, y) вычисляется как

$$mag(x, y) = \sqrt{gx^2 + gy^2}, \quad (7)$$

$$ori(x, y) = \frac{180}{\pi} \tan^{-1}(gy/gx). \quad (8)$$

Угол градиента находится в диапазоне $[0^\circ, 360^\circ]$. Из-за различных условий видеосъемки два соседних кадра могут иметь небольшие изменения значений отсчетов, которые могут сильно изменить гистограмму при значении угла 360° . Для того, чтобы компенсировать это изменение выбирается количество уровней квантования Q в зависимости от среднеквадратичного отклонения шума (1). Угол градиента для каждого отсчета квантуется как [21]

$$\varphi = \frac{Q}{360} ori(x, y) + \frac{1}{2}, \quad (9)$$

где коэффициент $1/2$ учитывает поворот начала координат против часовой стрелки, поэтому значения угла в начале и в конце гистограммы будут принадлежать первому уровню квантования. Таким образом, циклическое условие для угла градиента выполнено. Обработка с помощью графического процессора делает задачу тривиальной. Гистограммы направленных градиентов вычисляются как

$$HOG(a) = \begin{cases} \sum_{(x,y) \in R} \delta(\alpha - \varphi(x, y)), & \text{если } mag(x, y) \geq Med \\ 0, & \text{в противном случае} \end{cases}, \quad (10)$$

где $\alpha = (0, \dots, Q - 1)$ — номер дискретного уровня гистограммы, Med — медиана модулей градиентов внутри круглого окна и $\delta(z) = \begin{cases} 1, & z = 0, \\ 0, & \text{в противном случае} \end{cases}$ — дельта-функция Кронекера. Для обеспечения инвариантности к повороту используется нормализованный корреляционный оператор для сравнения гистограмм эталонного изображения и фрагмента сцены. Вычислим центрированную и нормализованную гистограмму направленных градиентов в следующем виде:

$$\overline{HOG}(\alpha) = \frac{HOG(\alpha) - Mean}{\sqrt{Var}}, \quad (11)$$

где $Mean$ и Var — выборочное среднее значение и дисперсия гистограммы, соответственно. Отметим, что вычисления в уравнении (10) требуют приблизительно $[\pi r^2]$ операций сложения. Для снижения вычислительной сложности, вычисление гистограммы в любой позиции скользящего окна может быть выполнено рекурсивным способом. Такой подход приводит к вычислительной сложности порядка $[2\pi r]$ операций сложения [15] (см. Рис. 4).

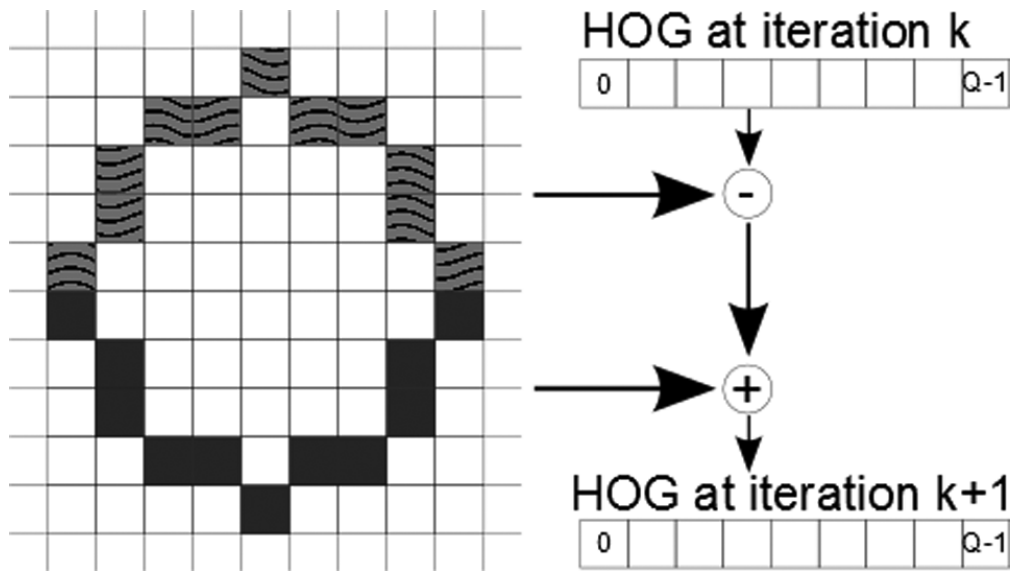


Рис. 4. Рекурсивное вычисление гистограммы вдоль столбцов.

2.3. Сопоставление фрагментов изображений

На k -ом шаге итерационного процесса вычисляется гистограмма направленных градиентов фрагмента сцены. Сопоставление выполняется с помощью корреляции в области Фурье между гистограммами i -го круглого окна в k -ой позиции фрагмента сцены и i -го круглого окна объекта слежения [20],

$$C_i^k(\alpha) = IFT[HS_{ik}(\omega)HO_i^*(\omega)], \quad (12)$$

where $HS_{ik}(\omega)$ — преобразование Фурье гистограммы направленных градиентов внутри i -го окна входной сцены, $HO_i(\omega)$ — преобразование Фурье гистограммы направленных градиентов внутри i -го круглого окна объекта слежения, а звездочка $*$ обозначает комплексное сопряжение, $IFT[\cdot]$ — обратное преобразование Фурье. Корреляционный пик является мерой схожести

двух гистограмм и может быть получен следующим образом:

$$P_i^k = \max_{\alpha} \{C_i^k(\alpha)\}. \quad (13)$$

Значения этих корреляционных пиков находятся в диапазоне $[-1, 1]$. Нормализованные корреляционные пики (13) обладают двумя важными свойствами: инвариантность к повороту, поскольку циклический сдвиг значений гистограммы соответствует циклическому сдвигу значений пиков; нормализация гистограмм (11) позволяет учесть небольшую разницу между масштабами эталонного изображения и сцены. Вычисление центрированных и нормализованных гистограмм для всех круглых окон по эталонному изображению и вычисление преобразований Фурье могут быть выполнены как предварительная обработка эталонных данных. Предлагаемый итерационный алгоритм, уменьшающий пространство признаков, кратко может быть описан в следующем виде.

1. Выполнить сопоставление первого круглого окна в структуре, получив корреляционную функцию C_1 и корреляционный пик P_1 ; цель этапа состоит в том, чтобы исключить из дальнейшего рассмотрения как можно больше точек на корреляционной плоскости, применяя некоторый заданный порог Th [14].
2. Оставленные на первом этапе точки используются как центры для сопоставления со вторым круглым окном структуры; для каждой такой точки строится кольцевая область с внешним и внутренним радиусами, равными $D_{1,2} \pm \varepsilon$, соответственно, для поиска пика внутри второго окна P_2 на плоскости C_2 . Здесь ε — малая величина изменения расстояния вследствие геометрических деформаций формы объекта слежения. Поиск последовательных корреляционных пиков может быть выполнен быстро при учете углов между центрами окон. Например, для третьего окна вычисляем пересечение между кольцевой областью с внешним и внутренним радиусами $D_{2,3} \pm \varepsilon$ и открытым круговым сектором с началом в позиции пика второго окна, и ограниченным внутренним и внешним радиусами с наклонами $\gamma_1 \pm \varepsilon_\gamma$. Такая процедура с каждым шагом уменьшает область поиска потенциальных точек для рассмотрения. Каждый шаг выполняется с помощью графических процессоров, то есть набор процессоров одновременно анализирует множество корреляционных пиков и окрестностей выбранных точек после каждой итерации, пока не будут проверены все гистограммы геометрической структуры (см. Рис. 5). Окончательное решение о наличии объекта принимается при учете совместного распределения корреляционных пиков для всех рассматриваемых окон.

Таким образом, снижение вероятности ошибок пропуска и вероятности ошибок ложной тревоги достигается на основе совместного анализа признаков во всех рассматриваемых круглых областях эталона.

2.4. Слежение за объектом

Для того чтобы увеличить скорость обработки, из изображения текущего кадра вырезается фрагмент, содержащий объект. Предполагается, что положение объекта и сам объект слежения в первом кадре определены. Геометрическая структура объекта и гистограммы направленных градиентов вычисляются из первого кадра. Геометрическая структура всегда строится внутри формы объекта, а размер фрагмента выбирается в 1.5 раз больше размера объекта, чтобы учесть возможную деформацию формы объекта. Предполагая равномерность движения объекта на нескольких соседних кадрах, предсказание положения объекта можно производить с помощью степенной модели движения объекта [16]. Как показывают многочисленные эксперименты, достаточно использовать квадратичную экстраполяцию:

$$x = a + bk + ck^2. \quad (14)$$

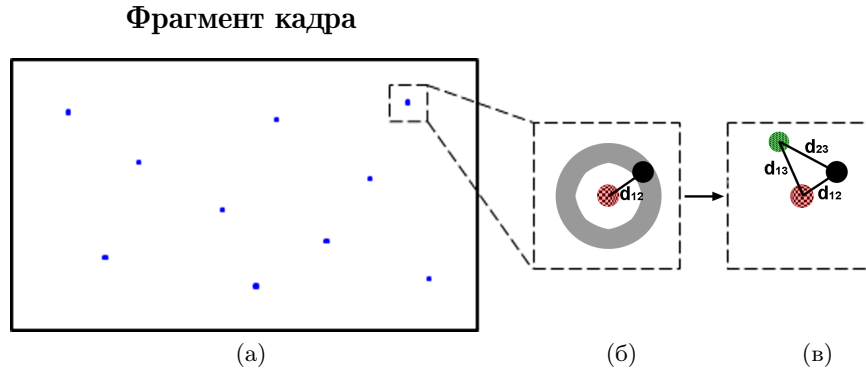


Рис. 5. Уменьшение пространства поиска для трех круглых окон геометрической структуры. а) Результат поиска кандидатов на первом этапе для первого окна структуры. б) Для каждой выбранной точки на первом этапе производится поиск кандидатов на втором этапе в кольцевой области. в) Используя углы и расстояния между окнами геометрической структуры эталона, производится дальнейшее уменьшение пространства поиска на третьем шаге.

где x, y — координаты объекта; k — номер кадра, где вычисляются координаты объекта; a, b, c — параметры движения объекта, которые являются и параметрами параболы. Используя метод наименьших квадратов, можно получить следующую систему уравнений:

$$\sum x = ma + b \sum k + c \sum k^2, \quad (15)$$

$$\sum xk = a \sum k + b \sum k^2 + c \sum k^3, \quad (16)$$

$$\sum xk^2 = a \sum k^2 + b \sum k^3 + c \sum k^4. \quad (17)$$

Если $m = 5$, то предсказание сильно упрощается: $k = [-2, 2]$, суммы $\sum k$ и $\sum k^3$ равны 0, и система уравнений легко решается:

$$a = \frac{\sum x + c \sum k^2}{m}, \quad (18)$$

$$b = \frac{\sum xk}{\sum k^2}, \quad (19)$$

$$c = \frac{m \sum xk^2 - \sum k^2}{m \sum k^4 - (\sum k^4)^2}. \quad (20)$$

Предсказание делается для всех круговых окон структуры объекта в каждом последующем кадре. Дополнительно используется информация о глубине сцены для выделения фрагмента и решения проблемы перекрытия объекта другими предметами сцены.

Отметим, что если происходит перекрытие объекта посторонним предметом, то гистограмма глубины сцены будет содержать две моды: первый локальный максимум будет соответствовать постороннему предмету, так как предмет находится ближе к камере глубины [11], а второй максимум соответствует объекту распознавания (см. Рис. 6). Используя информацию о глубине кадра, в случае перекрытия объекта посторонними предметами можно точнее выделить фрагмент кадра, полученный на стадии предсказания без учета информации о глубине сцены. Если объект не обнаруживается в предсказанном фрагменте (выпадает из поля зрения камеры на некоторое время, произошли полное перекрытие объекта другими предметами или ошибка предсказания из-за резкого изменения параметров движения объекта), то поиск объекта осуществляется по большим фрагментам, пока объект не будет обнаружен.

Мы различаем два типа перекрытия объекта: частичное и полное перекрытие объекта посторонними предметами. В первом случае алгоритм проверяет на соответствие как можно

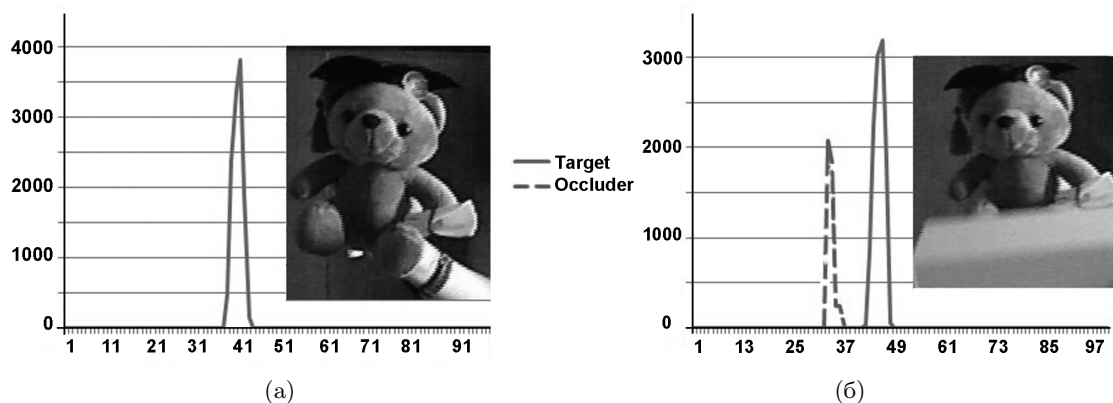


Рис. 6. а) Гистограмма глубины сцены, содержащей объект, б) гистограмма глубины сцены, содержащей объект и перекрывающий его предмет.

больше дисков геометрической структуры, и решение о наличии объекта принимается с учетом совместного распределения корреляционных пиков для всех рассматриваемых дисков. При сильном перекрытии объекта обнаружение возможно, если, по крайней мере, существует наилучшее соответствие для одного диска. Если происходит полное перекрытие объекта, алгоритм производит слежение за перекрывающим предметом, анализируя карту глубины. Как только появляется информация на карте глубины об объекте слежения, то рассматривается вариант частичного перекрытия объекта. Если объект не найден в текущем фрагменте кадра, модель предсказания и анализ карты глубины сцены позволяют оценить положение объекта в следующем кадре. Область поиска увеличивается на $0.1r$ каждый раз, когда объект не обнаружен, до максимального размера поиска $2r$. Если же объект не будет обнаружен, то он ищется по полным кадрам.

3. РЕЗУЛЬТАТЫ КОМПЬЮТЕРНОГО МОДЕЛИРОВАНИЯ

В этом разделе представлены результаты экспериментов, выполненных с использованием базы данных Princeton Tracking Benchmark [11]. Видеосъемки состоят из пяти видеофрагментов, снятых с помощью Microsoft Kinect. Количество кадров в видеофрагментах варьируется от 44 до 370. Каждая видеосъемка содержит RGBD изображения размером 640×480 пикселей и карту глубины, а также информацию для независимой проверки.

Результаты работы предложенного алгоритма сравнивались с результатами популярных алгоритмов слежения. Рис. 7 иллюстрирует некоторые использованные видеофрагменты и результаты, полученные с помощью предложенного алгоритма по слежению за различными объектами. Для сравнения были выбраны алгоритмы разных категорий с наилучшими результатами по слежению за объектом: алгоритмы, использующие информацию о глубине сцены, такие как RGBD Occlusion+Optical Flow (RGBDOcc+OF) [11] и Occlusion Aware Particle Filter (OAPF) [12]; алгоритмы, не использующие информации о глубине сцены, такие как RGBD+Optical Flow (RGBD+OF) [11], Tracking-Learning-Detection (TLD) [9] и Multiple Instance Learning (MIL) [10].

Параметры предложенного алгоритма: $M = 2$, $Th = 0.7$, $r = 32$ пикселя, и

$$Q = \begin{cases} 360/\sigma_n, & \text{если } 1.5 < \sigma_n < 40, \\ 64, & \text{в противном случае.} \end{cases}$$

Алгоритм был реализован на стандартном персональном компьютере с процессором Intel Core i7 с 3.2 ГГц and 8 Гб RAM и графической картой ATI RADEON HD 6450 с использованием OpenCV и OpenCL. Для оценки качества слежения

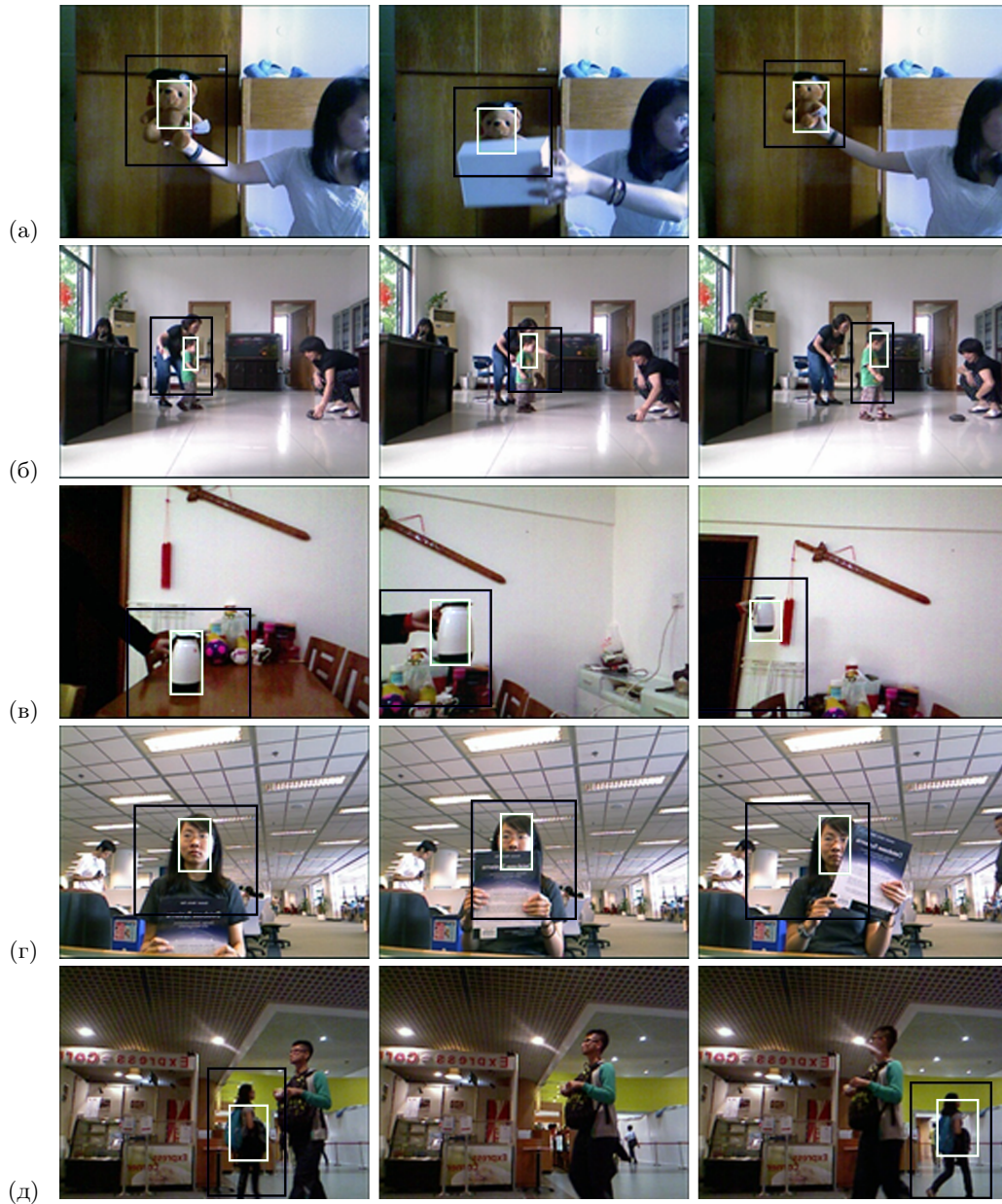


Рис. 7. Используемые видеофрагменты из базы данных Princeton Tracking Benchmark, содержащие кадры с изменением масштаба, освещения, поворотами и перекрытием объекта. Результаты слежения предложенным алгоритмом: белый прямоугольник соответствует обнаруженному объекту, а черный прямоугольник — предсказанная область поиска.

объектов тестируемым алгоритмом, было предложено измерять степень перекрытия между прямоугольной областью нахождения объекта, полученной с помощью алгоритма, и заранее точно известной прямоугольной областью нахождения объекта [11]:

$$O_k = \begin{cases} \frac{GT_k \cap T_k}{GT_k \cup T_k}, & \text{если обе области } GT_k \text{ и } T_k \text{ определены,} \\ 0, & \text{если обе области } GT_k \text{ и } T_k \text{ не определены,} \\ -1, & \text{если только одна из областей } GT_k \text{ или } T_k \text{ не определена,} \end{cases} \quad (21)$$

где GT_k — точно известная прямоугольная область нахождения объекта, T_k — прямоугольная область нахождения объекта, полученная с помощью алгоритма, k — номер кадра. В каче-

стве коэффициента успешного обнаружения (success rate) будем использовать нормированное количество пересечений с заданным порогом перекрытия прямоугольной области нахождения объекта, полученной с помощью алгоритма, и точно известной областью нахождения объекта. На Рис. 8 представлен этот показатель для тестируемых алгоритмов как функция порога перекрытия. Отметим, что алгоритмы, использующие информацию о глубине сцены, как правило,

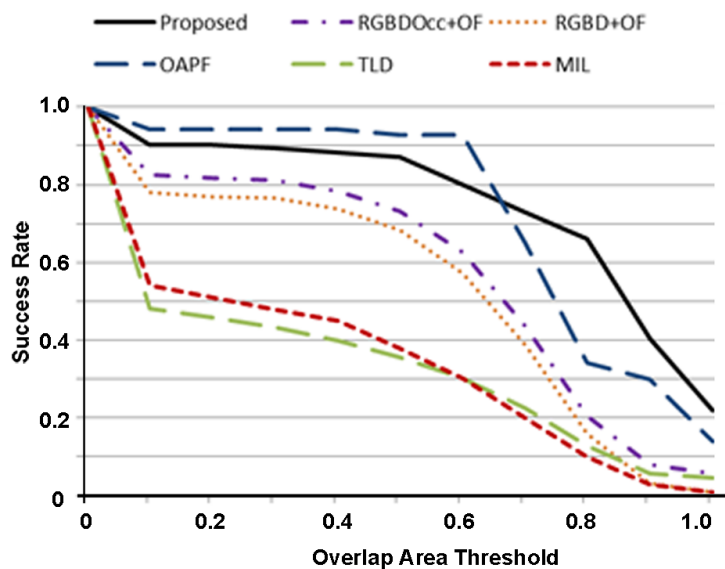


Рис. 8. Коэффициент успешного обнаружения объекта с помощью тестируемых алгоритмов в зависимости от порога площади перекрытия.

имеют лучшие результаты, чем алгоритмы, которые не используют информацию о глубине сцены. Предложенный алгоритм превосходит все алгоритмы при увеличении порога перекрытия, что означает, что он является наилучшим среди тестируемых алгоритмов для точного обнаружения и слежения за объектами. Отметим, что качество работы OAPF [12] резко ухудшается при пороге перекрытия 70%, в то время как предложенный алгоритм хорошо обнаруживает объект при этом пороге перекрытия (см. Рис. 9).

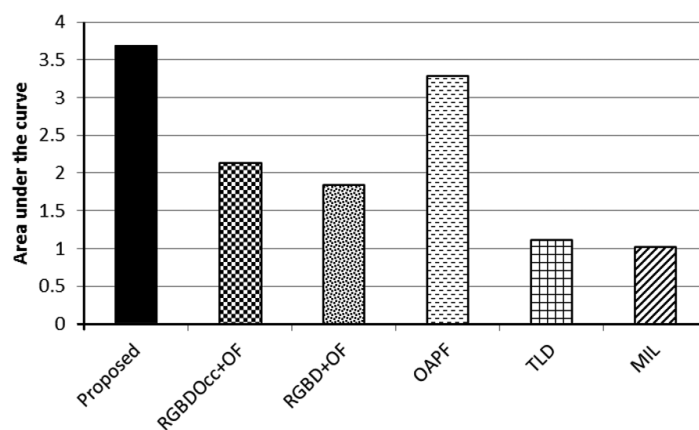


Рис. 9. Средняя степень перекрытия обнаружения объекта при пороге площади перекрытия 70%.

Для оценки качества работы тестируемых алгоритмов в данной работе используются три типа ошибок [11]: ошибка I типа возникает, когда объект является видимым на сцене, но нет пересечения между точно известной прямоугольной рамкой, ограничивающей объект на сцене, и прямоугольной рамкой, полученной тестируемым алгоритмом; ошибка II типа возникает, когда объект полностью закрыт посторонним предметом, но алгоритм обнаруживает объект в произвольном месте на сцене; ошибка III типа возникает, когда объект является видимым на сцене, но алгоритм не обнаруживает объекта. Все три ошибки проиллюстрированы на Рис. 10.

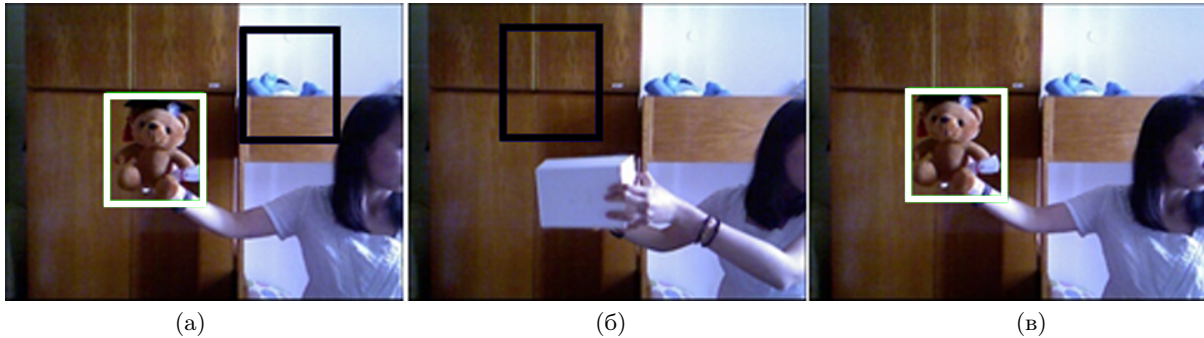


Рис. 10. Иллюстрация ошибок трех типов: прямоугольная белая рамка — точно известное положение объекта, прямоугольная черная рамка — обнаружение тестируемым алгоритмом. а) Ошибка I типа, б) ошибка II типа, в) ошибка III типа.

Как видно из Рис. 11, предлагаемый алгоритм выполняет слежения за объектами с меньшими ошибками по сравнению с другими тестируемыми алгоритмами. Интересно отметить, что

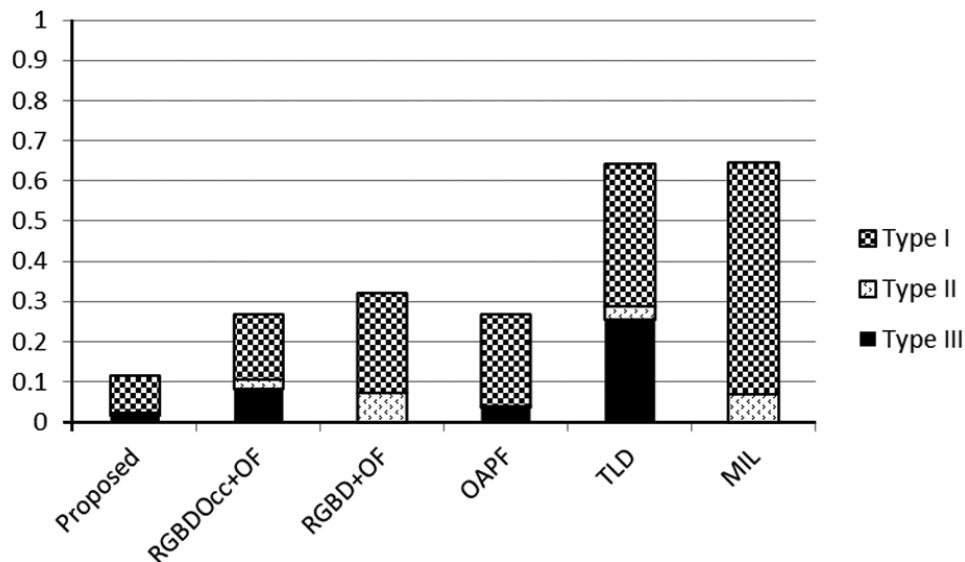


Рис. 11. Нормированные ошибки слежения трех типов для тестируемых алгоритмов.

результаты слежения алгоритмами, которые не учитывают информацию о глубине сцены, значительно хуже, чем результаты алгоритмов, учитывающих глубину сцены. Информация о глу-

бине сцены позволяет лучше выделять область поиска, а также решать успешно некорректную задачу обнаружения объекта при частичном перекрытии объекта посторонними предметами. Предлагаемый алгоритм анализирует результат совместного сопоставления круглых областей геометрической структуры, что позволяет обнаруживать объект при частичном перекрытии и легко восстановить всю геометрическую структуру после прохождения объектом препятствия (перекрытия).

ЗАКЛЮЧЕНИЕ

В работе представлен новый дескрипторный алгоритм слежения за объектом в реальном масштабе времени, инвариантный к неравномерному освещению и устойчивый к частичному перекрытию объекта посторонними предметами. Алгоритм использует информацию о глубине сцены от Kinect камеры. В качестве дескрипторного описания объекта используются гистограммы направленных градиентов, вычисленные по круговым скользящим областям изображения сцены. Для ускорения процесса слежения используется степенная модель предсказания движения объекта. По результатам компьютерного моделирования можно сделать вывод, что предложенный алгоритм является конкурентоспособным с лучшими современными алгоритмами слежения. Более того, точность слежения предложенного алгоритма на тестовых базах данных выше, чем точность тестируемых алгоритмов с точки зрения ошибок слежения трех типов. Для обеспечения режима реального времени предложенный алгоритм реализован на многоядерном графическом процессоре с интегрированной параллельной архитектурой, что позволяет проводить слежение за объектом со скоростью 30 кадров в секунду.

СПИСОК ЛИТЕРАТУРЫ

1. Yilmaz A., Javed O., Shah M. Object tracking: A survey. *ACM Computer Surveys*, 2006, Vol. 38, No. 4, p. 45.
2. Schweitzer H., Bell J., Wu F. Very fast template matching. *Eur. Conf. on Comp. Vis.*, 2002, pp. 358–372.
3. Nejhum S, Ho J., Yang M. Online visual tracking with histograms and articulating blocks. *Comp. Vis. and Img. Underst.*, 2010, pp. 901–914.
4. Diaz-Ramirez V.H., Picos K., Kober V. Target tracking in nonuniform illumination conditions using locally adaptive correlation filters. *Optics Communications*, 2014, Vol. 323, pp. 32–43.
5. Haritaoglu I., Harwood D., Davis L. W4: real-time surveillance of people and their activities. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 2000, Vol. 22, No. 8, pp. 809–830 .
6. Talu F., Turkoglu I., Cebeci M. A hybrid tracking method for scaled and oriented objects in crowded scenes. *Exp. Syst. with App.*, 2011, Vol. 38, pp. 13682–13687.
7. Buchanan A., Fitzgibbon A. Document image dewarping using robust estimation of curled text lines. *Combining local and global motion models for feature point tracking*, 2007, pp. 1–8.
8. Sbalzarini I., Koumoutsakos P. Feature point tracking and trajectory analysis for video imaging in cell biology. *Jour. of Struct. Bio*, 2005, Vol. 151, No. 2, pp. 182–195.
9. Kalal Z., Mikolajczyk K., Matas J. Tracking-learning-detection. *IEEE Trans. Pat- tern Anal. Mach. Intell.*, 2012, Vol. 34, No. 7, pp. 1409–1422.
10. Babenko B., Ming-Hsuan Y. C., Belongie S. Visual tracking with online multiple instance learning. *IEEE Conf. on Comp. Vis. and Patt. Rec.*, 2009, pp. 983–990.
11. Song S., Xiao J. Cluster based weighted SVM for the recognition of Farsi handwritten digits. *Tracking revisited using rgbd camera: Unified benchmark and base- lines*, 2013, pp. 233–240.
12. Meshgi K., Maeda S., Oba S., Ishii S. Fusion of multiple cues from color and depth domains using occlusion aware bayesian tracker. *IEICE Tech. Rep. Neurocomp*, 2014, Vol. 114, No. 500, pp. 127–132.

13. Dalal N., Triggs B. Histograms of oriented gradients for human detection. *Comp. Vis. and Patt. Rec.*, 2005, Vol. 1, pp. 886–893.
14. Miramontes-Jaramillo D., Kober V., Daz-Ramrez V. CWMA: Circular window matching algorithm. *Proc. 18th Iberoam. Cong. in Patt. Rec.*, 2013, LNCS 8258, pp. 439–446.
15. Miramontes-Jaramillo D., Kober V., Daz-Ramrez V. A novel Image Matching Algorithm Based on Sliding Histograms of Oriented Gradients. *Journal of Communications Technology and Electronics*, 2014, Vol. 59, No. 12, pp. 1446–1450.
16. Gupta V. *Nonlinear filters with spatially connected neighborhoods*. Laxmi Publications, 2005.
17. Ramos E.M., Kober V. Design of correlation filters for recognition of linearly distorted objects in linearly degraded scenes. *Journal OSA A*, 2007, Vol. 24, No. 11, pp. 3403–3417.
18. Yaroslavsky L.P., Eden M. *Fundamentals of Digital Optics*. Birkhäuser, Boston, 1996.
19. Po-Ming L., Hung-Yi C. Adjustable gamma correction circuit for TFT LCD. *IEEE Symp. On Circ. and Syst.*, 2005, pp. 780–783.
20. Pratt W.K. *Digital Image Processing*. John Wiley & Sons, 2007.
21. Takacs G., Chandrasekhar V., Tsai S., Grzeszczuk R., Girod B. Distortion invariant pattern recognition with local correlations. *Fast computation of rotation-invariant image features by approximate radial gradient transform*, 2013, Vol. 22, No. 8, pp. 2970–2982.

A new descriptor-based tracking algorithm using a depth camera

Miramontes-Jaramillo D., Kober V., Diaz-Ramirez V., Karnaukhov V.

Recently, many algorithms of tracking have been proposed owing to the release of cheap and high-quality video and depth cameras. In this paper, a new descriptor-based tracking algorithm using depth information from Microsoft Kinect is proposed. The descriptor used by the algorithm is the histograms of oriented gradients computed over several circular sliding windows of the image scene. Additionally, depth information of the scene helps us to solve the problem of partial occlusion of the target. To speed up the tracking, a motion prediction model is used. To provide real-time tracking, the proposed algorithm is implemented on multicore processing units such as GPUs.

KEYWORDS: tracking, histogram of oriented gradients, real-time algorithm.