

# Детектирование объектов переднего плана в видео

---

Александр Новиков

*Video Group*

*CS MSU Graphics & Media Lab*



# Содержание

---

- **Введение**
- Обработка пространственно-временных срезов
- Проверка условий 3D-геометрии сцены
- Распространение пространственной и цветовой информации
- Использование информации о смене цветов на областях открытия
- Заключение

# Постановка задачи

Задача — выделить объекты переднего плана в видео

Проблемы:

- Движение камеры
- Необходимость сохранения временной целостности
- Изменение внешнего вида объекта во времени



Кадр последовательности



Выделенный объект



# Применения

---

- Редактирование видео
- Слежение за объектами
- Удаление объектов
- Конвертация 2D видео в 3D



# Содержание

---

- Введение
- **Обработка пространственно-временных срезов**
- Проверка условий 3D-геометрии сцены
- Распространение пространственной и цветовой информации
- Использование информации о смене цветов на областях открытия
- Заключение

# Пространственно-временные срезы (1)

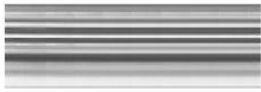


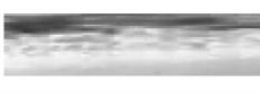

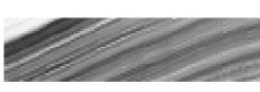
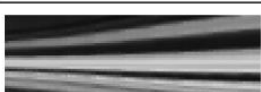

- Пространственно-временные срезы — это изображения, получаемые при перемещении в  $(x, t)$  [ $(y, t)$ ] измерениях при фиксированном  $y$  [ $x$ ]
- Срезы разделяются на
  - $H_{y=i}(x, t)$  — горизонтальные, фиксирован  $y=i$
  - $V_{x=j}(y, t)$  — вертикальные, фиксирован  $x=j$

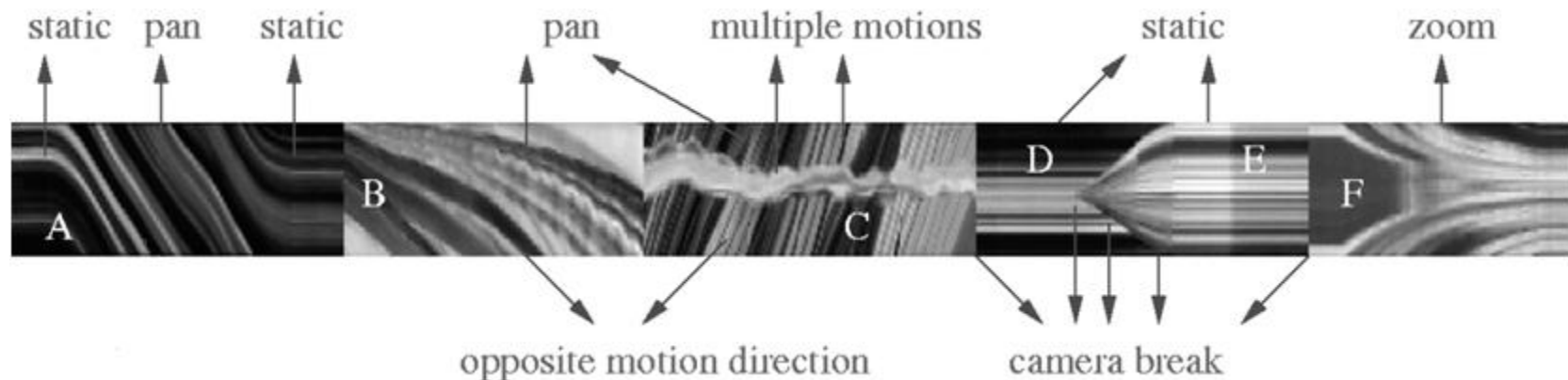


Горизонтальный срез видео

# Пространственно-временные срезы (2)

Структура пространственно-временных срезов позволяет определить вид и направление движения камеры

Motion Type	$t$	
	Horizontal Slice	Vertical Slice
static		
pan		
tilt		
zoom		



Ch.-W. Ngo, T.-Ch. Pong, H.-J. Zhang, "Motion Analysis and Segmentation Through Spatio-Temporal Slices Processing", IEEE ToIP, 2003

# Идея метода

- 1) Разбиение видео на участки с однородным движением камеры и классификация:
  - Отсутствие движения
  - Панорамирование
  - Приближение/отдаление
  - Сложное движение
- 2) Восстановление фона сцены
- 3) Background subtraction и выделение движущихся объектов



# Алгоритм

## Локальная структура срезов

Структура среза  $H$  в точке представляется матрицей  $\Gamma$

$$\Gamma = \begin{bmatrix} J_{xx} & J_{xt} \\ J_{xt} & J_{tt} \end{bmatrix}$$

$$J_{xx} = \sum_{x', t' \in w} \hat{H}_x^2(x - x', t - t')$$

$$J_{xt} = \sum_{x', t' \in w} \hat{H}_x(x - x', t - t') \hat{H}_t(x - x', t - t')$$

$$J_{tt} = \sum_{x', t' \in w} \hat{H}_t(x - x', t - t')$$

$$\hat{H}_x = \frac{\partial(G \otimes H)}{\partial x}, \hat{H}_t = \frac{\partial(G \otimes H)}{\partial t}$$

— частные производные срезов, сглаженных фильтром Гаусса

$w$  — окно  $3 \times 3$  с центром в некотором пикселе

# Алгоритм

## Локальная направленность

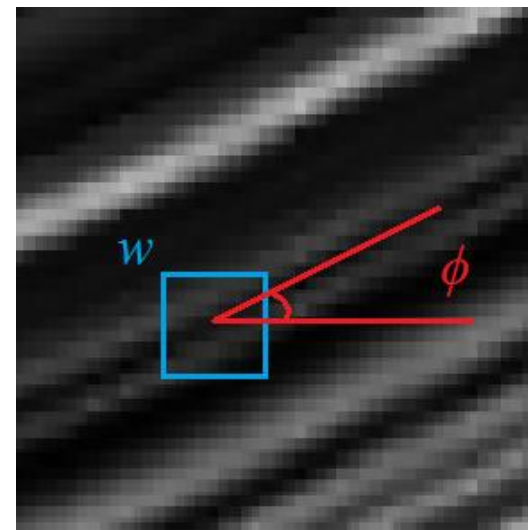
- Вычисляется локальная ориентация  $\phi$  в окне  $w$

$$\phi = \begin{cases} \theta - \frac{\pi}{2}, \theta > 0 \\ \theta + \frac{\pi}{2}, otherwise \end{cases} \quad \phi \in \left[ -\frac{\pi}{2}, +\frac{\pi}{2} \right] \quad \theta = \frac{1}{2} \tan^{-1} \frac{2J_{xt}}{J_{tt} - J_{xx}}$$

- $c$  — мера того, насколько точно  $\phi$  оценивает локальную ориентацию,

$c = 1$  — точная направленность

$$c = \frac{(J_{xx} - J_{tt})^2 + 4J_{xt}^2}{(J_{xx} + J_{tt})^2}, c \in [0,1]$$



# Алгоритм

## Гистограмма направленностей

- Строится двумерная гистограмма локальных направленностей

$$M(\hat{\phi}, t) = \begin{cases} \sum_i \sum_x c(x, t)_{y=i}, & \phi(x, t)_{y=i} = \hat{\phi} \\ 0, & otherwise \end{cases}$$

- На кадрах  $[k, k+N]$  выделяется доминирующая траектория  $\{\phi\}$ , если для них выполняется

$$\frac{\sum_{t=k}^{k+N} p(t)}{\sum_{t=k}^{k+N} \sum_{\phi} M(\phi, t)} > T$$

$$p(t) = \max_{-\pi/2 < \phi < \pi/2} \{M(\phi, t)\}$$

$N, T$  — константы, в статье 15 и 0.6

# Алгоритм

## Классификация движения камеры

- Неподвижная камера

$$\phi \in [-\phi_a, \phi_a] \quad \phi_a \approx 0$$

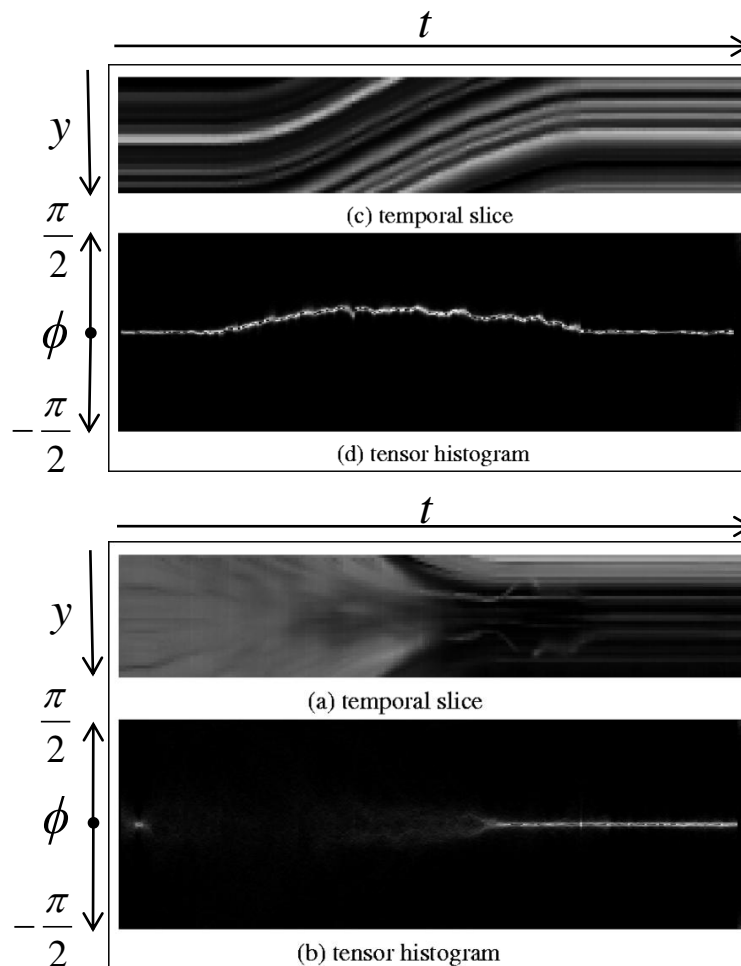
- Панорамирование

$$\phi > \phi_a, \phi < \phi_a$$

- Отдаление/приближение

$$\frac{\sum_t \sum_{\phi > 0} M(\phi, t)}{\sum_t \sum_{\phi < 0} M(\phi, t)} \approx 1$$

- Иначе — сложное движение



Ch.-W. Ngo, T.-Ch. Pong, H.-J. Zhang, "Motion Analysis and Segmentation Through Spatio-Temporal Slices Processing", IEEE ToIP, 2003

# Алгоритм

## Выделение фона на срезах

- Дискретизация гистограммы

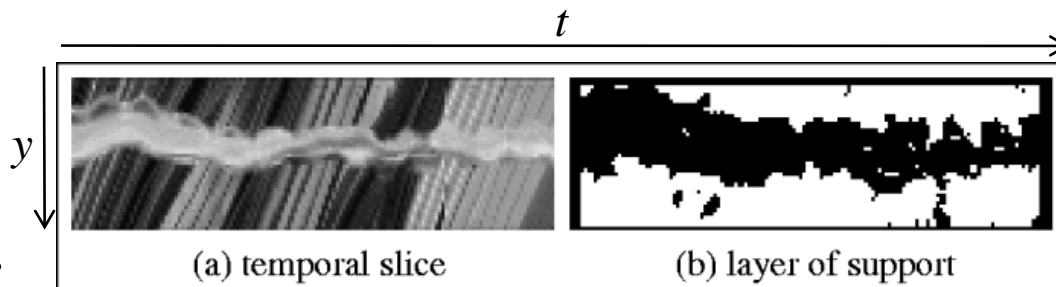
$$N(\Phi_k) = \sum_{\phi_i \in \Phi_k} \sum_t M(\phi_i, t)$$

$\Phi_1 = [-90^\circ, -45^\circ]$	$\Phi_5 = (5^\circ, 25^\circ]$
$\Phi_2 = [-45^\circ, -25^\circ]$	$\Phi_6 = (25^\circ, 45^\circ]$
$\Phi_3 = [-25^\circ, -5^\circ]$	$\Phi_7 = (45^\circ, 90^\circ]$
$\Phi_4 = [-5^\circ, 5^\circ]$	

- Выделение на срезах пикселей с движением близким к основному (*LoS* — layer of support)

$$\hat{\Phi} = \arg \left\{ \max_{\Phi_k} N(\Phi_k) \right\}$$

$$LoS(x, t)_{y=i} = \begin{cases} 1, & \phi \in \hat{\Phi} \\ 0, & otherwise \end{cases}$$



# Алгоритм

## Восстановление панорамы фона

- Пиксели с  $LoS = 1$  считаются принадлежащими фону
- Для каждого кадра  $t$  выбирается сдвиг

$$d(t) = \arg \min_d \left\{ \operatorname{med}_x \left| \hat{H}(x, t)_{y=i} - \hat{H}(x + d, t + 1)_{y=i} \right| \right\},$$

где  $\hat{H}_{y=i}$  — выбранный для кадра  $t$  срез с наибольшим количеством текстурной информации

$$\hat{H} = \arg \max_{H_{y=i}} \left\{ \frac{C(t)_{y=i} + C(t + 1)_{y=i}}{\left| n(t)_{y=i} - n(t + 1)_{y=i} \right| + 1} \right\}$$

$$C(t)_{y=i} = \sum_x c(x, t)_{y=i} LoS(x, t)_{y=i} \quad n(t)_{y=i} = \sum_x LoS(x, t)_{y=i}$$

# Алгоритм

## Color back-projection

- По  $LoS$  для фона инвертированием вычисляется  $LoS$  для переднего плана
- Строится дискретизованная гистограмма  $C$  цветов пикселей со значениями  $LoS$  переднего плана = 1
- Передний план  $F$  вычисляется по формуле:

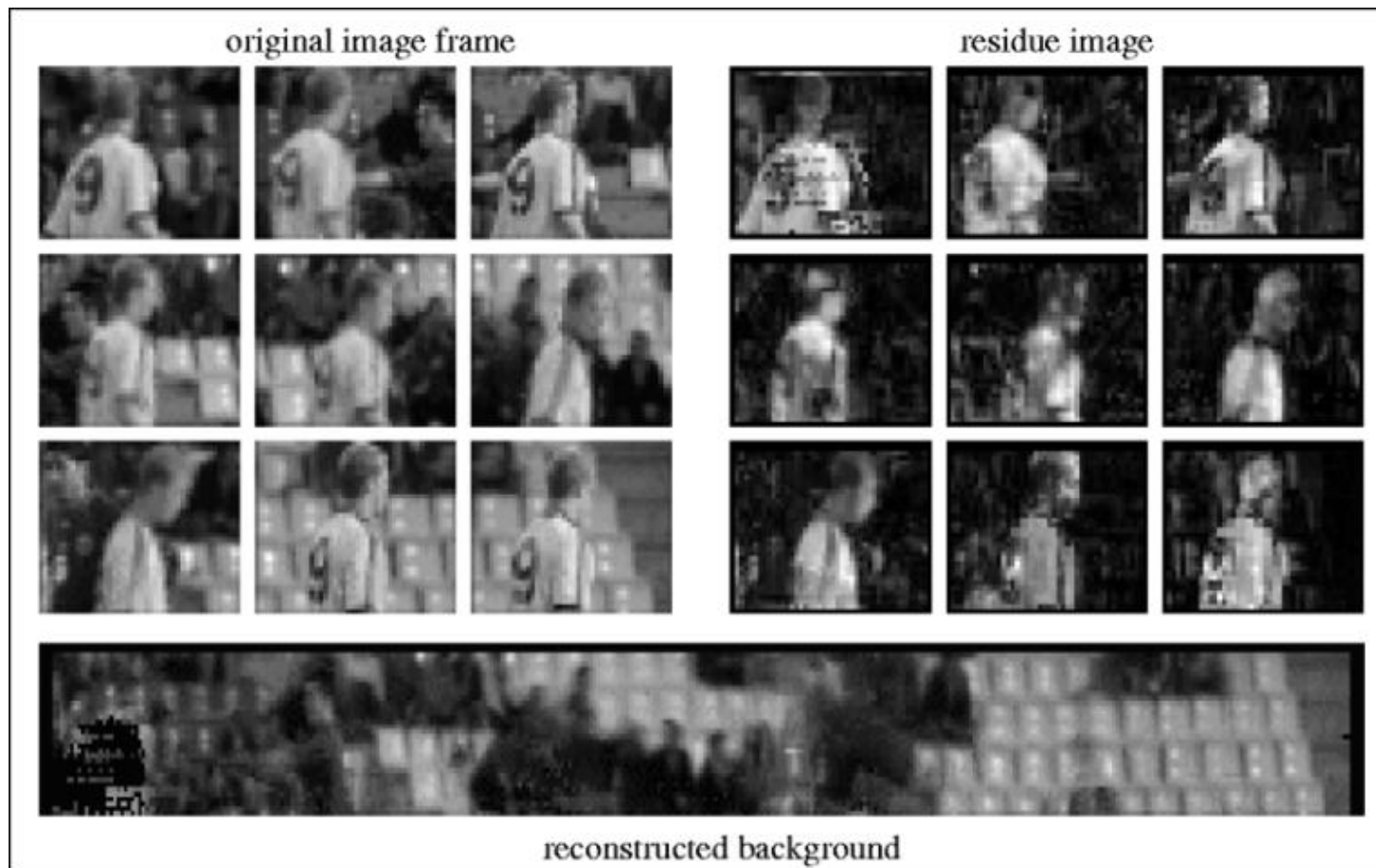
$$F(X, t) = \frac{1}{2} (\hat{C}(Image(X, t)) + \hat{R}(X, t)) \cdot Image(X, t)$$

$\hat{R}(X, t)$  — нормализованная карта разницы  $|Image-Background|$

$\hat{C}$  — нормализованная гистограмма  $C$

# Результаты (1)

## Background subtraction

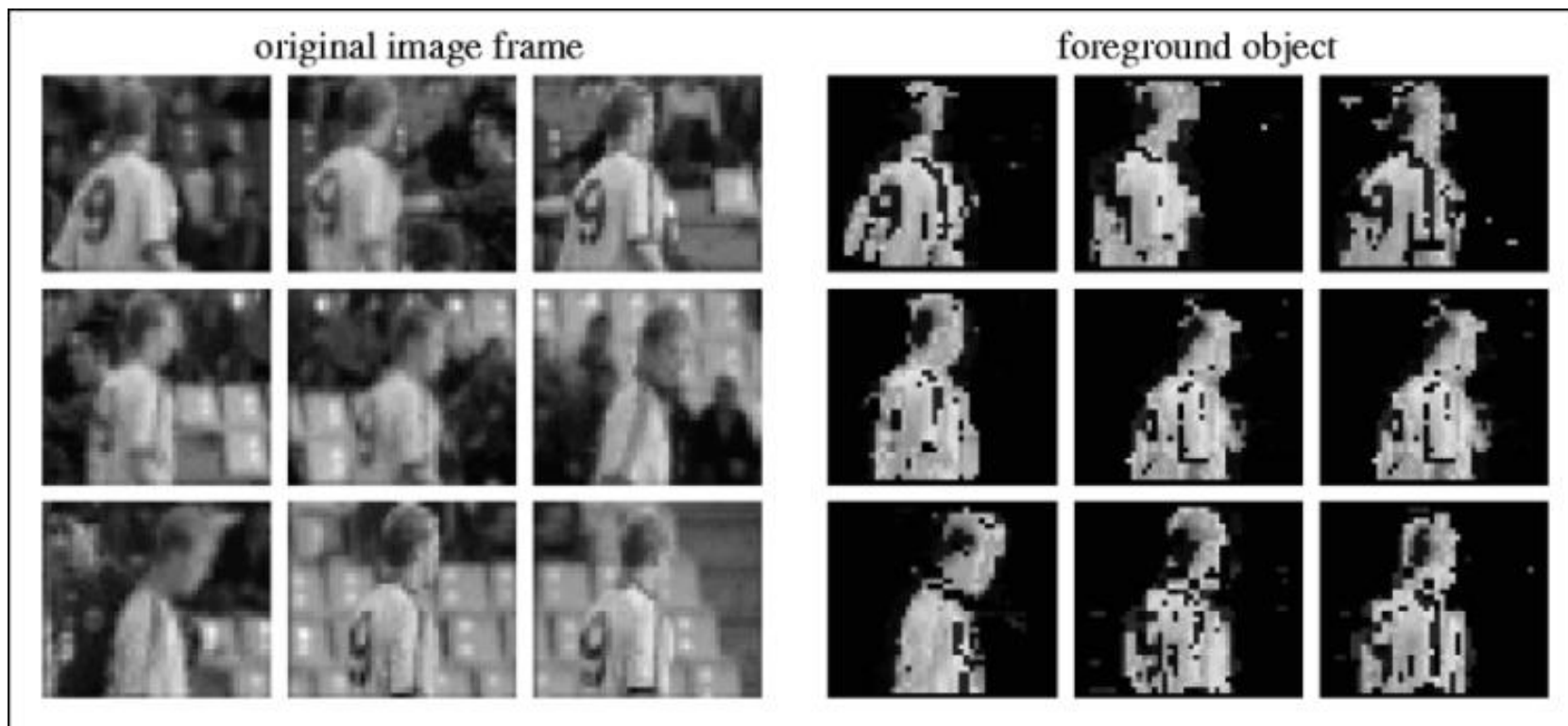


Ch.-W. Ngo, T.-Ch. Pong, H.-J. Zhang, "Motion Analysis and Segmentation Through Spatio-Temporal Slices Processing", IEEE ToIP, 2003



# Результаты (2)

## Color back-projection



Ch.-W. Ngo, T.-Ch. Pong, H.-J. Zhang, "Motion Analysis and Segmentation Through Spatio-Temporal Slices Processing", IEEE ToIP, 2003



# Выводы

---

## Достоинства:

- Скорость работы
- Восстановление фона

## Недостатки:

- Отсутствие временной целостности
- Недостаточно точная сегментация



# Содержание

---

- Введение
- Обработка пространственно-временных срезов
- **Проверка условий 3D-геометрии сцены**
- Распространение пространственной и цветовой информации
- Использование информации о смене цветов на областях открытия
- Заключение

# Схема работы алгоритма

- 1) Восстановление структуры сцены по движению
- 2) Оценка движения и глубины
  - При фиксированной сегментации оцениваются движение и области открытия
  - Строится карта глубины и карта ошибок геометрии
- 3) Выделение фона и переднего плана
- 4) Повторение второго и третьего шагов  $k$  (2-3) раз
- 5) Matting границ выделенных объектов

# Алгоритм

## Structure From Motion

- Выполняется поиск 3D-feature points
- Для каждого кадра вычисляются матрицы камеры

$$u_X^t = K^t (R^t X + T^t)$$

$u_X^t$  — проекция 3D-точки  $X$  на кадр  $t$   
 $K^t, R^t, T^t$  — матрицы камеры для кадра  $t$



Вставленные объекты демонстрируют определенную геометрию сцены

G. Zhang, X. Qin, W. Hua, T.-T. Wong, P.-A. Heng, H. Bao,  
"Robust metric reconstruction from challenging video sequences",  
IEEE CVPR, 2007

# Алгоритм

## Оценка движения (1)

Оценка движения и поиск областей закрытия выполняется решением задачи минимизации

$$\arg \min_{d,o} \sum_{t=1}^{n-1} (E^{t,t+1}(d,o) + E^{t+1,t}(d,o))$$

$$E^{t,t+1}(d,o) = \sum_{i \in f^t} [m^{t,t+1}(i) + \sum_{j \in N_i} s^{t,t+1}(i,j)] + Q^{t,t+1}(D)$$

$d^{t,t+1}(i)$  — motion-вектор пикселя  $i$  при переходе с кадра  $t$  к  $t+1$

$o_i^{t,t+1}(i)$  — индикатор принадлежности пикселя  $i$  области открытия при переходе от кадра  $t$  к  $t+1$

$N_i$  — множество 8 пикселей, соседних с  $i$ -м

$m^{t,t+1}(i), s^{t,t+1}(i,j)$  — условия соответствия данных, гладкости поля motion-векторов и соответствия 3D-feature points

$Q^{t,t+1}(D)$

# Алгоритм

## Оценка движения (2)

$$E^{t,t+1}(d, o) = \sum_{i \in f^t} [m^{t,t+1}(i) + \sum_{j \in N_i} s^{t,t+1}(i, j)] + Q^{t,t+1}(D)$$

Условие соответствия данных:

$$m^{t,t+1}(i) = \begin{cases} \rho_d^{t,t+1}(i), & o_i^{t,t+1} = 0, \alpha_i^t = \alpha_{i'}^{t+1} \\ \min\{\rho_d^{t,t+1}(i), \eta_o\}, & o_i^{t,t+1} = 0, \alpha_i^t \neq \alpha_{i'}^{t+1} \\ \eta_o, & o_i^{t,t+1} = 1 \end{cases} \quad \rho_d^{t,t+1}(i) = \frac{\|f^t(i) - f^{t+1}(i')\|^2}{\eta_d + \|f^t(i) - f^{t+1}(i')\|^2}$$

$i'$  — пиксель кадра  $t + 1$  соответствующий  $i$ -тому по  $d^{t,t+1}$

$f^t(i)$  — цвет пикселя  $i$  кадра  $t$

$\alpha_i^t(i)$  — индикатор принадлежности пикселя  $i$  кадра  $t$  переднему плану

$\eta_o, \eta_d$  — константы, в статье 0.5 и 400 соответственно

# Алгоритм

## Оценка движения (3)

$$E^{t,t+1}(d, o) = \sum_{i \in f^t} [m^{t,t+1}(i) + \sum_{j \in N_i} s^{t,t+1}(i, j)] + Q^{t,t+1}(D)$$

Условие гладкости поля motion-векторов:

$$s^{t,t+1}(i, j) = \beta_s \rho_s^{t,t+1}(i, j) + \beta_o |o_i^{t,t+1} - o_j^{t,t+1}| + \beta_w |o_i^{t,t+1} - W_i^{t,t+1}|$$

$$\rho_s^{t,t+1}(i, j) = \begin{cases} \min\{\|d^{t,t+1}(i) - d^{t,t+1}(j)\|^2, \eta_s\} & \alpha_i^t = \alpha_j^{t+1} \\ 0, & \alpha_i^t \neq \alpha_j^{t+1} \end{cases}$$

$W_i^{t,t+1}$  — 1, если для пикселя  $f^t(i)$  нет соответствующего по  $d^{t+1,t}$  пикселя в  $f^{t+1}$ , 0 — иначе

$\rho_s^{t,t+1}(i, j)$  — мера сонаправленности векторов  $d^{t,t+1}(i)$  и  $d^{t,t+1}(j)$

$\beta_s, \beta_o, \beta_w$  — балансирующие коэффициенты, 0.1, 0.21 и 0.6 в статье

$\eta_s$  — ограничение сверху, 4 в статье



# Алгоритм

## Оценка движения (4)

$$E^{t,t+1}(d, o) = \sum_{i \in f^t} [m^{t,t+1}(i) + \sum_{j \in N_i} s^{t,t+1}(i, j)] + Q^{t,t+1}(D)$$

Условие соответствия 3D-feature points:

$$Q^{t,t+1} = \beta_D \sum_{X \in D} \sum_{f_t, f_{t+1} \in \varphi(X)} \left\| d^{t,t+1}(u_X^t) - (u_X^{t+1} - u_X^t) \right\|^2$$

$D$  — множество всех 3D-feature points

$u_X^t$  — проекция 3D-точки  $X$  на кадр  $t$

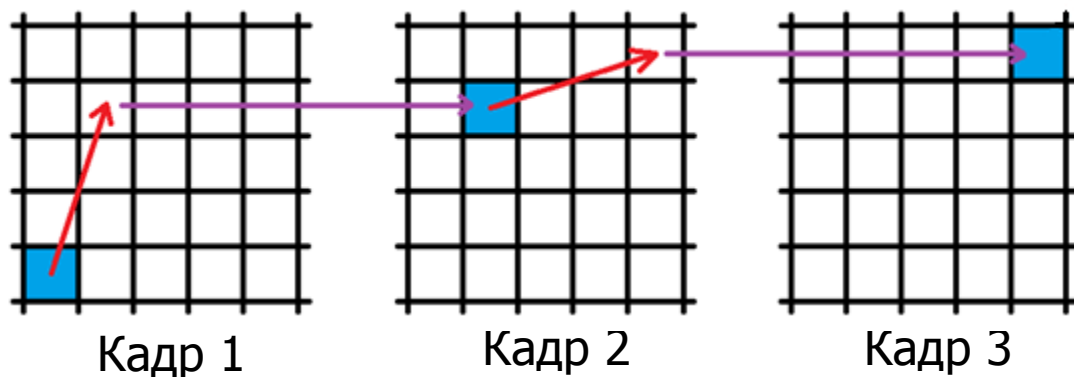
$\varphi(X)$  — множество кадров, на которых  $X$  имеет соответствующую feature point

$\beta_D$  — коэффициент, берется большим, в статье равен 100

# Алгоритм

## Карта глубины и карта ошибок (1)

- Трек — последовательность пикселей видео, связанных по векторам движения
- Связь между пикселями  $i$  и  $i'$  разрывается, если:
  - Хотя бы один из них отмечен как область открытия
  - Ошибка согласованности optical flow  $e_{flow}^{t,t+1}(i) = \|d^{t,t+1}(i) + d^{t+1,t}(i')\|$  выше некоторого предела (2.0 в статье)



Красные стрелки —  
motion-векторы  
Синим отмечены  
пиксели трека

G. Zhang, J. Jia, W. Xiong, T.-T. Wong, P.-A. Heng, H. Bao,  
"Moving object extraction with a hand-held camera", IEEE ICCV,  
2007

# Алгоритм

## Карта глубины и карта ошибок (2)

- Все пиксели  $x_p^t$  трека  $p$  от кадра  $f_l$  до  $f_r$  должны соответствовать одной 3D-точке  $X_p$

$$x_p^t = K^t (R^t X_p + T^t)$$

- $X_p$  оценивается минимизацией среднеквадратического отклонения (RMSE)

$$\arg \min_{X_p} \sqrt{\frac{1}{r-l+1} \sum_{t=l}^r \|x_p^t - K^t (R^t X_p + T^t)\|^2}$$

- Карта глубины для кадров берется из  $z$  координат точек треков, для всех пикселей записывается RMSE в карту ошибок  $\gamma^t$

$$[x_p^t, y_p^t, z_p^t] = R^t X_p + T^t$$

# Алгоритм

## Результаты (1)

- Большое значение ошибки — пиксель находится на объекте переднего плана
- Проблема — карта ошибок зашумлена на границах объектов



Кадр



Полученная карта глубины  
Ярче — дальше



Карта ошибок  
Ярче — больше

G. Zhang, J. Jia, W. Xiong, T.-T. Wong, P.-A. Heng, H. Bao,  
"Moving object extraction with a hand-held camera", IEEE ICCV,  
2007

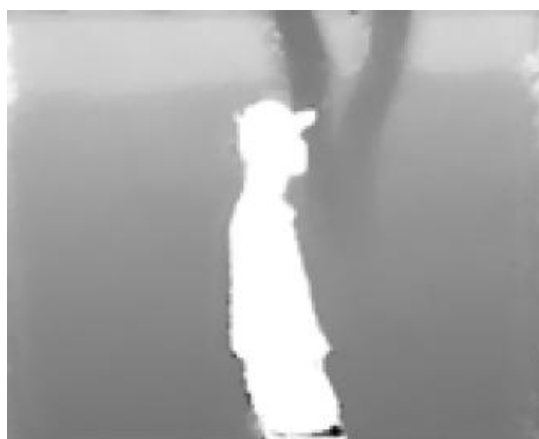
# Алгоритм

## Результаты (2)

В случае слежения камеры за объектом строится неверная карта глубины, и по карте ошибок нельзя выделить объект

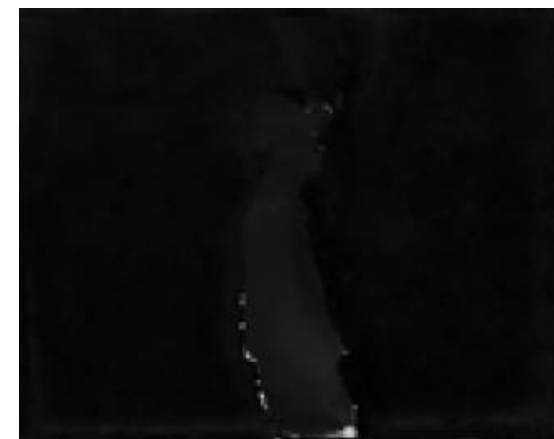


Кадр



Полученная карта глубины

Ярче — дальше



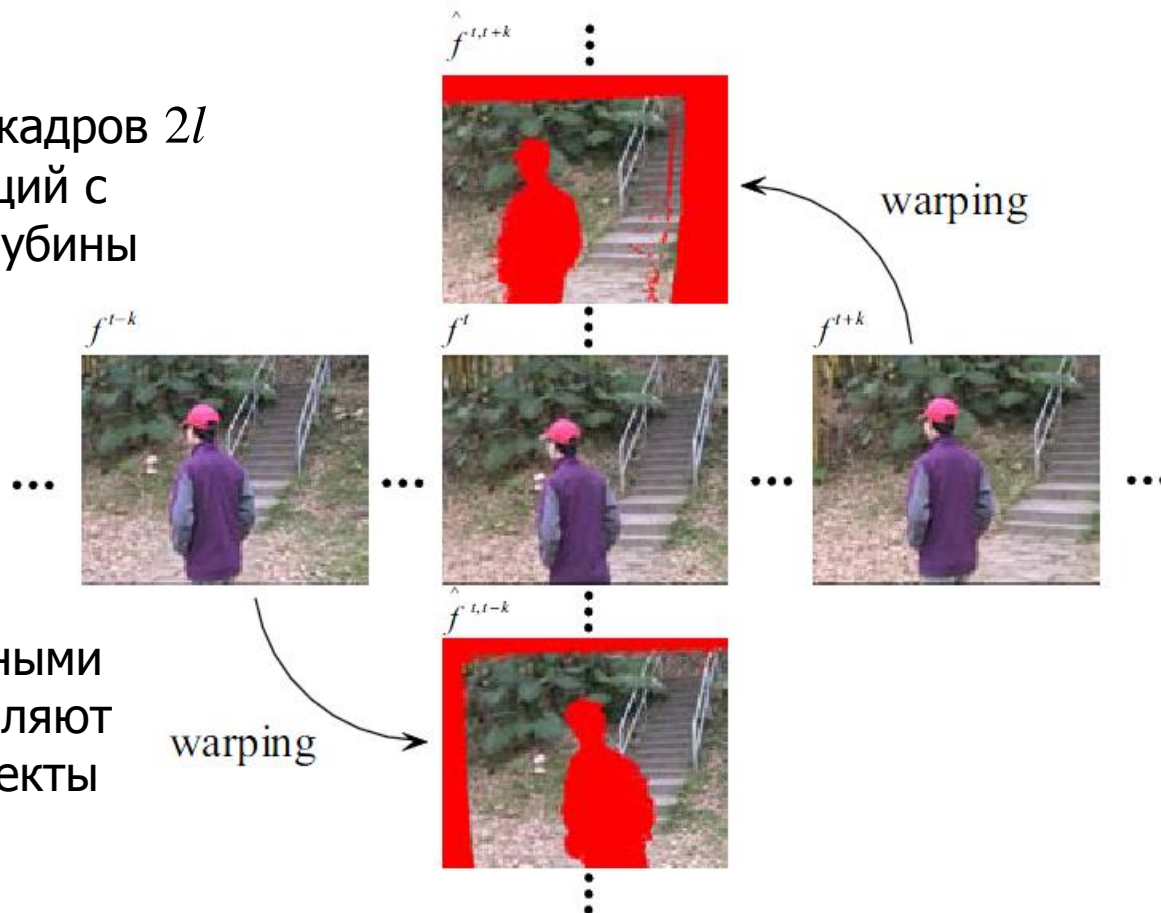
Карта ошибок

Ярче — больше

G. Zhang, J. Jia, W. Xiong, T.-T. Wong, P.-A. Heng, H. Bao,  
"Moving object extraction with a hand-held camera", IEEE ICCV,  
2007

# Алгоритм 3D-warping

Выполняется 3D-warping кадров  $2l$  соседних кадров на текущий с использованием карты глубины



Различия между полученными кадрами и текущим позволяют выявить движущиеся объекты

Красным отмечены точки, не имеющие проекции

G. Zhang, J. Jia, W. Xiong, T.-T. Wong, P.-A. Heng, H. Bao,  
"Moving object extraction with a hand-held camera", IEEE ICCV,  
2007

# Алгоритм Сегментация

Разделение на фон и передний план выполняется минимизацией функции энергии по  $\alpha$ :

$$E_B(\alpha) = \sum_{t=1}^n \sum_{i \in f^t} (L^t(i) + \lambda_T G_T^t(i) + \lambda_S \sum_{j \in N_i} G_S^t(i, j))$$

$L^t$  — функция правдоподобия данных

$G_T^t$  — условие временной гладкости

$G_S^t$  — условие пространственной гладкости

$N_i$  — множество 8 соседних с  $i$ -м пикселей

$\lambda_T, \lambda_S$  — балансирующие коэффициенты

# Результаты (1)



Кадр последовательности

Выделенный объект переднего плана

G. Zhang, J. Jia, W. Xiong, T.-T. Wong, P.-A. Heng, H. Bao,  
"Moving object extraction with a hand-held camera", IEEE ICCV,  
2007



# Результаты (2)



Кадр последовательности

Выделенный объект переднего плана

G. Zhang, J. Jia, W. Xiong, T.-T. Wong, P.-A. Heng, H. Bao,  
"Moving object extraction with a hand-held camera", IEEE ICCV,  
2007



# Выводы

---

## Достоинства:

- Высокая точность сегментации
- Восстановление фона за объектами

## Недостатки:

- Очень низкая скорость работы
- Основан на Structure From Motion



# Содержание

---

- Введение
- Обработка пространственно-временных срезов
- Проверка условий 3D-геометрии сцены
- **Распространение пространственной и цветовой информации**
- Использование информации о смене цветов на областях открытия
- Заключение

# Идея метода

- Выделить кадры со значительными областями с движением, отличным от глобального
- Выделить на них двигающиеся части объектов переднего плана (подобъекты)
- Собрать информацию о цвете и расположении подобъектов, распространить их во времени

# Алгоритм

## Motion cues

Motion cue – разница между локальным движением и глобальным смещением камеры

- Оценка движения
  - Глобального (ищутся feature points алгоритмом SIFT, устанавливаются соответствия и строится преобразование) —  $m_o(x, y)$
  - Локального (optical flow) —  $m_g(x, y)$
- Вычисление motion cues

$$mc(x, y) = \left\| m_o(x, y) - m_g(x, y) \right\|_2^2$$

# Алгоритм

## Выделение ключевых кадров

Ключевым называется кадр, на котором объект, либо его часть совершает заметное движение, отличное от глобального

Критерии:

- Большое количество пикселей с движением, отличным от глобального

$$\sum (mc(x, y) \geq \delta) > minArea$$

- Их компактное расположение

$$\text{Var}((x, y) | mc(x, y) \geq \delta) < maxSpan$$

$maxSpan, minArea, \delta$  — константы

# Алгоритм

## Выделение движущихся подобъектов (1)

Для каждого кадра строится функция правдоподобия для произвольного распределения меток “фон” и “движущийся объект”

На основании информации о цвете:

$$p(l_i | i = 1, \dots, M) = \prod_{i \in \{1, \dots, M\}} \prod_{j \in N_i} \psi(l_i, l_j) \quad \psi(l_i, l_j) = \exp\left(\frac{\lambda_i l_j}{(\alpha + d(i, j))}\right)$$

$l_i$  — метка  $i$ -го пикселя (-1 — движущийся объект, 1 — фон)

$N_i$  — множество 8 пикселей, соседних с  $i$ -м

$d(i, j)$  — разница по цвету между пикселями  $i$  и  $j$

$\lambda, \alpha$  — константы

# Алгоритм

## Выделение движущихся подобъектов (2)

На основании информации о движении:

$$p(I | \{l_i, i = 1, \dots, M\}) = \prod_{i \in \{1, \dots, M\}} p_m(mc_i, l_i)$$

$$p_m(mc_i | l_i) = \exp(l_i(mc_i - \delta_m))$$

$M$  — количество пикселей в изображении

$l_i$  — метка  $i$ -го пикселя (-1 — движущийся объект, 1 — фон)

$N_i$  — множество 8 пикселей, соседних с  $i$ -м

$\delta_m$  — константа



# Алгоритм

## Выделение движущихся подобъектов (3)

- Объединяя информацию о цвете и движении, получим функцию правдоподобия

$$P(L | I) = \prod_{i \in I} p_m(mc_i | l_i) \prod_{i \in I} \prod_{j \in N_i} \psi(l_i, l_j)$$

$L$  — множество всех меток

$I$  — множество всех пикселей кадра

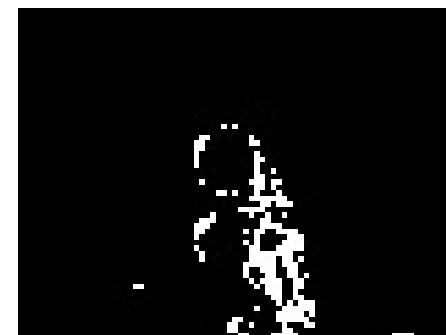
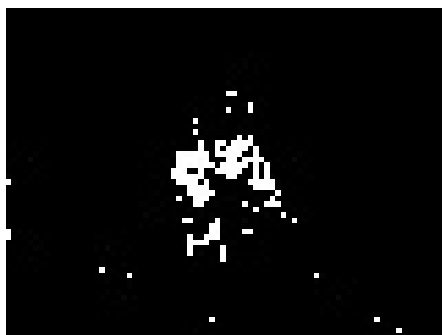
- Для каждого ключевого кадра находится набор меток  $l_i$ , её максимизирующий
- Выделяются подобъекты — движущиеся на конкретном ключевом кадре части объектов

# Алгоритм

## Ключевые кадры и motion cues



Key frames



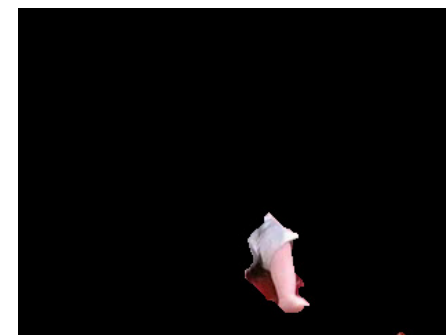
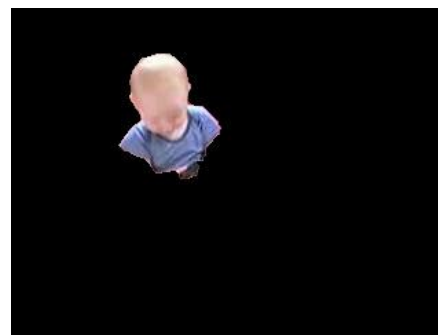
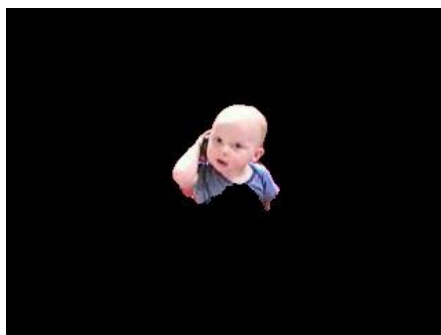
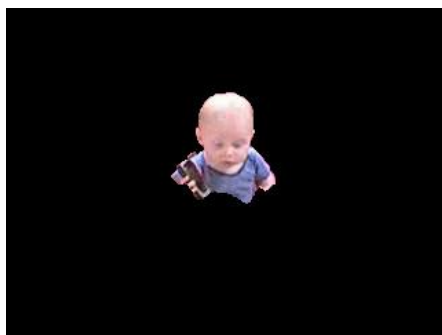
Motion cues

# Алгоритм

## Ключевые кадры и подобъекты



Key frames



Moving sub-objects

# Алгоритм

## Близость по цвету

- Строится распределение цветов в виде Gaussian Mixture Model (GMM)

$$G_f = \{g_i\}, i = 1, \dots, n$$

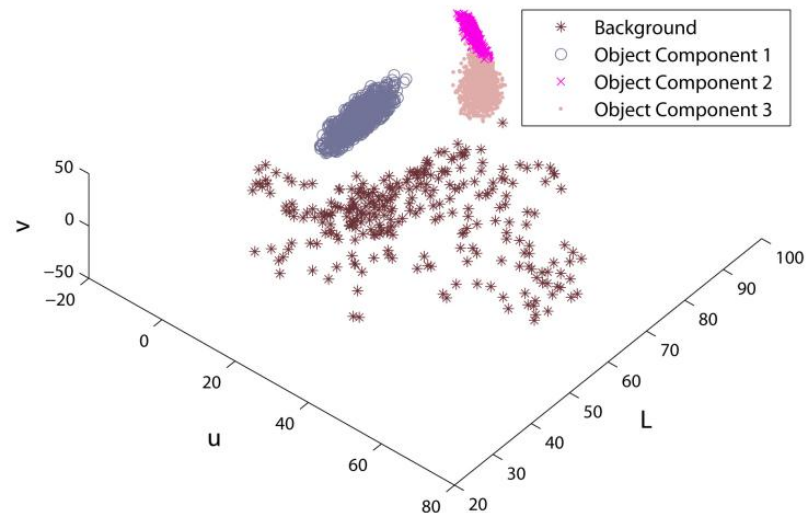
- Функция близости пикселя с цветом  $c$  к цветовой модели

$$\text{affi}_c(c) = \max_{g_j \in G_f} g_j(c)$$

- Функция правдоподобия сегментации с точки зрения цветовой модели

$$p_c(c_i | l_i) = \exp(l_i (\log(\text{affi}_c(c_i)) - \delta_c))$$

$\delta_c$  — константа

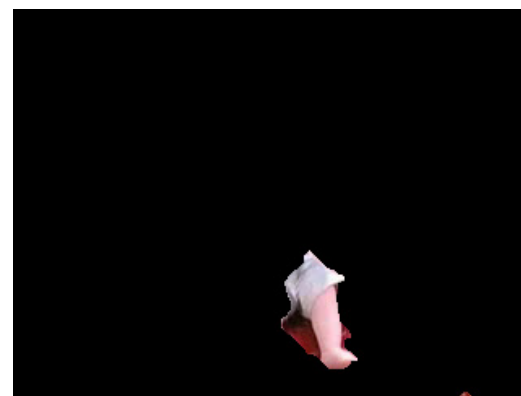


Распределение  
цветов объектов

# Алгоритм

## Удаление фоновых компонент из GMM

- Из-за ошибок выделения подобъектов в GMM могли появиться компоненты  $g_i$  связанные с фоном
- Вычисляется мера близости компоненты GMM к фону сцены
- Из GMM исключаются все  $g_i$  со значениями близости, превышающими порог



# Алгоритм

## Пространственная близость

- Вводится мера пространственной близости точки к выделенным подобъектам и пространственная функция правдоподобия сегментации

$$\text{affi}_s^t(x_i) = \frac{1}{2\pi\sigma^2} \max_{j \in F} \exp\left(-\frac{(x_i - x_j)^T (x_i - x_j)}{2\sigma^2}\right)$$

$$p_s^t(x_i | l_i) = \exp(l_i (\log(\text{affi}_s^t(x_i)) - \delta_s))$$

$F$  — движущийся объект

$\sigma, \delta_s$  — константы

- Вводится обновлённая функция правдоподобия, и еще раз выполняется сегментация ключевых кадров

$$\log p^t(f_i | l_i) = \log p_m^t(mc_i | l_i) + \lambda_c \log p_c^t(c_i | l_i) + \lambda_s \log p_s^t(x_i | l_i)$$

$\lambda_s, \lambda_c$  — балансирующие коэффициенты

# Алгоритм

## Распространение пространственной информации



Для каждого ключевого кадра  $k$ :

- Распространить пространственную информацию на кадры  $k+1, k+2, \dots$  следующим образом:
  1. Инициализировать пространственную функцию правдоподобия на кадре  $t+1$  функцией кадра  $t$
  2. Выделить движущиеся объекты обновленной функцией правдоподобия
  3. По выделенным объектам обновить пространственную информацию и функцию правдоподобия кадра  $t+1$
  4. Если кадр  $t+1$  ключевой и выделенные на шаге 2 объекты покрывают выделенные ранее, удалить кадр из ключевых
- Аналогично распространить на кадры  $k-1, k-2, \dots$

# Результаты (1)



Кадры последовательности



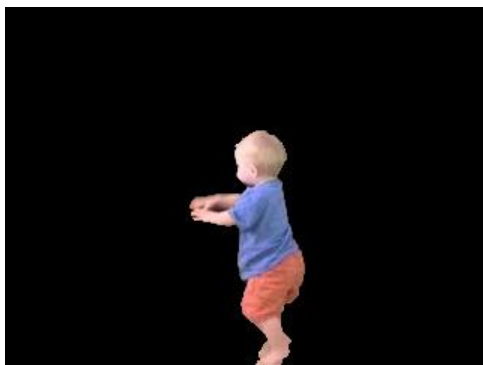
Результаты сегментации



# Результаты (2)



Кадры последовательности



Результаты сегментации



# Выводы

---

## Достоинства:

- Качество сегментации
- Временная целостность выделенных объектов

## Недостатки:

- Время работы
- Проблемы при близости объектов к фону по цвету



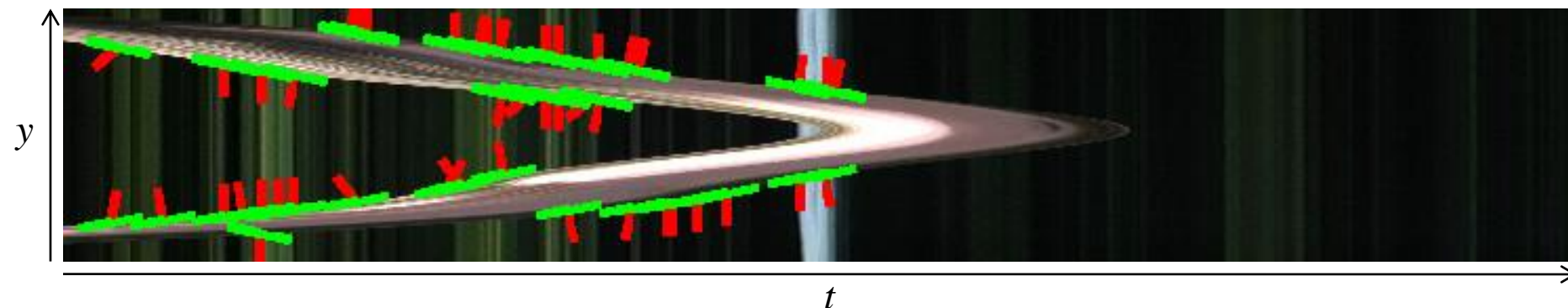
# Содержание

---

- Введение
- Обработка пространственно-временных срезов
- Проверка условий 3D-геометрии сцены
- Распространение пространственной и цветовой информации
- **Использование информации о смене цветов на областях открытия**
- Заключение

# T-Junctions

T-Junction — это фигура из двух отрезков, формирующих букву Т, расположенная на местах резкой смены градиентов на изображении

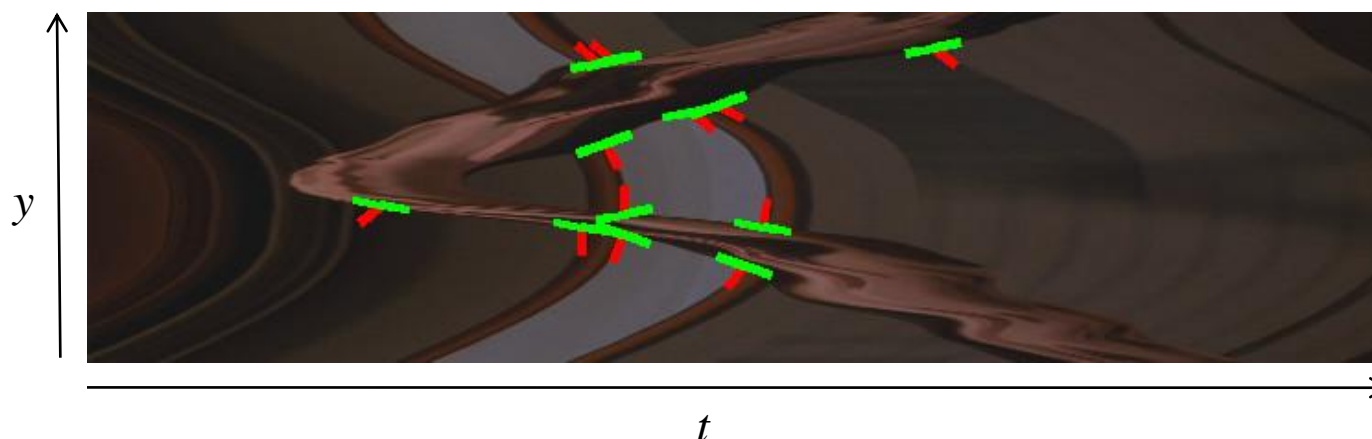


Пример выделенных T-Junctions. Заметим, что красные отрезки лежат на фоне сцены, зеленые — на движущихся объектах

# Алгоритм

## Поиск T-Junctions

- Первым шагом выполняется поиск T-Junctions, которые выступают как индикаторы областей открытия
- Таким образом получены спаренные пробы пикселей из переднего плана и фона



# Алгоритм

## Использование цветовой информации

По выделенным T-Junctions строятся GMM-модели

- Распределения цветов фона и переднего плана сцены
- **Смены цвета на областях открытия**



Кадры последовательности



Модель областей закрытия: каждый столбец соответствует одному варианту смены цвета

# Алгоритм Сегментация

Разделение на фон и передний план происходит минимизацией для каждого кадра функции энергии:

$$E = U + W + V$$

$U$  — отвечает за близость выделенных объектов по цвету к цветовой модели

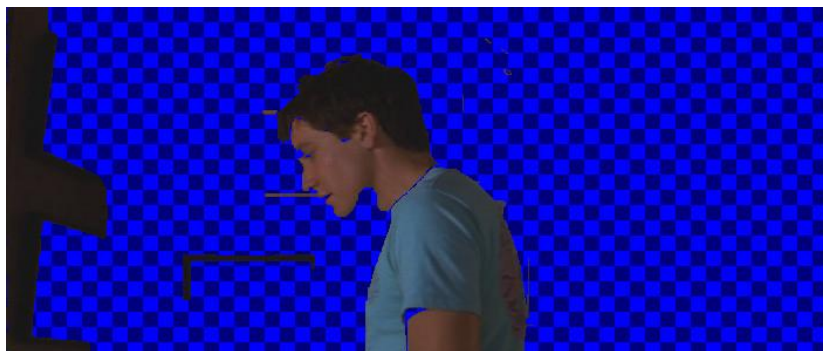
$W$  — условие соответствия модели смены цветов пикселей на областях открытия

$V$  — условие пространственной гладкости

# Результаты (1)



Кадр последовательности



Сегментация без условия  
смены цветов



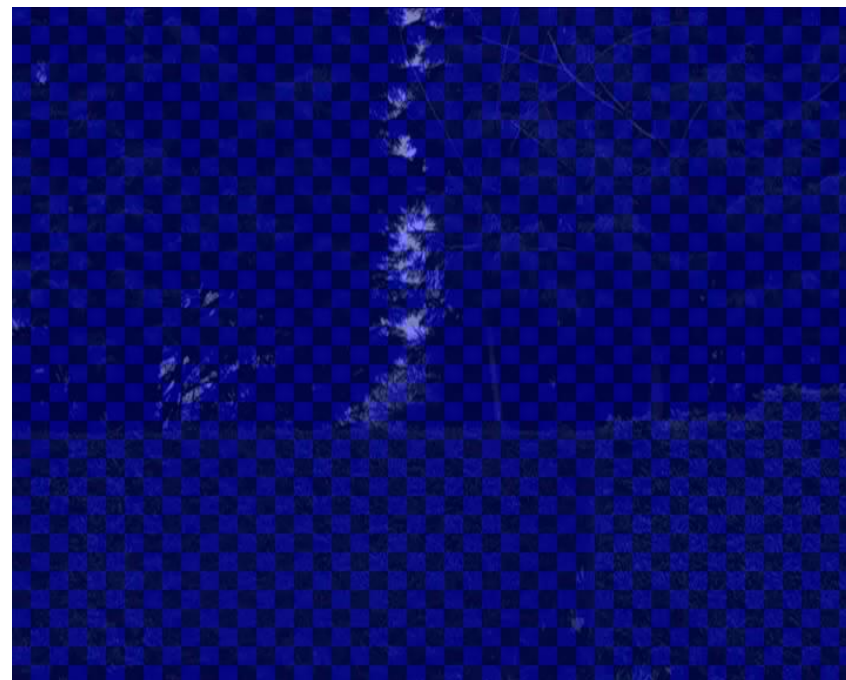
Сегментация с использованием  
всех условий



# Результаты (2)



Исходное видео



Результат сегментации

# Результаты (3)



Исходное видео



Результат сегментации



# Выводы

---

## Достоинство:

- Скорость работы
- Интересная идея решения задачи

## Недостатки:

- Основное движение должно быть горизонтальным
- Проблемы с временной целостностью



# Содержание

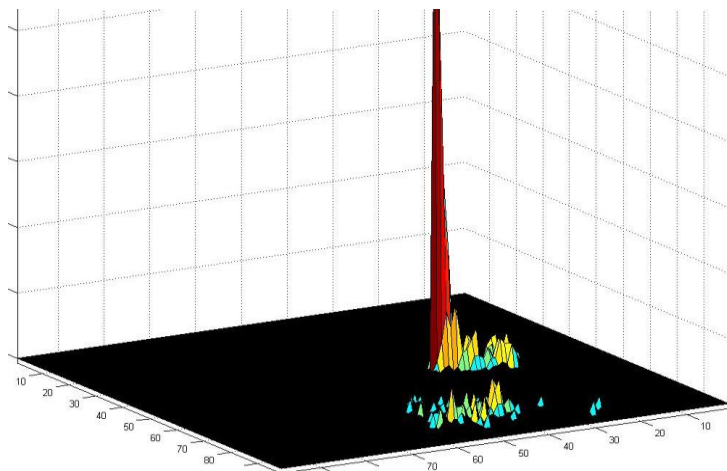
---

- Введение
- Обработка пространственно-временных срезов
- Проверка условий 3D-геометрии сцены
- Распространение пространственной и цветовой информации
- Использование информации о смене цветов на областях открытия
- **Заключение**

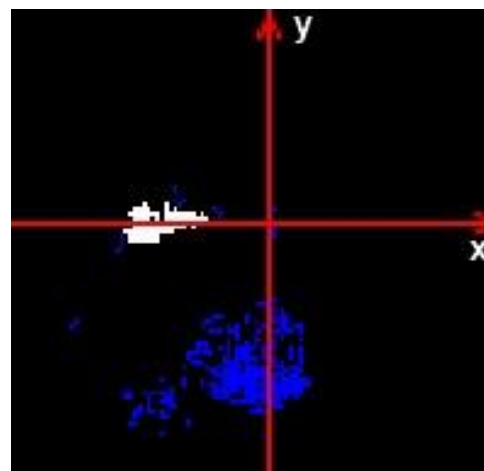
# Дальнейшие планы (1)

Алгоритм:

1. Оценка движения
2. Построение гистограммы векторов ME
3. Кластеризация гистограммы (Mean-shift)
4. Наибольший кластер — фон, остальные — движущиеся объекты



Визуализация гистограммы (высоты соответствуют количеству векторов с направлением)



Ортогональная проекция гистограммы (белые – векторы движения фона, синие – переднего плана)



# Дальнейшие планы (2)



Результаты сегментации

# Дальнейшие планы (3)

Решить проблемы:

- Разбиения на чрезмерно большое количество кластеров (пространственным объединением кластеров)
- Отсутствия временной целостности

# Литература

1. Ch.-W. Ngo, T.-Ch. Pong, H.-J. Zhang, "Motion Analysis and Segmentation Through Spatio-Temporal Slices Processing", IEEE Transactions on Image Processing, vol. 2, 2003
2. G. Zhang, X. Qin, W. Hua, T.-T. Wong, P.-A. Heng, H. Bao, "Robust metric reconstruction from challenging video sequences", IEEE CVPR, 2007
3. G. Zhang, J. Jia, W. Xiong, T.-T. Wong, P.-A. Heng, H. Bao, "Moving object extraction with a hand-held camera", IEEE ICCV, 2007
4. F. Liu, M. Gleicher, "Learning color and locality cues for moving object detection and segmentation", IEEE CVPR, 2009
5. N. Apostoloff, A. Fitzgibbon, "Automatic video segmentation using spatiotemporal T-junctions", BMVC, 2007
6. N. Apostoloff, A. W. Fitzgibbon, "Learning spatiotemporal T-junctions for occlusion detection", IEEE CVPR, 2005
7. V. Kolmogorov, R. Zabih, "What energy functions can be minimized via graph cuts", IEEE ECCV, 2002



# Лаборатория компьютерной графики и мультимедиа



Видеогруппа — это:

- Выпускники в аспирантурах Англии, Франции, Швейцарии (в России в МГУ и ИПМ им. Келдыша)
- Выпускниками защищено 5 диссертаций
- Наиболее популярные в мире сравнения видеокодеков
- Более 3 миллионов скачанных фильтров обработки видео