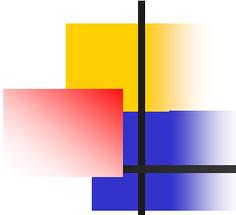


Вычитание фона с использованием карт глубины

Максим Харенко

Video Group

CS MSU Graphics & Media Lab



Содержание

- **Multi-modal background subtraction using Gaussian mixture models**
- Robust background subtraction for depth generation

Multi-modal background subtraction using GMM

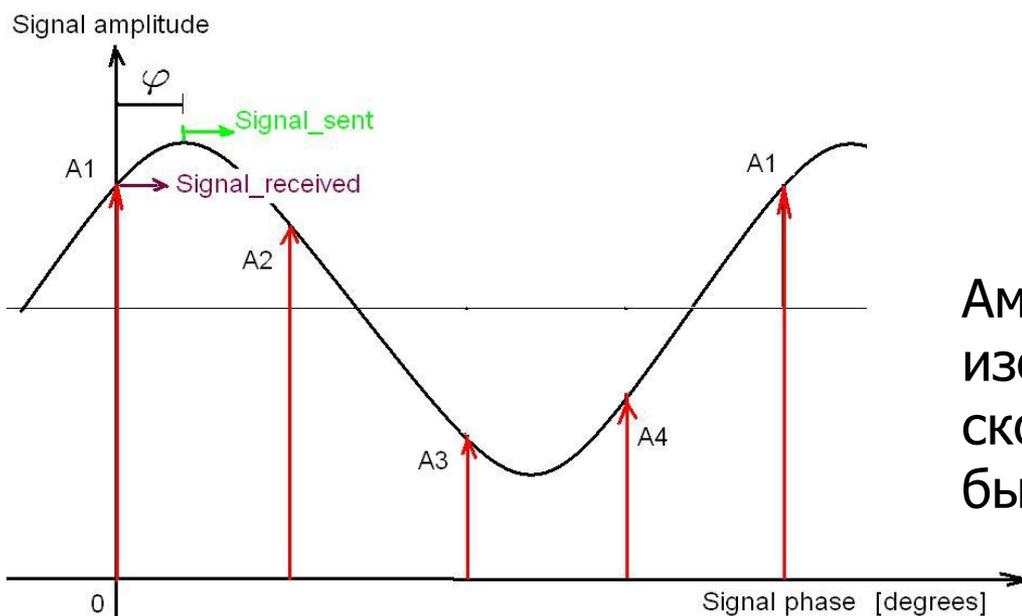
- Описание задачи
- GMM
- Реализация
- Сравнение результатов
- Выводы



Описание задачи

PMD-камера

2D/3D камера с PMD технологией, использующей ToF-принцип, выдает цветные изображения 640x480, а так же изображения с низким разрешением 64x48 глубины и амплитудной модуляции



Амплитудная модуляция – изображение, показывающее сколько инфракрасного света было принято PMD-чипом

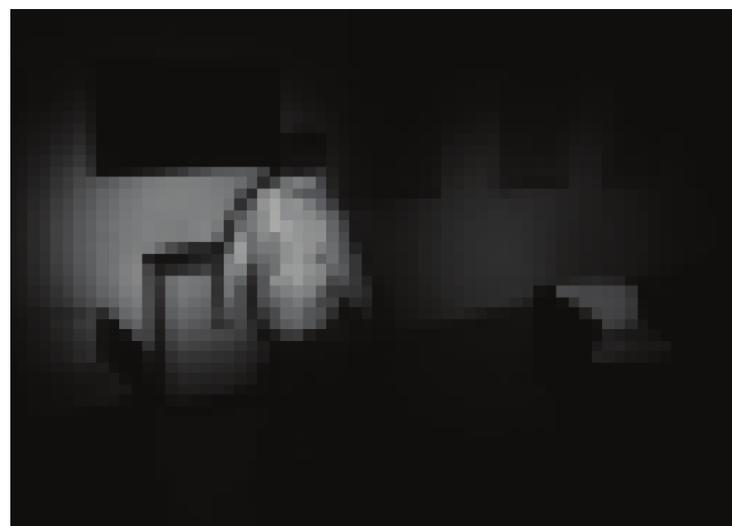
Описание задачи

Амплитудная модуляция

Исходное изображение



Амплитудная модуляция



Описание задачи

Введение

Требуется изолировать регионы, которые определяются движением переднего плана относительно неподвижного фона

Для этого:

- оценивается и сегментируется фон
- сегментируются объекты переднего плана

Описание задачи

Введение

Фон оценивается смесью нормальных распределений (GMM), работающей с данными из:

- цветного изображения
- карты глубины
- амплитудной модуляции

Описание задачи

Введение

Результаты, в виде отсегментированного переднего плана оцениваются по:

- распознаванию и обработке теней
- снижению количества блочных артефактов
- верному определению движения

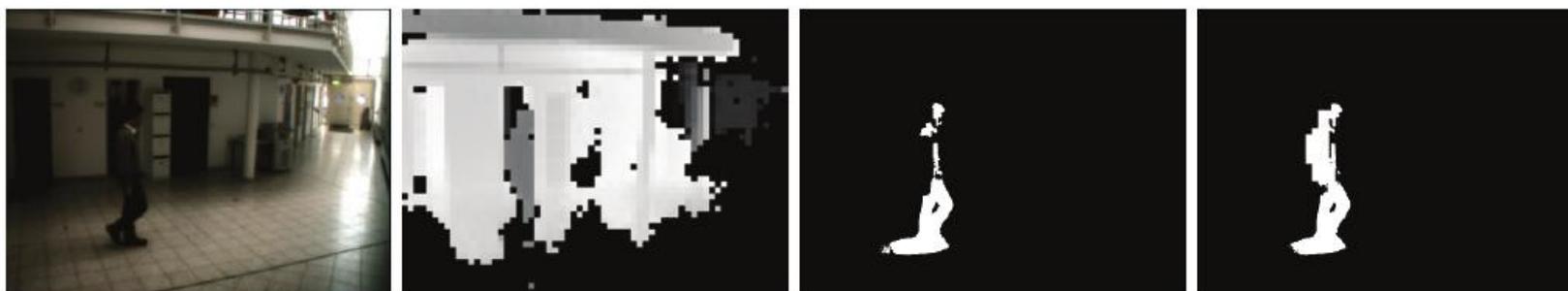
Описание задачи

Введение

- Тени и шум



- Блочные артефакты



(a) Input video

(b) Depth image

(c) GMM on color

(d) GMM on color and depth

B. Langmann S. E., Ghobadi, K. Hartmann, "Multi-modal background subtraction using Gaussian mixture models", IAPRS, Vol. XXXVIII, Part 3A, 2010

GMM

- Все наблюдения о каждом пикселе в положении $\underline{x} = [x, y]^T$ моделируются с помощью смеси гауссиан
- Каждое наблюдение моделируется одной гауссианой, чьи матожидание и дисперсия адаптированы по времени
- Наблюдение пикселя \underline{x} в момент времени t :

$$\underline{s}^t(\underline{x}) = [s_1^t(\underline{x}), s_2^t(\underline{x}), \dots, s_n^t(\underline{x})]^T$$

GMM

Плотность распределения вероятности:

$$f_{st}(\underline{x}) (\underline{\xi}) = \sum_i \omega_i \cdot N_{\underline{\xi}} (\underline{\mu}_i, \Sigma_i)$$

ω_i – коэффициенты смешивания, $\sum_i \omega_i = 1$

$\underline{\mu}_i$ – математическое ожидание

Σ_i – матрица ковариации

$$N_{\underline{\xi}} (\underline{\mu}_i, \Sigma_i) = \left[(2\pi)^{\frac{\dim(s)}{2}} \cdot \det(\Sigma_i)^{-\frac{1}{2}} \right]^{-1} \cdot \exp \left\{ -\frac{1}{2} [\underline{\xi} - \underline{\mu}_i]^T \cdot \Sigma_i^{-1} \cdot [\underline{\xi} - \underline{\mu}_i] \right\}$$

GMM

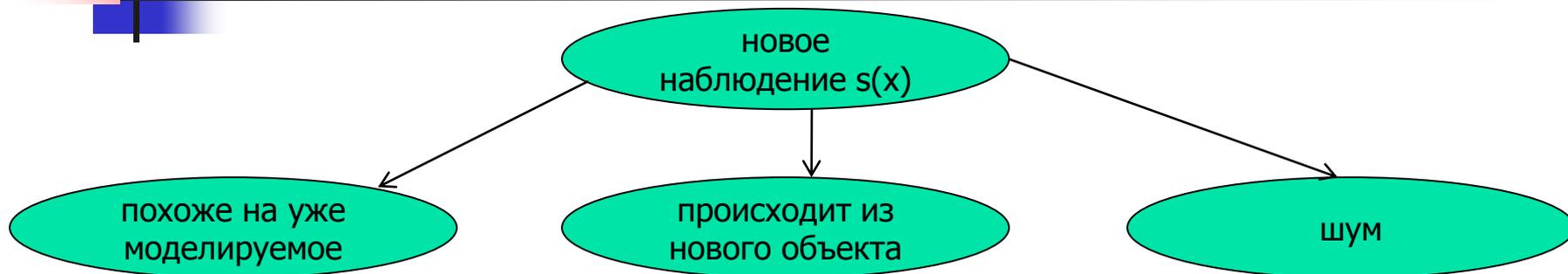
Проблемы

- Сколько нужно гауссиан для моделирования наблюдения?
ограниченное количество гауссиан на пиксель
- Какую гауссиану следует рассматривать в качестве фона?
Минимальное число гауссиан с весами, сумма которых выше заданного порога, рассматриваются в качестве фона
- Как адаптировать гауссианы с течением времени?

"Multy-modal background substruction using gaussian mixture models" , Benjamin Langmann, Seyed E. Ghobadi, Klaus Hartmann, ZEISS - Center for Sensor Systems, University of Siegen, Germany

GMM

Адаптация гауссиан с течением времени



- Сравнение происходит путем вычисления расстояния Махаланобиса на связанных гауссианах $N_{\underline{x}}(\underline{\mu}_i, \Sigma_i)$

$$\delta(\underline{x}, \underline{\mu}_i) = \sqrt{(\underline{\mu}_i - \underline{x})^T \Sigma_i^{-1} (\underline{\mu}_i - \underline{x})} < T_{near}$$

T_{near} - константа

- Если похожие наблюдения уже были записаны, их гауссианы адаптируются с использованием данных нового наблюдения
- В противном случае, новая гауссиана создается и добавляется к смеси

GMM

Реализация

- Наблюдение для пикселя задается как $\underline{s}(\underline{x}) = (y, c_b, c_r, z, a)^T$
- Разрешение глубины и амплитудной модуляции меняется в соответствии с разрешением цветного изображения, по методу ближайшего соседа
- Дисперсии всех гауссиан сводятся к диагональному виду

GMM

Обработка значений глубины



Карта глубины

Значения глубины считаются неправильными, если соответствующая амплитуда ниже заданного порога

GMM

Обработка значений глубины

В таких случаях вводится мера

$$\widehat{\delta}(\underline{x}, \underline{\mu}_i)^2 = (\underline{\mu}_i - \underline{x})^T \Sigma_i^{-1} \begin{pmatrix} 1 & & & \\ & \lambda_c & & \\ & & \lambda_c & \\ & & & \lambda_z \\ & & & & \lambda_a \end{pmatrix} (\underline{\mu}_i - \underline{x})$$

и проверяющее условие

$$\begin{aligned} \widehat{\delta}(\underline{x}, \underline{\mu}_i)^2 &< T_{near}^2 \cdot Tr \begin{pmatrix} 1 & & & \\ & \lambda_c & & \\ & & \lambda_c & \\ & & & \lambda_z \\ & & & & \lambda_a \end{pmatrix} \\ &= T_{near}^2 (1 + 2\lambda_c + \lambda_z + \lambda_a) \end{aligned}$$

где $\lambda_z \in \{0, 1\}$, $\lambda_c \in \{0, 1\}$, $\lambda_a \in \{0, 1\}$

GMM

Связь между измерениями

- Сопоставляющая функция учитывает факт, что наблюдения в трех измерениях практически являются зависимыми
- Объект переднего плана отличается не только глубиной, но и цветом, и свойствами отражения
- Слабое отражение объекта имеет влияние на измерения глубины
- Низкий уровень освещенности ухудшает значения цветности

GMM

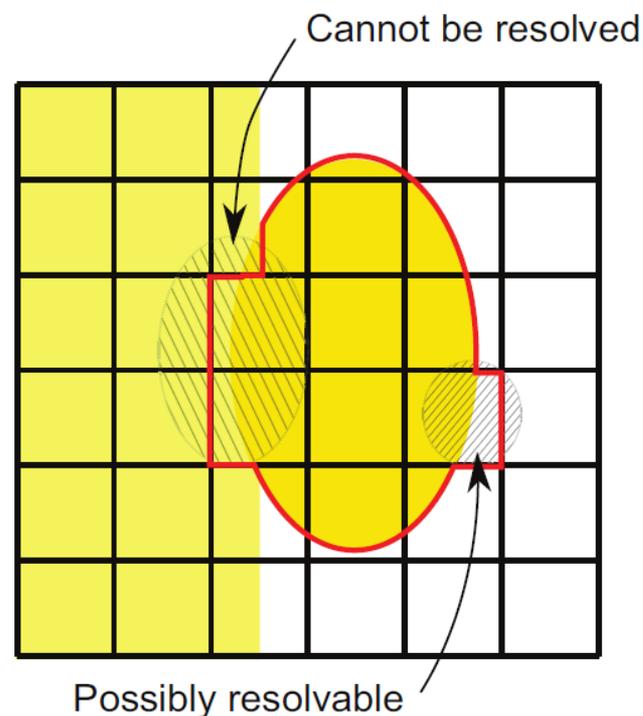
Связь между измерениями

Связь между измерениями снижает:

- уровень шума на маске переднего плана
- количество ошибочных классификаций, связанных с тенями и блочными артефактами

Ограничения

Из-за попиксельной обработки основанное на цвете GMM вычитание фона не может различить фон и передний план при схожести цветов

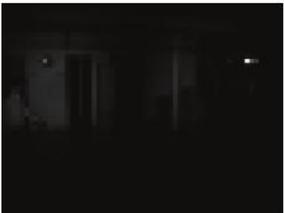
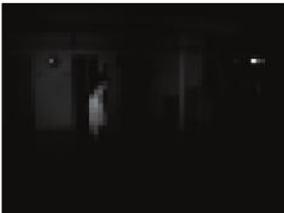
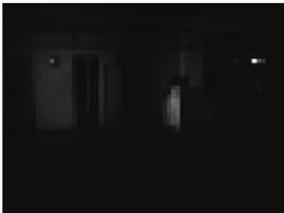
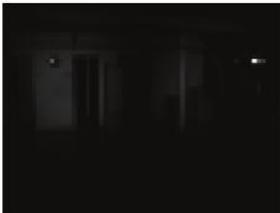


Результаты

Сравниваемый метод	Описание
GMM	Стандартный метод, основанный на цвете (Stauffer and Grimson, 1999)
Original GMMD	Метод, основанный на цвете и глубине (Harville et al., 2001)
GMMD without depth	Метод авторов с параметрами $\lambda_z = 0, \lambda_a = 0$
GMMD	Метод авторов, λ_z определяется зависимости от амплитудной модуляции для каждого пикселя
MMGMM	Метод авторов с параметрами $\lambda_a = 1$
OpenCV GMM	Метод авторов с постобработкой

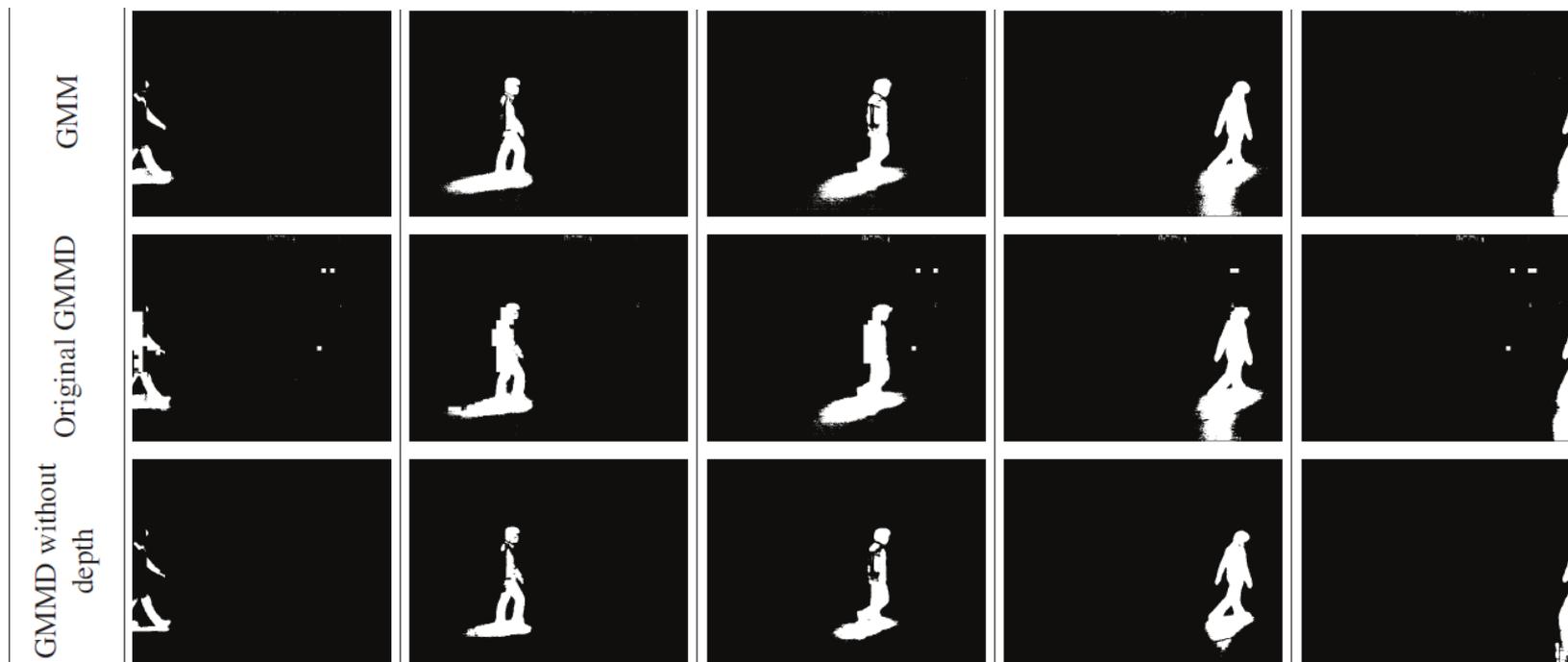
Результаты

Table 1: Walk-by video with difficult lighting conditions

Method	Frame 195	Frame 210	Frame 225	Frame 240	Frame 255
Input					
Depth					
Modulation amplitude					

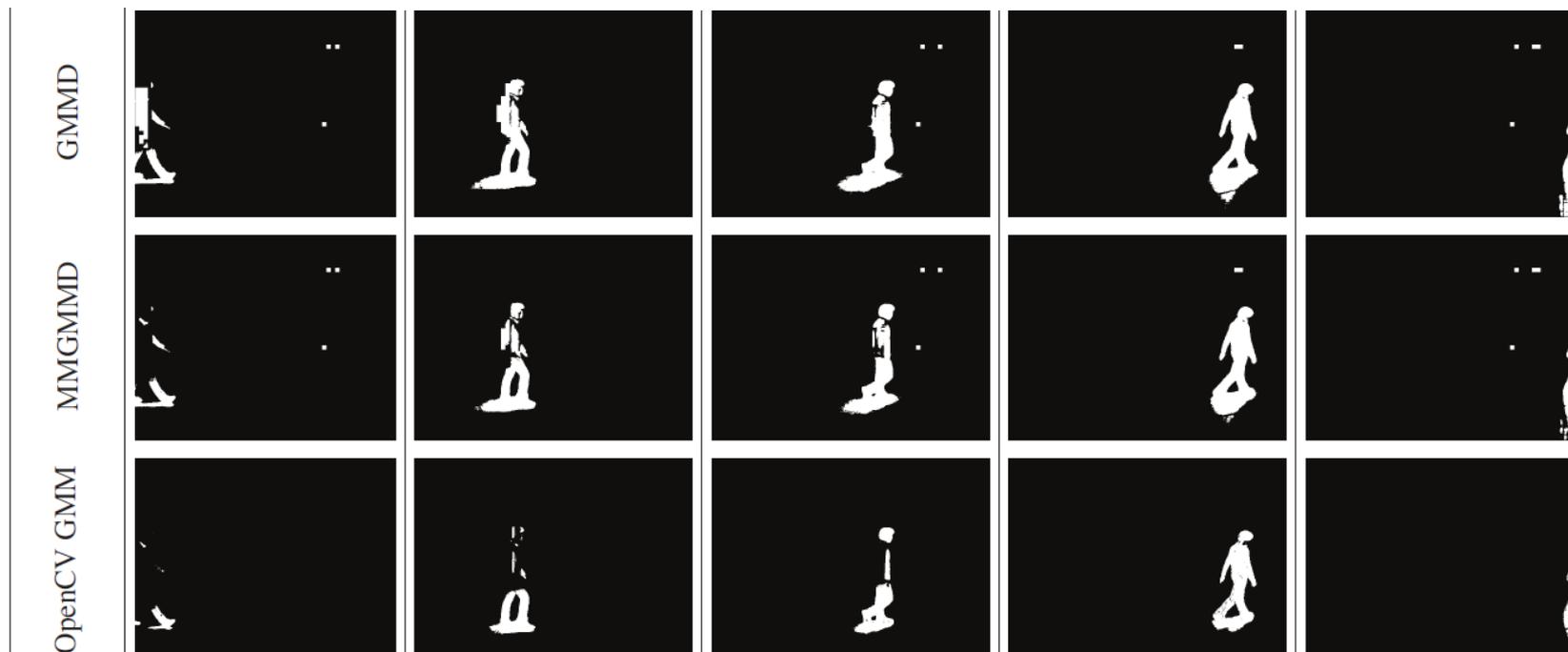
B. Langmann S. E., Ghobadi, K. Hartmann, "Multi-modal background subtraction using Gaussian mixture models", IAPRS, Vol. XXXVIII, Part 3A, 2010

Результаты



B. Langmann S. E., Ghobadi, K. Hartmann, "Multi-modal background subtraction using Gaussian mixture models", IAPRS, Vol. XXXVIII, Part 3A, 2010

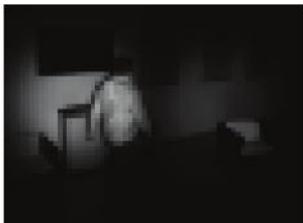
Результаты



B. Langmann S. E., Ghobadi, K. Hartmann, "Multi-modal background subtraction using Gaussian mixture models", IAPRS, Vol. XXXVIII, Part 3A, 2010

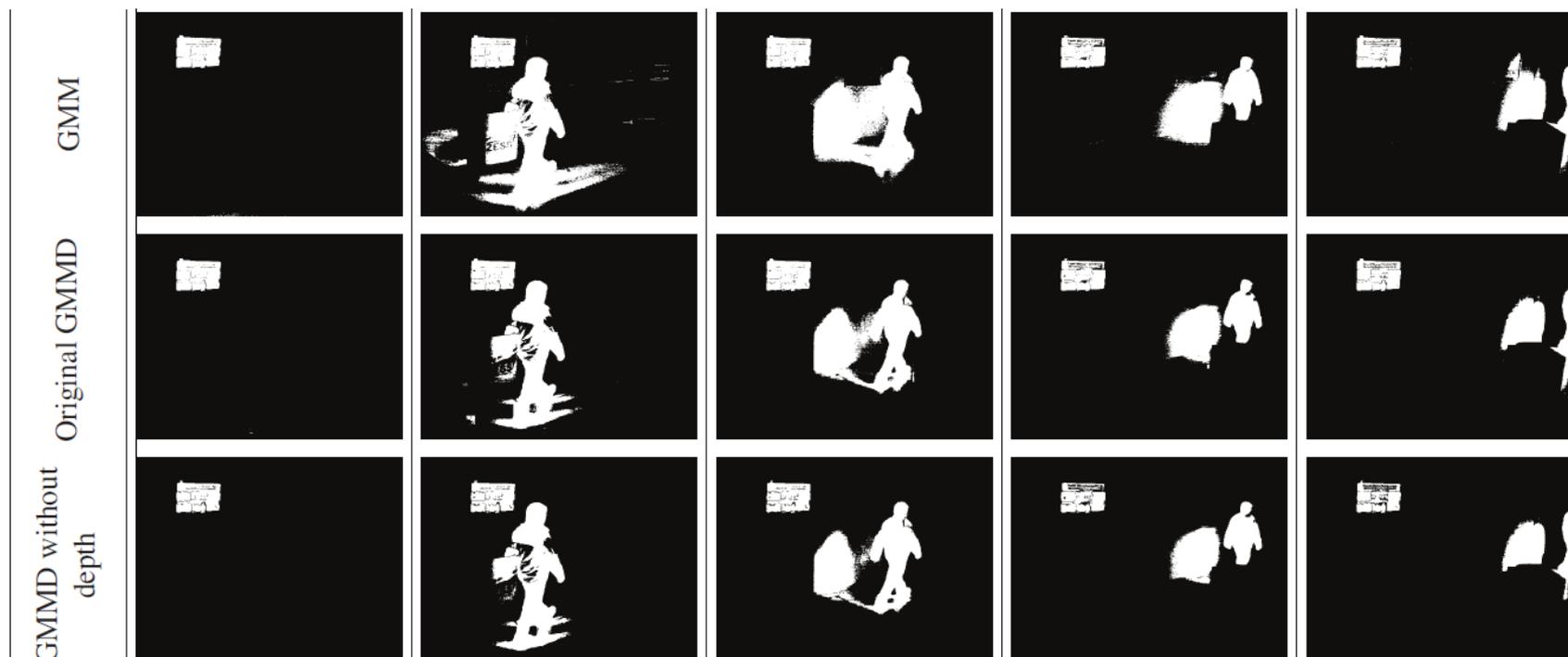
Результаты

Table 2: Simple video with good lighting conditions but difficult foreground

Method	Frame 80	Frame 100	Frame 120	Frame 140	Frame 160
Input					
Depth					
Modulation amplitude					

B. Langmann S. E., Ghobadi, K. Hartmann, "Multi-modal background subtraction using Gaussian mixture models", IAPRS, Vol. XXXVIII, Part 3A, 2010

Результаты



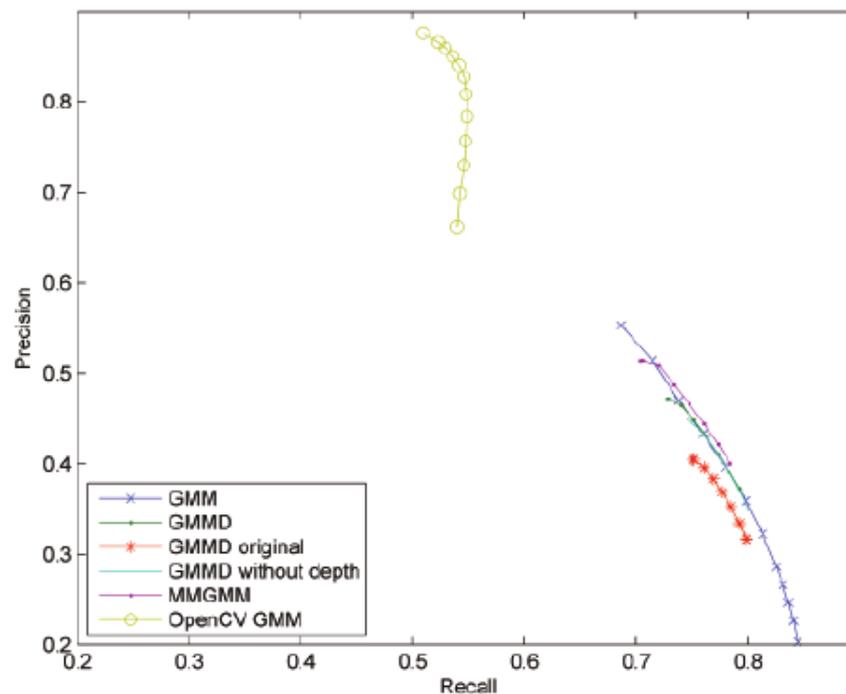
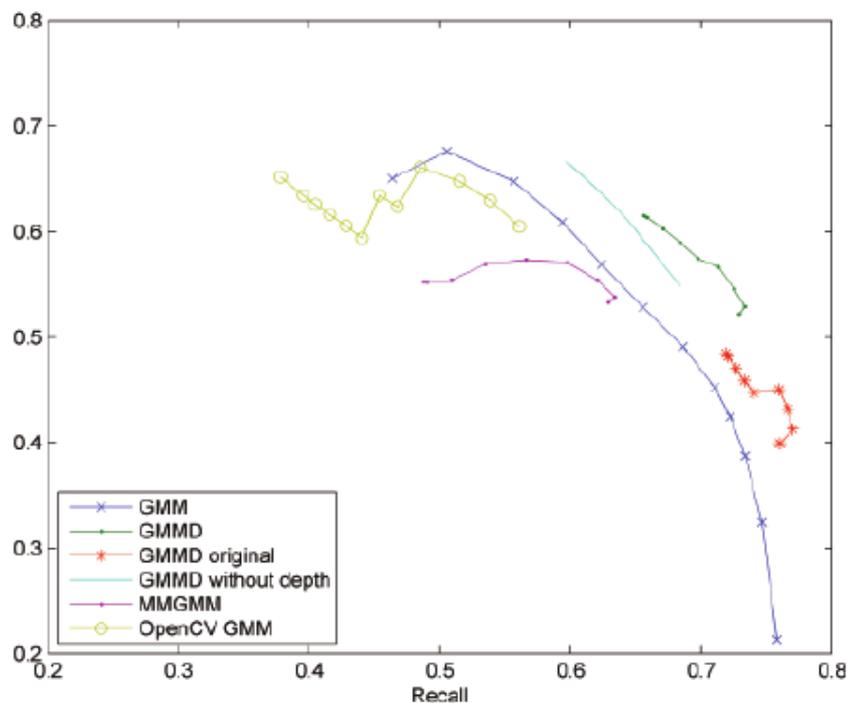
B. Langmann S. E., Ghobadi, K. Hartmann, "Multi-modal background subtraction using Gaussian mixture models", IAPRS, Vol. XXXVIII, Part 3A, 2010

Результаты

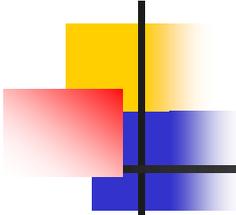


B. Langmann S. E., Ghobadi, K. Hartmann, "Multi-modal background subtraction using Gaussian mixture models", IAPRS, Vol. XXXVIII, Part 3A, 2010

Результаты

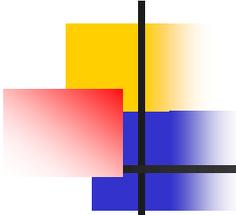


Результаты recall/precision для первого и второго видео



Выводы

- **Достоинства**
 - Снижена ошибочная классификация теней
 - Уменьшены повреждения натуральных границ переднего плана
 - Когда классификация по цвету не срабатывает, эти области заполняются частично
- **Недостатки**
 - Блочность при использовании глубины
 - Проблемы при близких цветах фона и объекта



Содержание

- Multi-modal background subtraction using Gaussian mixture models(GMM)
- **Robust background subtraction for depth generation**

Robust background subtraction for depth generation

Два метода сегментации переднего плана основанные на background subtraction

Цели:

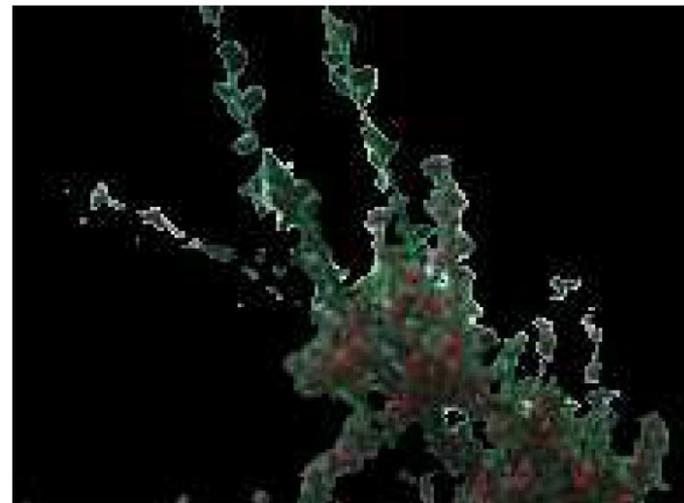
- Уменьшить вмешательство пользователя
- Улучшить качество полученных карт глубин



Original frame

Генерация карт глубин

- При статической камере только движущиеся объекты меняют глубину
- Карта глубины генерируется не для каждого кадра, а только для тех частей, где объекты меняют глубину



Segmented foreground

KLT



KLT Tracker:

- Автоматически выбирает особые точки на изображении
- Оценивает их перемещение между кадрами

Общий метод

- KLT corner tracker распознает глобальное движения камеры
- Для каждого кадра вычисляется преобразование
- Кадр наносится на мозаику фона
- Значение цвета каждого пикселя мозаики фона – медиана по каждому каналу

Общий метод

- Отделение фона от объекта: попиксельное расстояние Махаланобиса от входного кадра с фоновой мозаикой

- Пиксель с цветовыми значениями (R, G, B) считается фоном если

$$((R, G, B) - \mu)^T \Sigma^{-1} ((R, G, B) - \mu) \leq c$$

Σ, μ получены из фоновой мозаики

- Определяется бинарная маска

Общий метод

- Окончательная сегментация получается с использованием морфологических операций с n различными бинарными масками и соответствующих им n различных порогам
- Связанные регионы идентифицируются и прослеживаются во всей последовательности

Ограничения

- Алгоритм построен на следующих предположениях:
 - Фон занимает большую часть кадра
 - Движущиеся объекты отличны от фона
 - Движущиеся объекты статичны в <50% кадров
 - Максимум два объекта перекрываются одновременно
- Алгоритм не рассчитан на изменения в освещении

Выводы

- Достоинства
 - Возможность работы с движущейся камерой
 - Простота
- Недостатки
 - Пользователь должен вручную определить параметр n и пороги c_1, \dots, c_n
 - Проблемы при близких цветах фона и объекта

Метод авторов

- Статическая камера
- Фон определяется, если он встречается во времени хотя бы раз
- Один параметр: фон детализированный или грубый
- Noise-free сегментация

Метод авторов

Функция энергии

- Для заданных $M \times N$ размеров кадра, пикселя p , $l_p \in \{0, 1\}$ определим сегментацию как:

$$\{p_1 = l_{p_1}, \dots, p_{MN} = l_{p_{MN}}\}$$

- Для хорошей сегментации необходимы два свойства:
 - E1: чем более вероятно p описывает передний план, тем больше вероятность $l_p = 1$
 - E2: низкий "средний локальный контраст"

Метод авторов

Функция энергии

- Для каждого пикселя определяется

$$P_{FG}(p) := e^{-T \|I_p - BG_p\|^2}$$

где I_p – значение цвета пикселя в кадре

BG_p – значение цвета пикселя фона

T – константа

- Вероятность фона в пикселе

$$P_{BG}(p) := 1 - P_{FG}(p)$$

Метод авторов

Функция энергии

- Если $\frac{P_{FG}(p)}{P_{BG}(p)} < 1$, то $\tau_p := \log \frac{P_{FG}(p)}{P_{BG}(p)} < 0$
- Если p – передний план, то $\tau_p > 0$
- Определим свойства:

$$E_1(l_p) := \begin{cases} 0, & l_p = 1, \tau_p > 0 \\ -\tau_p, & l_p = 1, \tau_p \leq 0 \\ \tau_p, & l_p = 0, \tau_p > 0 \\ 0, & l_p = 0, \tau_p \leq 0 \end{cases}$$

$$E_2(p, q) := \begin{cases} p(1 - q) + q(1 - p), & (p, q) \in N_p \text{ где } p, q \text{ соседние} \\ 0, & \text{otherwise} \end{cases}$$

Метод авторов

Функция энергии

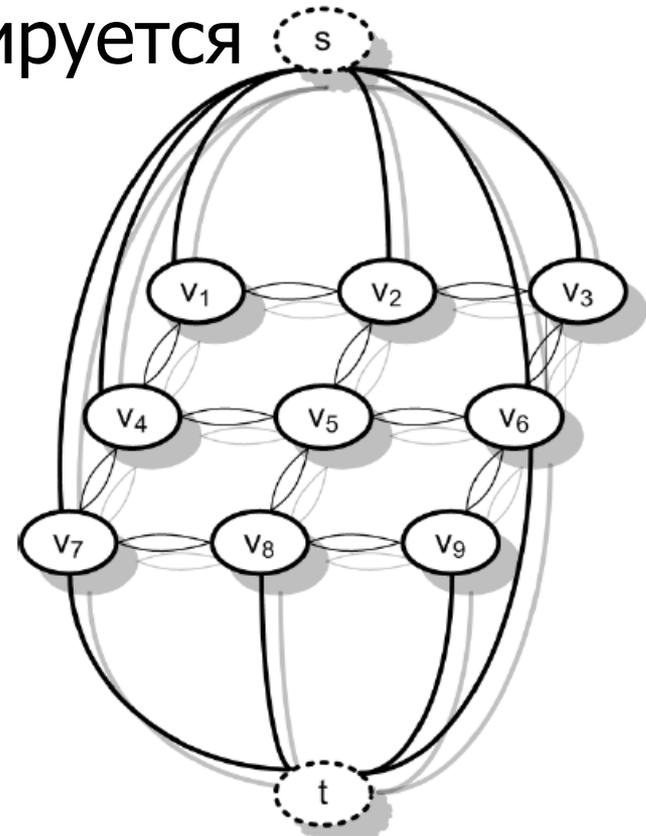
- Для каждой сегментации \mathcal{M} определим функцию энергии

$$\mathcal{E}(\mathcal{M}) := \sum_{p \in \mathcal{S}} E_1(l_p) + \lambda \sum_{p, q \in \mathcal{S}} E_2(l_p, l_q)$$

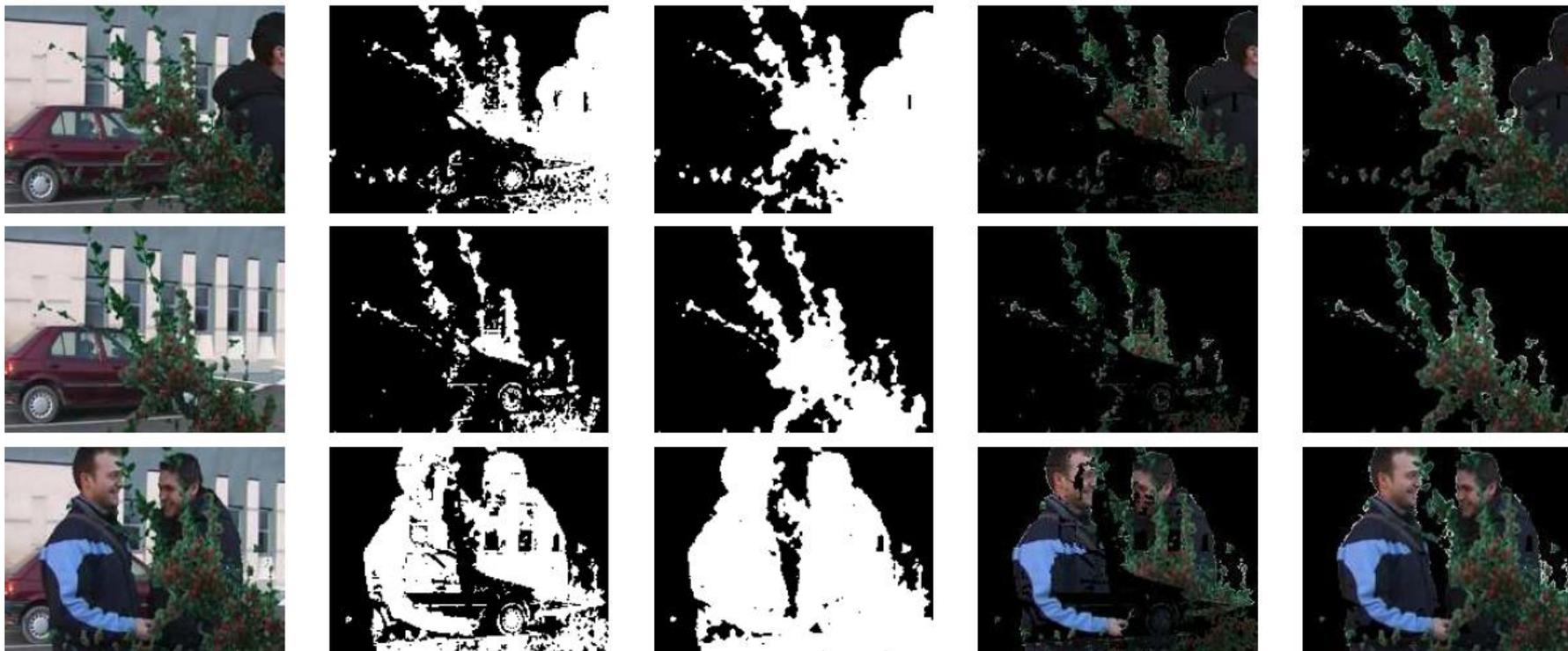
- $\mathcal{S} := \{(x, y) : x \in \{1, \dots, M\}, y \in \{1, \dots, N\}\}$
- λ показывает зависимость между свойствами E_1 и E_2

Минимизация функции энергии

- Функция энергии минимизируется с помощью Graph cut
- Для поиска минимального разреза используется maximum-flow



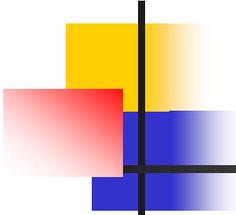
Результаты



M. Schmeing, X. Jiang. Robust Background Subtraction for Depth Map Generation, Proceedings of 3D Stereo MEDIA 2009

Выводы

- Достоинства
 - Меньше проблем при близких цветах фона и объекта
 - Фон строится если появляется хотя бы 1 раз
 - Пользователь задает только один параметр λ
 - Сегментация близка в реальном времени
- Недостатки
 - Проблемы при близких цветах фона и объекта остаются



Дальнейшая работа

- При работе с видео с Kinect можно использовать сегментацию методом вычитания фона
- Методы из статей могут оказаться полезными

Литература

1. "Multi-modal background subtraction using gaussian mixture models" , Benjamin Langmann, Seyed E. Ghobadi, Klaus Hartmann, ZESS - Center for Sensor Systems, University of Siegen, Germany
2. "Robust background subtraction for depth map generation", Michael Schmeing and Xiaoyi Jiang, Department of Computer Science, University of M.unster, 48149 M.unster, Germany
3. S. Rothaus, K. Rothaus, and X. Jiang. Synthesizing 3D, videos by a motion-conditioned background mosaic. In International Conference on Pattern Recognition, pages 1–4. IEEE, 2008

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