

Способы построения saliency map

Максим Харенко

Video Group
CS MSU Graphics & Media Lab



Содержание

- **Введение**
- **Методы построения**
 - Локальные
 - Глобальные
 - Правила визуального восприятия сцены
- **Оценка качества**

Введение

Что привлекает человеческое внимание?



Введение

Салиентность

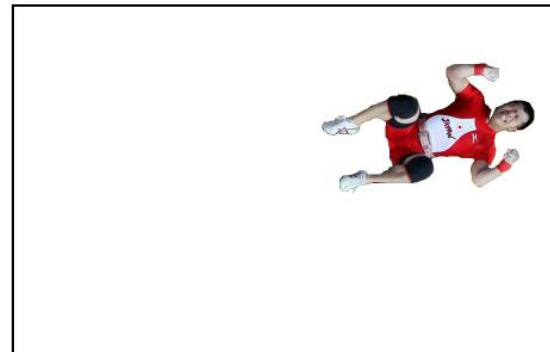
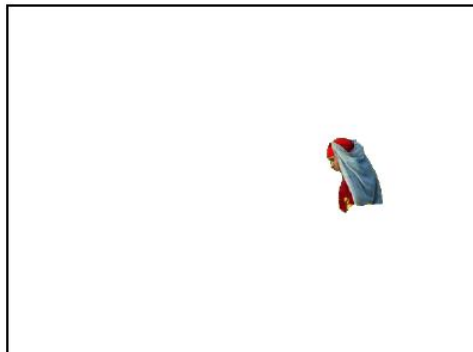
Descriptions

*happy girl
smiling kid
cute girl*

*man in flower field
in the fields
spring blossom*

*Olympic weight lifter
Olympic victory
Olympic achievement*

Salient object

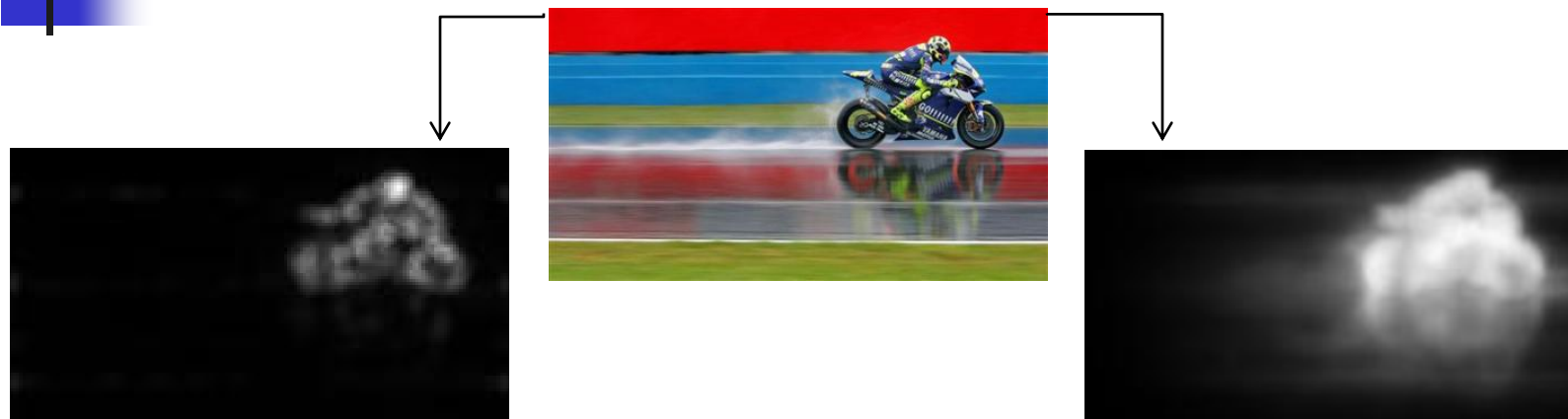


Saliency



Введение

Применение карт салиентности



- Сегментация
- Распознавание объектов
- Удаление объектов из видео

- Измерение качества 3D видео
- Создание коллажей
- Image thumbnailing
- Image and video retargeting

Введение

Методы построения карт салиентности

Методы моделирования человеческого визуального внимания:

- Временные
- Локальные
- Глобальные
- Основанные на правилах организации сцены
- Учет высокоуровневых факторов



Содержание

- Введение
- Методы построения
 - **Временные**
 - Локальные
 - Глобальные
 - Основанные на правилах организации сцены
- Оценка качества

Временные методы

Временная карта saliентности(1)

- Для каждого пикселя кадра считается optical flow и строится поле векторов $\vec{V}_\theta(s_i)$
- Оценивается global motion $\vec{V}_G(s_i)$
- Остаточное движение считается как:

$$\vec{V}_R(s_i) = \vec{V}_\theta(s_i) - \vec{V}_G(s_i)$$



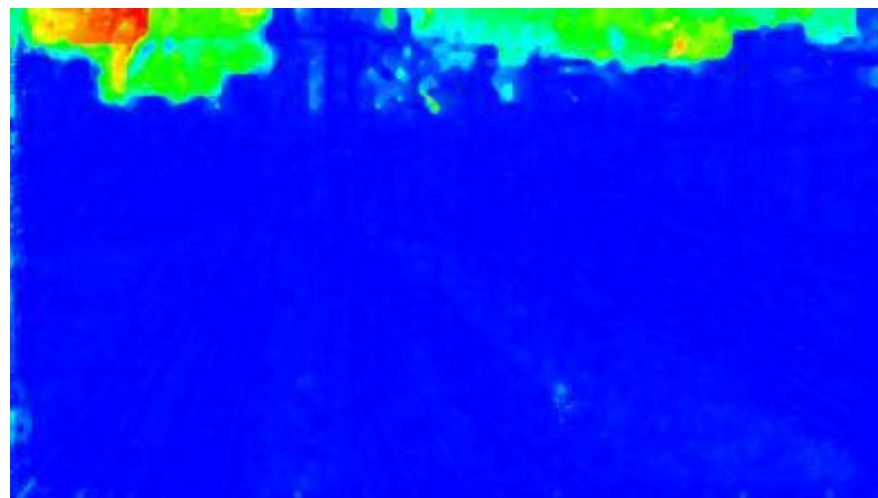
Исходный кадр

Временные методы

Временная карта saliентности(2)

Психо-визуальные ограничения:

- Человеческий глаз не успевает следить за движением скоростью которого превышает 80 deg./s
- Визуальная saliентность достигает максимума, когда значение движения находятся между 6 deg./s и 30 deg./s



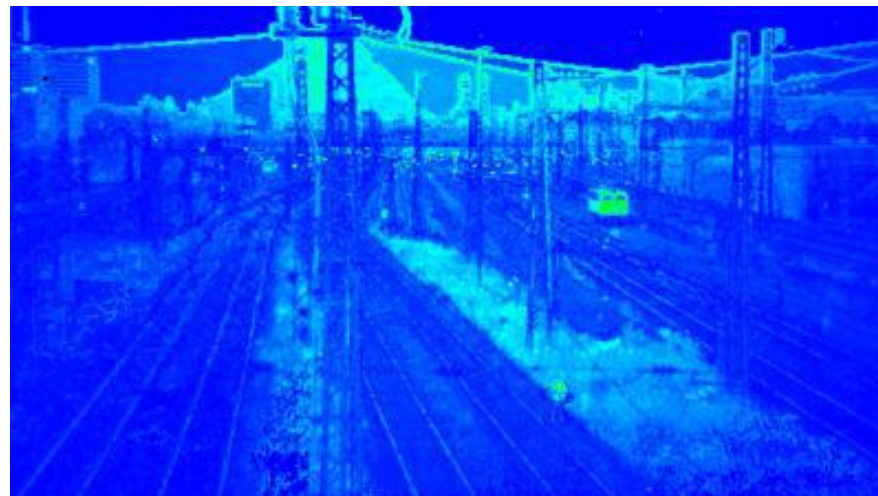
Временная карта saliентности
до фильтрации

Временные методы

Временная карта салиентности(3)

Фильтрация:

$$S^T(s_i) = \begin{cases} \frac{1}{7}\vec{V}_R(s_i) & \text{if } 0 \leq \vec{V}_R(s_i) < \vec{v}_1 \\ 1 & \text{if } \vec{v}_1 \leq \vec{V}_R(s_i) < \vec{v}_2 \\ \frac{1}{60}\vec{V}_R(s_i) + \frac{8}{5} & \text{if } \vec{v}_2 \leq \vec{V}_R(s_i) < \vec{v}_{max} \\ 0 & \text{if } \vec{V}_R(s_i) \geq \vec{v}_{max} \end{cases}$$



Временная карта салиентности
после фильтрации

$$\vec{v}_1 = 6 \text{ deg./s}, \vec{v}_2 = 30 \text{ deg./s}, \vec{v}_{max} = 80 \text{ deg./s}$$



Содержание

- Введение
- Методы построения
 - Временные
 - **Локальные**
 - Глобальные
 - Правила визуального восприятия сцены
- Оценка качества

Локальные методы

Пространственно-временной алгоритм

- Для каждого кадра строится пространственная карта салиентности
- Строится временная карта салиентности
- Путем их объединения получается финальная карта салиентности



Исходный кадр

Локальные методы

Пространственная карта saliентности

Суммируются 7 характеристик V_δ :

- контраст насыщенности
- контраст яркости
- контраст оттенка
- контраст противоположных цветов
- контраст теплых и холодных цветов
- доминирование теплых цветов
- доминирование яркости и оттенка

$$S^{SP}(s_i) = \frac{1}{7} \sum_{\delta=1}^7 V_\delta(s_i)$$



Пространственная карта saliентности

Локальные методы

Карта saliентности

Перемножив пространственную и временную карту saliентности с весами 2D Gauss, получаем финальную карту saliентности



Карта saliентности

$$S_{mul}^{SP-T}(s_i) = S^T(s_i) \times S^{SP}(s_i) \times 2DGauss(s)$$

$$g(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$



Содержание

- Введение
- Методы построения
 - Временные
 - Локальные
 - **Глобальные**
 - Правила визуального восприятия сцены
- Оценка качества

Глобальные методы

Основные принципы

- Используются соображения, основанные на частотно-фазовых характеристиках изображения
- Часто встречающиеся особенности подавляются
- Важные места определяются как отклонение от нормы

Глобальные методы

Логарифмический спектр

$$\mathcal{A}(f) = \Re\left(\mathfrak{F}[\mathcal{I}(x)]\right),$$

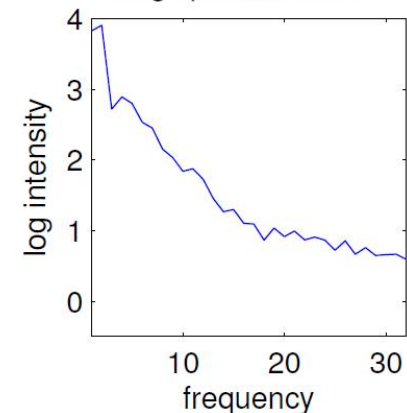
$$\mathcal{P}(f) = \Im\left(\mathfrak{F}[\mathcal{I}(x)]\right),$$

$$\mathcal{L}(f) = \log(\mathcal{A}(f))$$

Input image



Log spectrum curve



$\mathfrak{F}[\mathcal{I}(x)]$ – преобразование Фурье

$\Re\left(\mathfrak{F}[\mathcal{I}(x)]\right)$ – спектр амплитуд

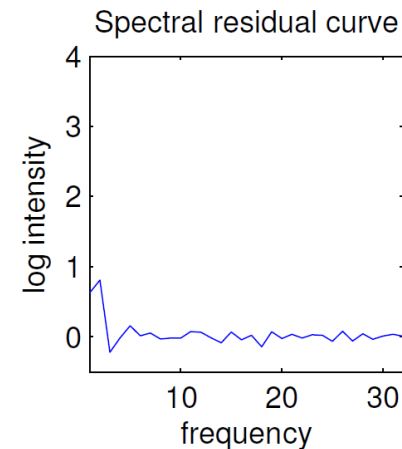
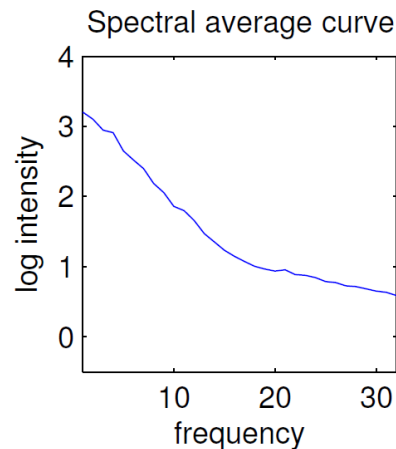
$\Im\left(\mathfrak{F}[\mathcal{I}(x)]\right)$ – спектр фаз

Глобальные методы

Остаточный спектр

$$\mathcal{A}(f) = h_n(f) * \mathcal{L}(f),$$

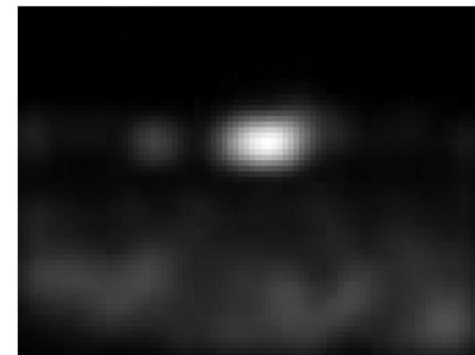
$$h_n(f) = \frac{1}{n^2} \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix}$$



$$\mathcal{R}(f) = \mathcal{L}(f) - h_n(f) * \mathcal{L}(f),$$

$$\mathcal{S}(x) = g(x) * \mathfrak{F}^{-1} \left[\exp(\mathcal{R}(f) + \mathcal{P}(f)) \right]^2$$

Saliency map



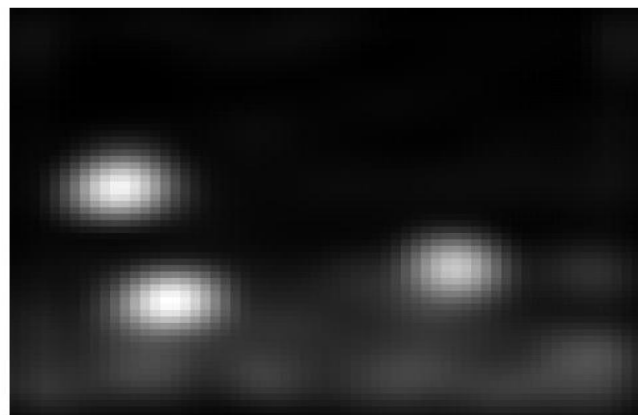
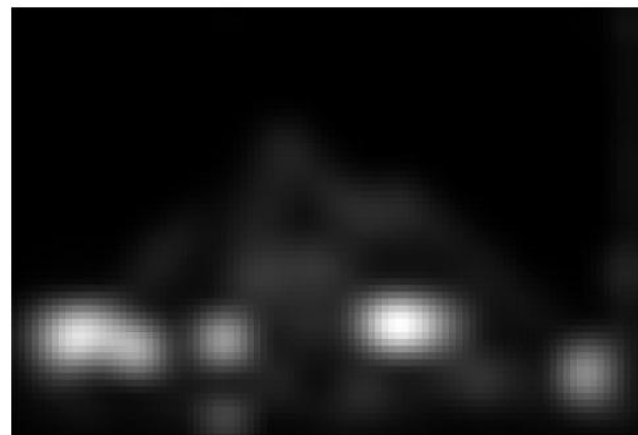
Глобальные методы

SR: Результаты

Input image



Saliency map

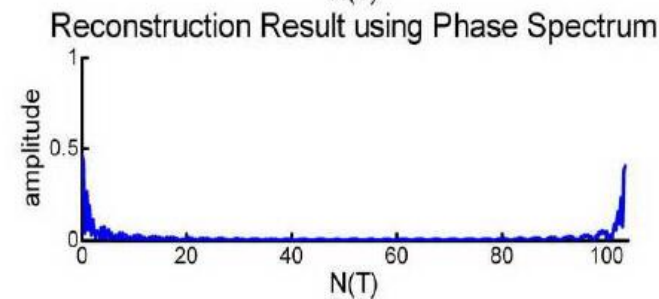
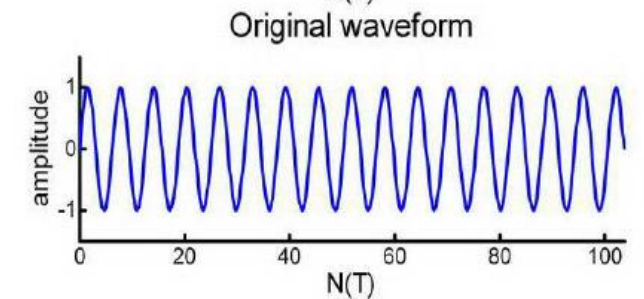
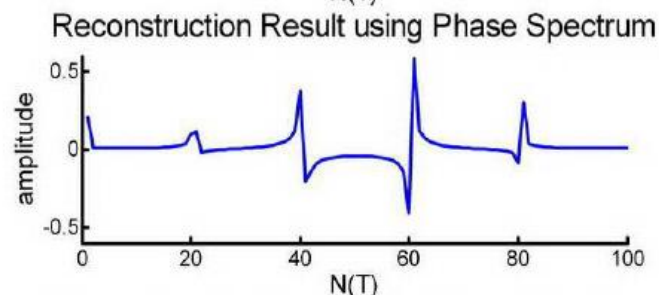
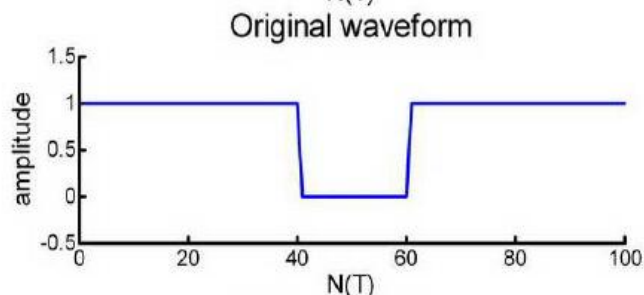
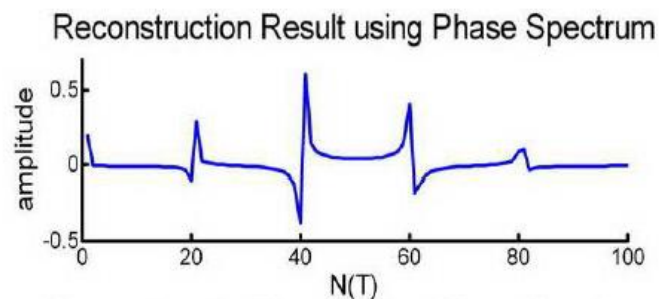
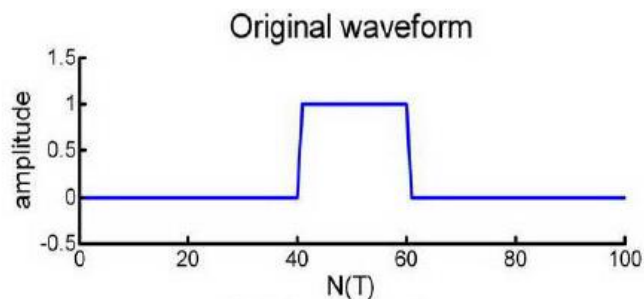


Глобальные методы РФТ и PQFT

- Использует только фазовые характеристики
- Работает быстрее SR
- Имеет временную реализация

Глобальные методы

Реконструкция сигнала по спектру фаз



C. Guo, Q. Ma, and L. Zhan, "Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform," CVPR, 2008

Глобальные методы

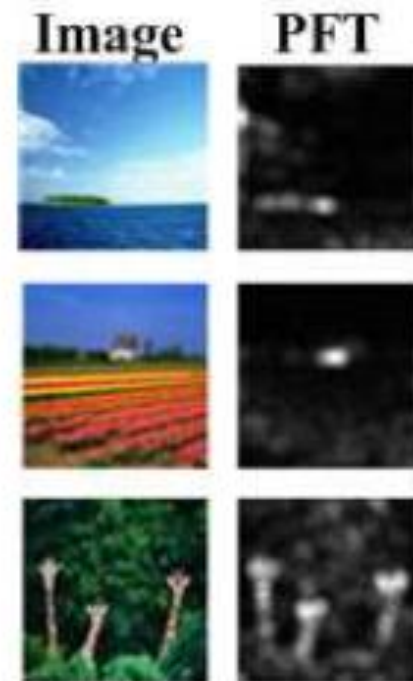
Построение saliency map

$$f(x, y) = F(I(x, y))$$

$$p(x, y) = P(f(x, y))$$

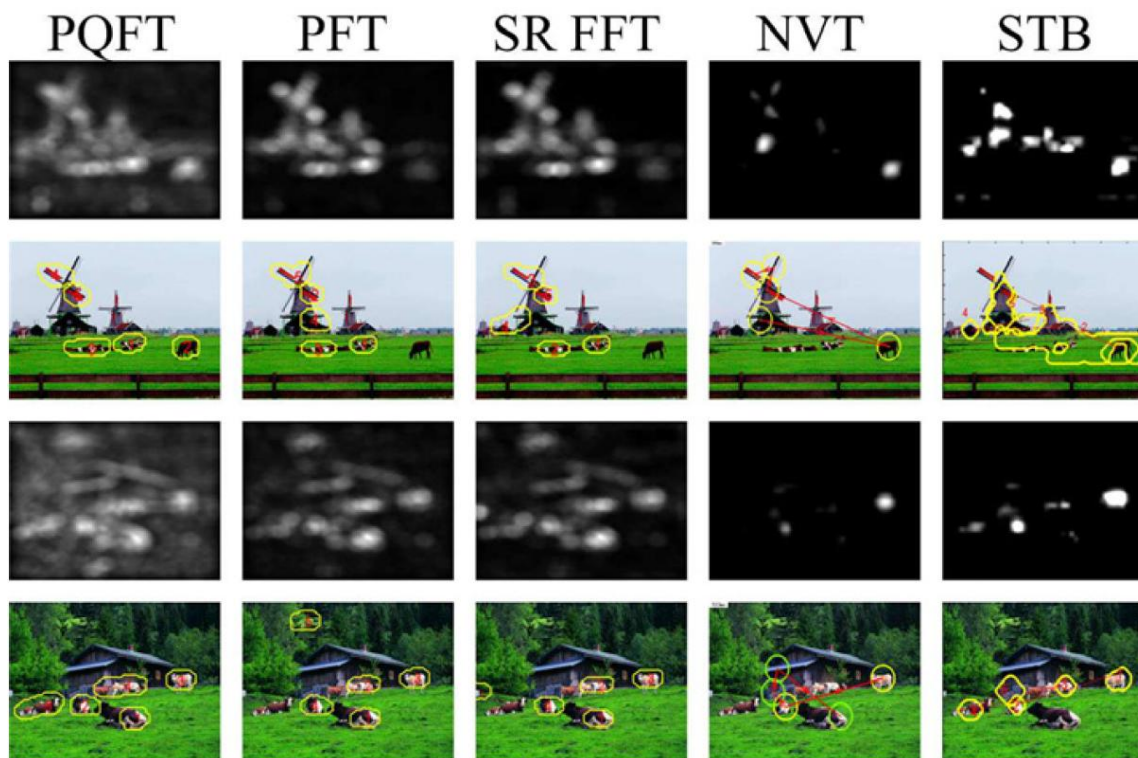
$$sM(x, y) = g(x, y) * \|F^{-1}[e^{i \cdot p(x, y)}]\|^2$$

$$g(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$



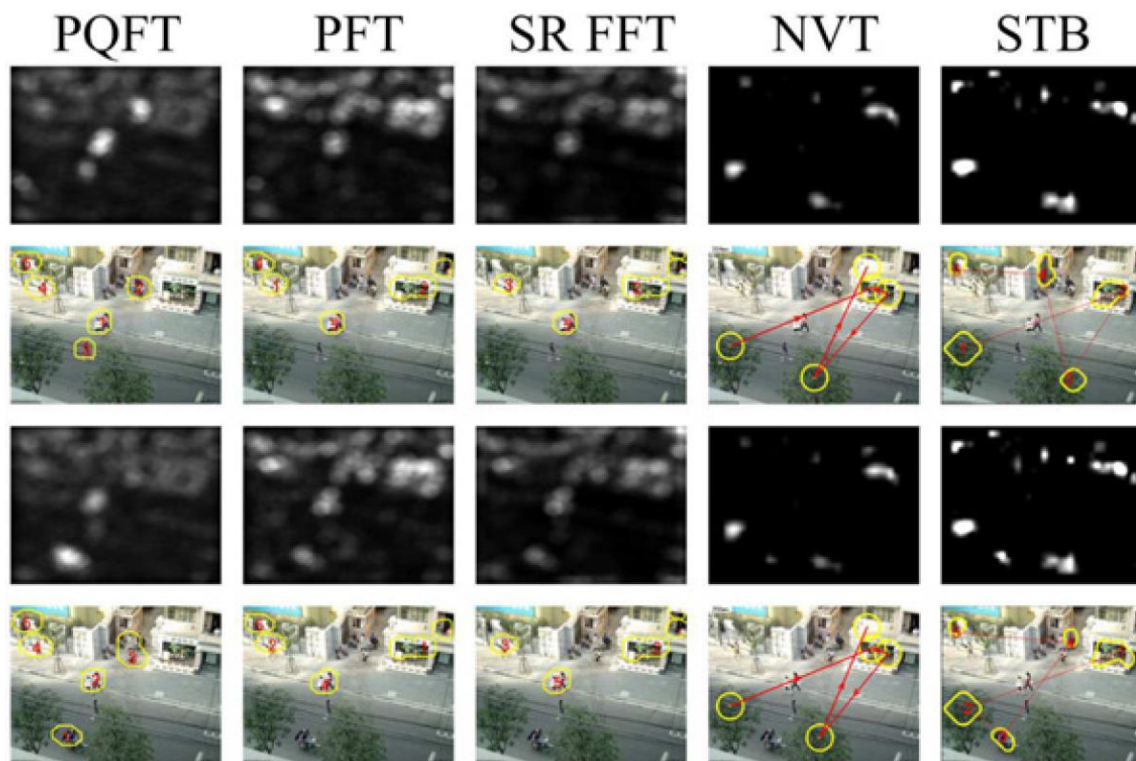
Глобальные методы

PFT: результаты на изображениях



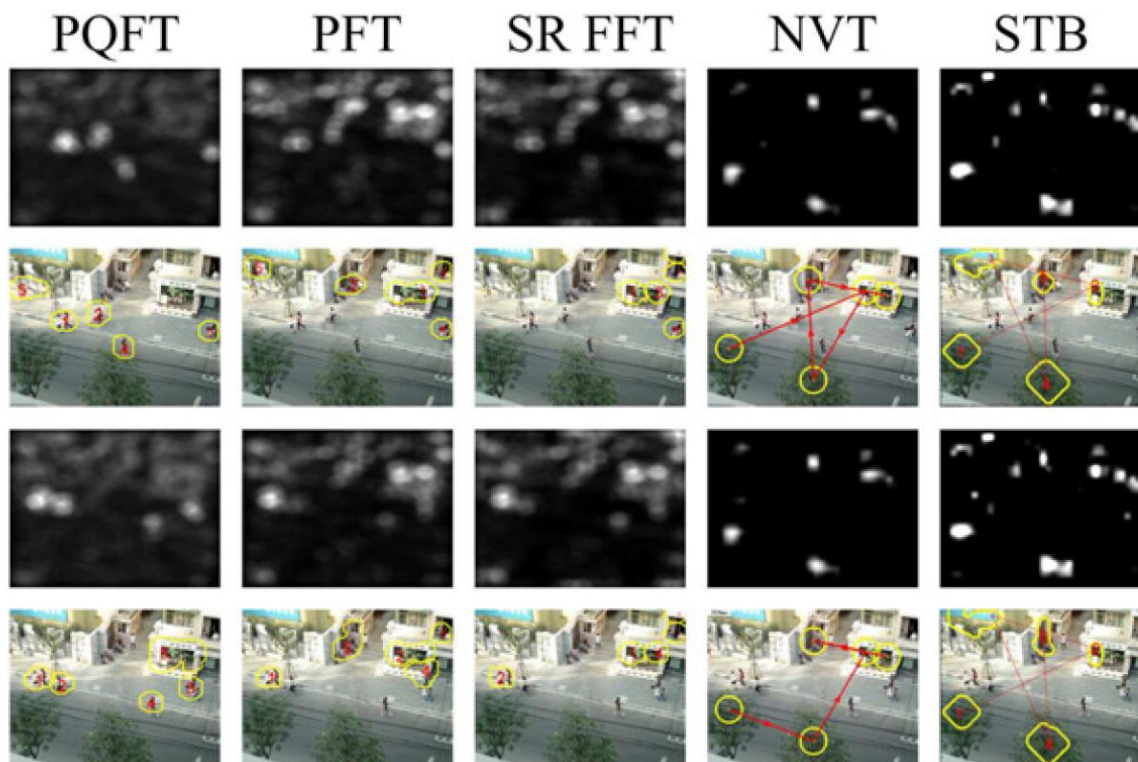
Глобальные методы

PFT: результаты на видео (1)



Глобальные методы

PFT: результаты на видео (2)



C. Guo, Q. Ma, and L. Zhan, "Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform," CVPR, 2008

Глобальные методы

Сравнение (1)

Model	1st	2nd	3rd	4th	5th	ANODF
PQFT	921	704	405	274	182	2.52
PFT	132	349	283	260	262	1.30
SR	142	208	229	234	218	1.04
NVT	24	56	70	91	75	0.32
STB	138	58	58	58	63	0.38

Table 2. The number of correct objects detected at each fixation in the test video. Note that ANODF represents the Average Number of Object Detected *per* Frame.

Глобальные методы

Сравнение (2)

Model	Average Time Cost (s)
PQFT	0.0565
PFT	0.0106
SR	0.0141
NVT	0.4313
STB	3.5337

Table 3. Average time cost *per* frame in the test video



Содержание

- Введение
- Методы построения
 - Временные
 - Локальные
 - Глобальные
 - **Основанные на правилах организации сцены**
- Оценка качества saliency map

Визуальное восприятие

Основные принципы



- Важны не только объекты, но и часть их окружения, передающая суть изображения
- Салиентные пиксели должны быть сгруппированы вместе, а не разбросаны по всему изображению
- Не салиентные области могут находиться и рядом и далеко друг от друга

Визуальное восприятие

Local-global single-scale saliency(1)



- Рассматривается патч размером r вокруг каждого пикселя i
- Пиксель i – салиентен, если непохожесть его патча ко всем другим патчам изображения высока, т.е. евклидово расстояние по цвету $d_{color}(p_i, p_j)$ велико $\forall j$
- Патч p_i салиентен, если похожие патчи располагаются близко и менее салиентен, если похожие патчи далеко
- $d_{position}(p_i, p_j)$ – евклидово расстояние между положениями p_i и p_j

Визуальное восприятие

Local-global single-scale saliency(2)

- Тогда мера похожести между парой патчей:

$$d(p_i, p_j) = \frac{d_{color}(p_i, p_j)}{1 + c \cdot d_{position}(p_i, p_j)} \quad c = 3$$

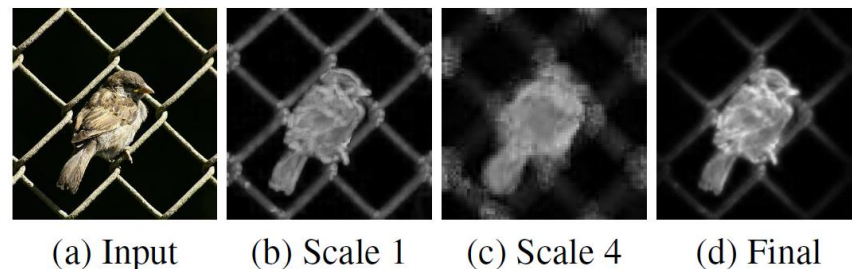
- Для каждого патча p_i в изображении проводится поиск K самых похожих на него патчей $\{q_k\}_{k=1}^K$
- Значение saliентности пикселя i при размере r :

$$S_i^r = 1 - \exp\left\{-\frac{1}{K} \sum_{k=1}^K d(p_i^r, q_k^r)\right\}$$

Визуальное восприятие

Multi-scale saliency

- При однородном или размытом фоне удобно использовать патчи разных размеров
- Патч p_i размера r сравнивается с патчами-кандидатами со всего изображения размером



$$R_q = \left\{ r, \frac{1}{2}r, \frac{1}{4}r \right\}$$

$$S_i^r = 1 - \exp\left\{ -\frac{1}{K} \sum_{k=1}^K d(p_i^r, q_k^{r_k}) \right\} \quad \bar{S}_i = \frac{1}{M} \sum_{r \in R} S_i^r$$

Сравнение результатов: Одиночный объект на простом фоне



input



local



global



context-aware

Визуальное восприятие

Добавление ближайшего фона

Близкие к фокусу внимания области салиентны:

- Из полученной карты салиентности извлекаются самые «привлекательные» области
- Пиксель «привлекателен», если $\bar{S}_i > 0.8$
- Остальные пиксели взвешиваются с $d_{foci}(i)$
- Салиентность пересчитывается:

$$\hat{S}_i = \bar{S}_i(1 - d_{foci}(i))$$

$d_{foci}(i)$ – евклидово расстояние между позициями пикселей i и ближайшего «привлекательного»

Сравнение результатов

Области вокруг фокуса внимания тоже салиентны



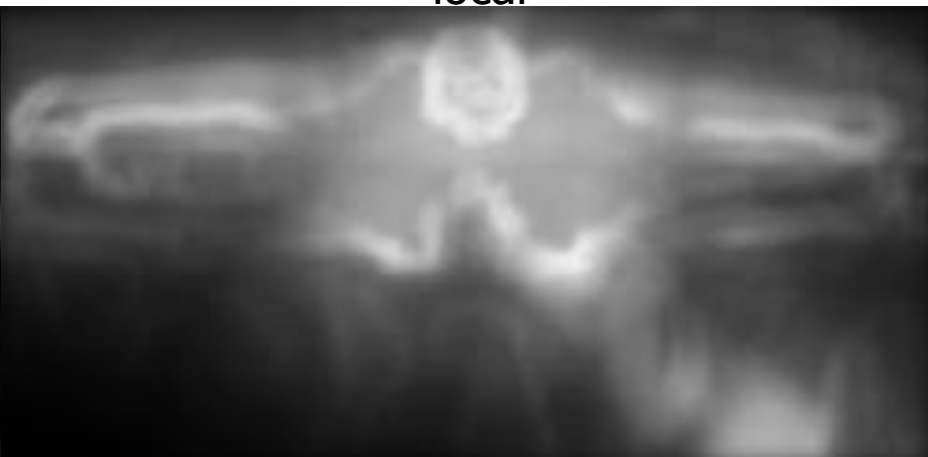
input



local



global



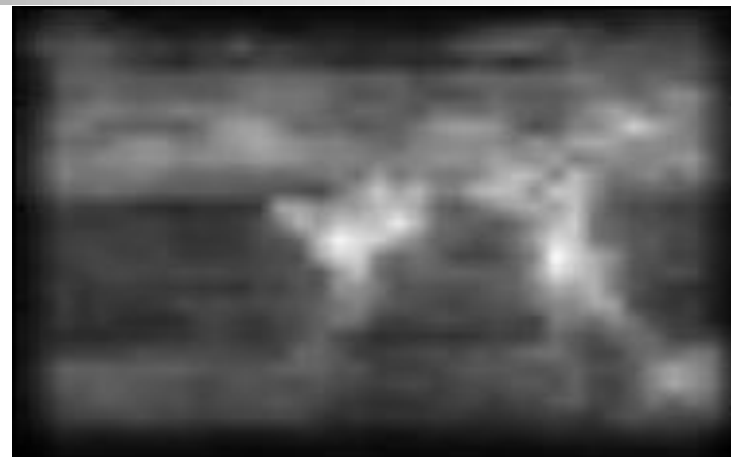
context-aware

Сравнение результатов

Сложные сцены (1)



input



local



global



context-aware

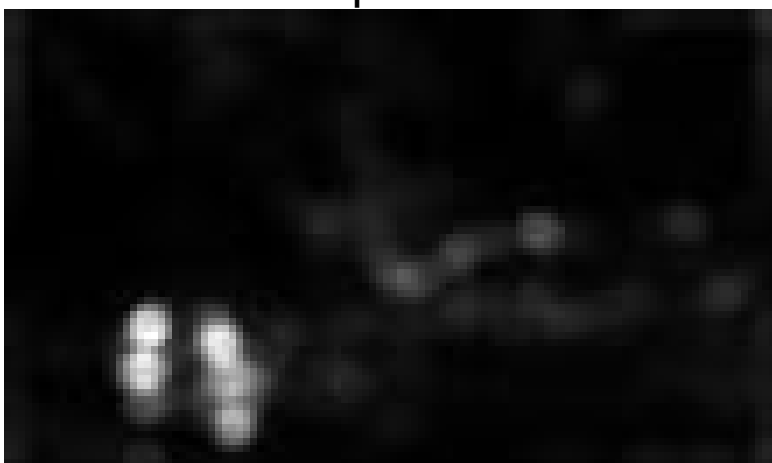
Сравнение результатов Сложные сцены (2)



input



local



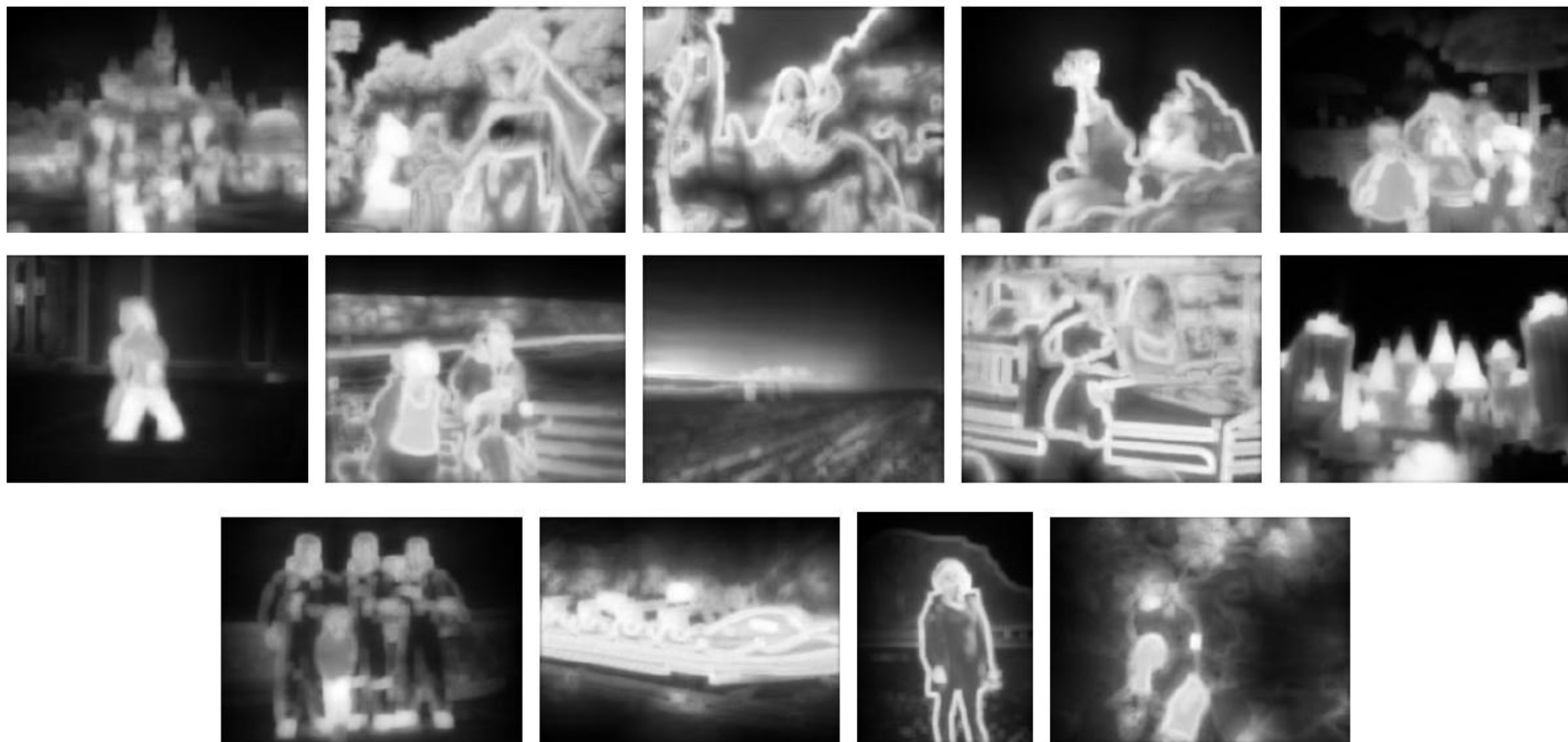
global



context-aware

Визуальное восприятие

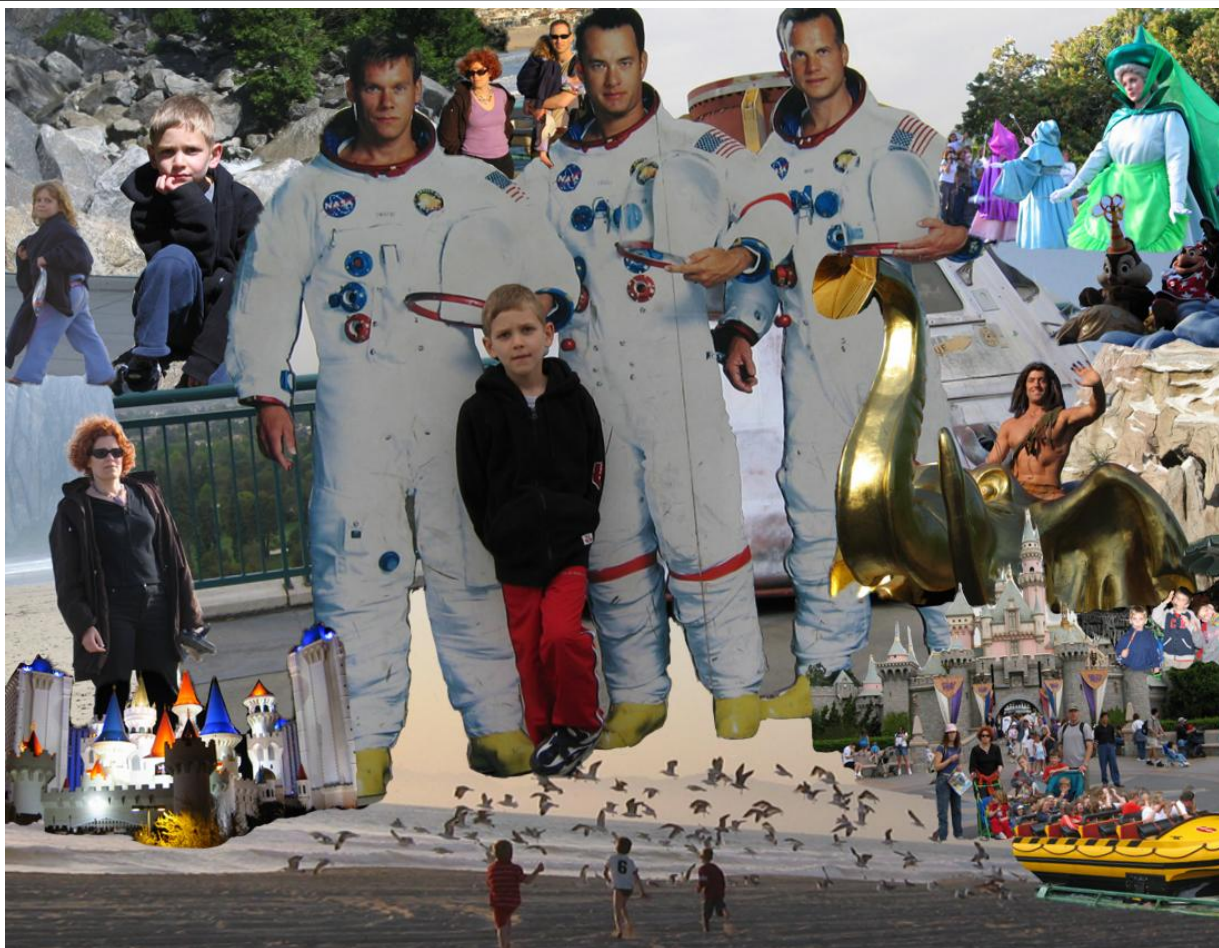
Создание коллажей (1)



(b) The saliency maps of the input images

Визуальное восприятие

Создание коллажей (2)



Суммарный коллаж



Содержание

- Введение
- Методы построения
 - Временные
 - Локальные
 - Глобальные
 - Правила визуального восприятия сцены
- **Оценка качества**

Оценка качества saliency map

Критерии оценки и ground truth



- В задачах обнаружения объектов – количество правильно распознанных объектов
- В задачах сегментации – сравнение границ сегментов
- Для коллажей – красота, корректность и информативность
- Визуальное сравнение: группу людей просят показать «важные» регионы изображения

Оценка качества saliency map

Эксперимент



Ход эксперимента: испытуемый смотрит в экран, eye tracker записывает данные

- Движения глаз записываются бинокулярно с частотой 250 Hz с помощью инфракрасного video-based eye tracker
- Монитор: 17 дюймов, 1280x960 pixels

Hani Alerza, Judith A. Redia, and Ingrid Heynderickxa, "Examining the effect of task on viewing behavior in videos using saliency maps," Human Vision and Electronic Imaging XVII, 2012

Оценка качества saliency map

Видеопоследовательности



- Выбраны 25 динамичных сцен из боевиков длительностью по 20 секунд
- Видео: 1280x720 pixels, 25 fps
- Половина тестируемых просто смотрела видео, другая – оценивала качество



Кадр из видео
с битрейтом 1237 bit/s



Кадр из видео
с битрейтом 209 bit/s

Оценка качества saliency map

Построение карты

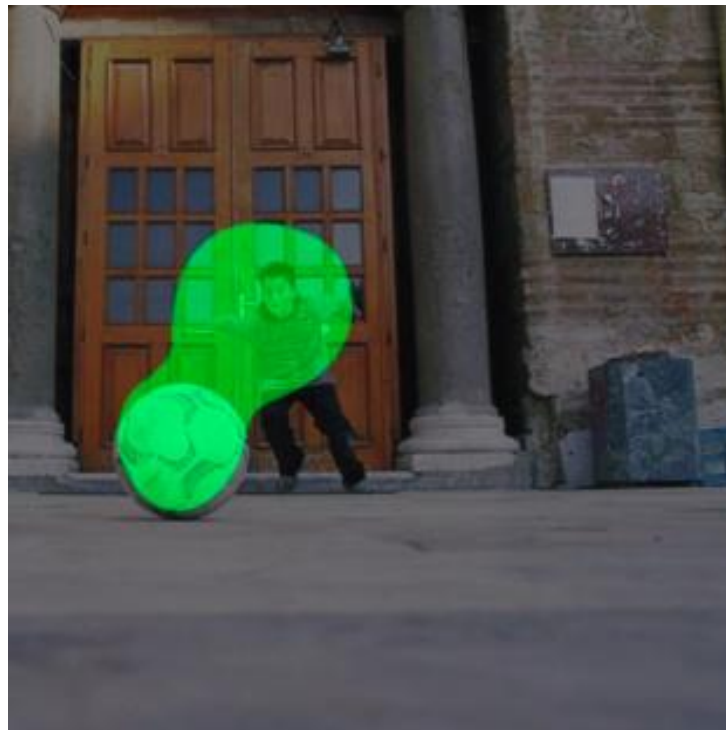
- Eye tracker фиксирует движения глаза, следящего за объектами
- Для каждой секунды каждого видео создается общая карта фиксаций, усредняющая информацию от всех испытуемых
- Значение каждой координаты – общая продолжительность фиксации глаз всех испытуемых на этой координате
- 20 карт фиксаций на видео
- Карта фиксаций трансформируются в карту saliency map:

$$S_{t,i}(k,l) = \sum_{j=1}^T \exp \left[-\frac{(x_j - k)^2 + (y_j - l)^2}{\sigma^2} \right]$$

(x_j, y_j) – пространственные координаты j фиксации
 T – общее количество фиксаций
 $S_{t,i}(k,l)$ – saliency map размера $M \times N$
 σ^2 – дисперсия

Оценка качества saliency map

Результаты



Полученные карты салиентности

Оценка качества saliency map

Выводы



- Карты салиентности, построенные по видео низкого качества, схожи у большинства испытуемых: люди в первую очередь обращают внимание на артефакты
- У групп людей с разным характером просмотра получаются разные карты салиентности
- Результаты свободного просмотра видео высокого качества больше всего приближены к законам салиентности

Литература

1. H. Boujut, O. Hadar, J. Benois-Pineau, T. Ahmed, and P. Bonnet, "Weighted-MSE based on saliency map for assessing video," *Electronic Imaging*, 2011.
2. S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *CVPR*, 2010.
3. X. Hou and L. Zhang, "Saliency detection: A spectral residual Approach", *CVPR*, 2007.
4. C. Guo, Q. Ma, and L. Zhan, "Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform," *CVPR*, 2008.
5. Hani Alersa, Judith A. Redia, and Ingrid Heynderickxa, "Examining the effect of task on viewing behavior in videos using saliency maps," *Human Vision and Electronic Imaging XVII*, 2012.

The image is a movie poster for the 2012 film 'Titanic'. It features a large, detailed view of the ship's hull and upper decks, set against a dramatic, sunset-colored sky. In the foreground, a woman's hands are shown holding a man's hands, symbolizing the central romance. The ship is shown sailing on a dark, choppy sea.

ВОПРОСЫ?

....Или сразу в кино?

ЛЕОНАРДО
ДИКАПРИО

КЕЙТ
УИНСЛЕТ

АВТОР СЦЕНАРИЯ И РЕЖИССЕР
ДЖЕЙМС КЭМЕРОН

ТИТАНИК

ЛЕГЕНДА ВОЗВРАЩАЕТСЯ НА БОЛЬШОЙ ЭКРАН
В 3D

Лаборатория компьютерной графики и мультимедиа



Видеогруппа — это:

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