

# Алгоритм тональной компрессии на основе восприятия человека\*

К.С. Зипа, А.В. Игнатенко

kzipa@graphics.cs.msu.ru | ignatenko@graphics.cs.msu.ru

Факультет Вычислительной Математики и Кибернетики

Московский государственный университет им. М.В.Ломоносова

*Алгоритмы тональной компрессии преобразуют изображения широкого динамического диапазона к изображениям узкого диапазона, представимому на экране. Основные недостатки существующих локальных алгоритмов – это большое число не интуитивных параметров и привнесение дефектов. В данной статье предлагается модификация алгоритма тональной компрессии на основе глобальной оптимизации линейных окон (Globally Optimized Linear Windowed Tone Mapping [1]). Модификация основана на устройстве зрительной системы человека и позволяет улучшить результат алгоритма, а также заменяет не интуитивные параметры на несколько интуитивных, вариация которых в достаточно широком диапазоне не приводит к визуальному искажению изображения.*

**Ключевые слова:** тональная компрессия, восприятие.

## Tone mapping algorithm based on human perception\*

K.S. Zipa, A.V. Ignatenko

Faculty of Computational Mathematics and Cybernetics

Lomonosov Moscow State University, Moscow, Russia

*Tone-mapping algorithms transform high dynamic range images to low dynamic range images that can be displayed on screen. The main shortcomings of the existing local algorithms are a large number of non-intuitive parameters and bringing in artifacts. In this paper, we propose a modification of tone-mapping algorithm based on global optimization of linear windows. The modification is relied on human visual system functioning and allows to improve both global lightness and local contrast of algorithm's result. We also replaced non-intuitive settings by few intuitive. Variation of these parameters over a wide range does not lead to any visible distortions in output image.*

**Keywords:** tone mapping, perception.

### 1. Введение

Динамический диапазон яркости окружающего мира очень широк, в процессе эволюции человеческий глаз научился воспринимать его большую часть. В то же время, диапазоны современных устройств вывода ограничены всего двумя-тремя порядками. Это ограничение привело к тому, что большинство получаемых изображений содержат в себе лишь узкий диапазон яркости, что может приводить к потере информации в некоторых областях изображения (см. Рис.1).

В то же время, с помощью обычной бытовой камеры [2] можно получить изображение широкого динамического диапазона (high dynamic range image – HDR image), то есть такое изображение, количество градаций яркости в котором превышает 256 значений. В противопоставление, изображение с меньшим количеством градаций называют изображением узкого динамического диапазона (low dynamic range image – LDR image).



Рис. 1: Пример фотографий разных выдержек, в которых потеряна детализация в различных частях сцены.

Чтобы отобразить HDR-изображение на экране монитора, нужно осуществить сжатие яркости. Такое преобразование называют алгоритмом тональной компрессии. Качество алгоритма тональной компрессии определяется степенью схожести результата на результат восприятия наблюдателем сцены. Поэтому алгоритмы тональной компрессии связаны с изучением восприятия человека.

Работа опубликована при финансовой поддержке РФФИ, грант 15-07-20347

## 2. Постановка задачи

Общая цель заключается в создании алгоритма тональной компрессии, обеспечивающего сохранение информации, существенной с точки зрения человеческого восприятия.

$$O = TM(I), \|O - I\|_{human} \rightarrow 0$$

$$I = \{(p, i) \in [0, M], M < \infty, p \in P\} - \text{HDR-изображение}$$

$$O = \{O(p, i) \in [0, 1], \dots, N\}, p \in P\} - \text{LDR-изображение}$$

$$P - \text{множество пикселей изображения}$$

$$i = \{R, G, B\} - \text{цветовой канал изображения}$$

Для решения этой задачи из всех существующих алгоритмов был выбран базовый алгоритм, который был модифицирован для повышения качества результата и уменьшения сложности использования.

## 3. Обзор существующих методов

Существует множество алгоритмов тональной компрессии, обзоры которых можно найти в работах [3] и [4]. Различают две большие группы методов – глобальные и локальные. Алгоритмы из первой группы преобразуют яркость каждого пикселя изображения единообразно, используя общую для всех пикселей кривую сжатия. Достоинство этих методов в простоте, высокой скорости. Недостаток же в том, что при достаточно большом диапазоне входных значений яркости начинает пропадать детализация в некоторых областях.

С целью предотвратить это и начали использоваться локальные операторы, которые модифицировали значение яркости уже исходя из некоторой окрестности пикселя. Методы этого класса значительно сложнее, но позволяют обеспечивать лучшее визуальное качество результата. Однако общая проблема локальных методов состоит в привнесении искусственных искажений в результат. Иногда эти искажения могут быть настолько выраженными и значительными, что это сводит на нет их преимущество.

Для сохранения достоинств и устранения недостатков был предложен метод [1]. В отличие от других локальных методов алгоритм он не требует декомпозиции или сегментации входного изображения, что позволяет избежать артефактов. Обработка карты излучательности производится напрямую. Благодаря процессу глобальной оптимизации, обеспечивается локальная согласованность в каждом пикселе изображения.

Основные недостатки алгоритма – это, во-первых, необоснованный выбор цветового пространства для построения карты излучательности, что приводит к необходимости использовать эвристические способы нормировки, а во-вторых – необходимость

спецификации нескольких параметров, уникальных для каждой сцены.

Специфичность параметров можно проиллюстрировать на следующем примере (Рис. 2). Вверху дано изображение, полученное на стандартных параметрах, рекомендованных в статье ( $[\beta_1 = 0.6, \beta_2 = 0.2, \beta_3 = 0.1]$ ), а внизу – изображение, полученное при индивидуальном наборе параметров ( $[\beta_1 = 1.3, \beta_2 = 0.0, \beta_3 = 0.1]$ ). Видно, что использование стандартных параметров в данном случае привело к значительной яркостной деформации области неба.

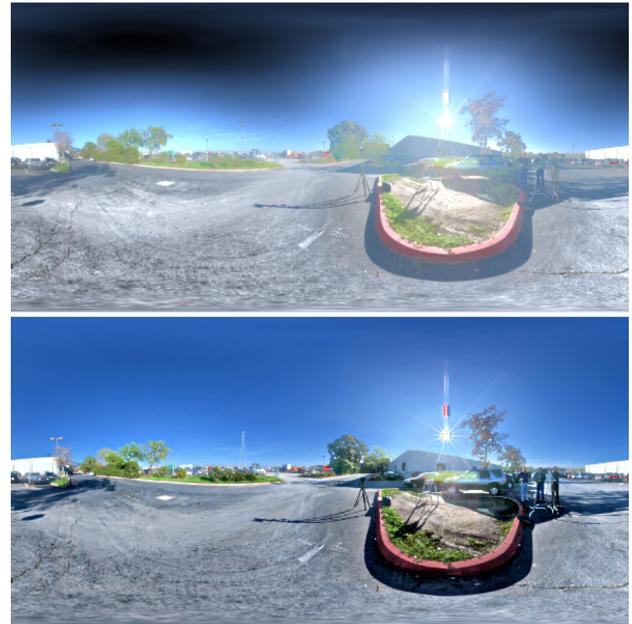


Рис. 2: Результаты работы базового метода (вверху – параметры  $[0.6, 0.2, 0.1]$ , внизу – параметры  $[1.3, 0.0, 0.1]$ ).

## 4. Предложенный алгоритм

Перейдем к описанию алгоритма. Вначале будет дана общая схема алгоритма, представленная в статье [1] и указаны блоки, которые предлагается видоизменить. Затем будут более подробно разъяснены произведенные модификации.

### 4.1 Общая схема алгоритма GOLWTM

На вход алгоритму подается изображение широкого динамического диапазона в формате RGB. На выходе алгоритм возвращает изображение узкого динамического диапазона в формате RGB. Основные шаги алгоритма следующие:

1. Построение карты излучательности  $Q$ . В каждом пикселе карты излучательности хранится информация о яркости для этого пикселя. Карта излучательности получается эвристическим

алгоритмом, использующим цветовое пространство HSV и приводящим итоговую среднюю яркость к фиксированному значению.

- Генерация карты рекомендации  $C$  по карте излучательности  $Q$ . Качество этой карты имеет решающее значение в качестве итогового алгоритма тональной компрессии. Авторы предлагают рассчитывать её по следующей формуле:

$$C(p) = (\mu(p)^{\beta_1} \sigma(p)^{\beta_2} Q(p)^{\beta_3} + k)^{-1}$$

$p$  – пиксель изображения

$Q$  – карта излучательности

$C$  – карта рекомендации

$\mu$  – локальное среднее (в окне  $3 \times 3$ )

$\sigma$  – локальное стандартное отклонение (в окне  $3 \times 3$ )

$\beta_1, \beta_2, \beta_3, k$  – параметры алгоритма

- Всё пространство изображения разбивается на множество пересекающихся окон размером  $3 \times 3$ , то есть каждый пиксель одновременно принадлежит 25 окнам. Внутри каждого локального окна предполагается наличие линейного преобразования яркости пикселя. Таким образом, требуется лишь найти параметры этого преобразования. Для этого предлагается в каждой точке минимизировать функционал разности:

$$f(q) = \sum_{p \in w(q)} (I(p) - u(p)Q - v(p))^2 + \varepsilon C^{-2} (u(p) - v(p))^2$$

$q$  – рассматриваемый пиксель изображения

$w(q)$  – множество окон, в которое входит  $q$

$p$  – пиксель изображения, входящий в окно

$I$  – выход этапа (излучательность узкого диапазона)

$Q$  – карта излучательности

$C$  – карта рекомендации

$u, v$  – параметры линейного преобразования

$\varepsilon$  – коэффициент регуляризации, позволяющий избежать тривиального решения

Минимизация функционала происходит путем рассмотрения системы линейных алгебраических уравнений и её решения. Детали этого процесса можно найти в статье [1].

Отметим, что из этого функционала виден смысл карты рекомендаций. Она показывает, насколько сильно можно сжать карту излучательности  $Q$ . Идея авторов статьи состояла в сжатии сильных перепадов яркости с сохранением детализации изображения, что обеспечивалось особенным построением карты, описанным в предыдущем пункте.

- Преобразование исходного HDR изображения в формате RGB по формуле Шлика [5]:

$$Output(p, i) = \left( \frac{Input(p, i)}{Q} \right)^s I$$

$p$  – пиксель изображения

$i$  – индекс цветового канала

$I$  – излучательность узкого диапазона

$Q$  – карта излучательности

*Input* – входные данные (HDR)

*Output* – выходные данные (LDR)

$s$  – степень насыщенности (задается пользователем).

После этого выходное изображение нормализуется по гистограмме с отсечением верхних и нижних 0.5%, а затем семплируется на 256 яркостных уровней.

Для преодоления недостатков предлагается заменить в алгоритме шаги (1) и (2) и (4). Шаги (1) и (4) комплиментарны и определяются взаимосвязью между цветовыми и яркостными значениями в выбранной цветовой модели. Они будут рассмотрены в следующем пункте 4.2. Модификация шага (2) является ключевой модификацией в предложенном методе, поскольку позволяет сместить фокус карты рекомендации с локальной четкости (обеспеченной статистическими величинами внутри малого окна) в сторону объема передаваемой информации, связанной с человеческим восприятием. Описание шага 2 представлено в пункте 4.3.

## 4.2 Выбор цветового пространства

Для расчета яркости пикселя в базовой статье предлагалось использовать компоненту  $V$  из цветового пространства HSV. Хотя это цветовое пространство достаточно широко используется в силу своей интуитивности, оно не обладает такими важными чертами, как, скажем, линейность к человеческому зрению. Кроме того, оно не отражает адаптивность восприятия и другие важные феномены зрительной системы. Предлагается использовать другое стандартизированное цветовое пространство – CIECAM02 [6]. Это самая последняя модель, одобренная CIE. В настоящее время, она является компромиссом между сложностью настройки и точностью передачи процессов, происходящих в зрительной системе. CIECAM02 позволяет вычислить 6 атрибутов цвета: яркость, светлоту, цветность (colorfulness), хроматическую насыщенность (chroma), яркостную насыщенность (saturation) и оттенок. Именно яркость в этой модели связана с восприятием отражающей способности объектов. И именно её логично использовать в качестве  $Q$  в алгоритме. Основные шаги расчета величин исходя из цвета в XYZ следующие:

- Конвертация цвета в CAT02 LMS. При этом увеличивается спектральная четкость цвета.
- Применяется хроматическая адаптация на основе информации об окружении и фоне.
- Происходит конвертация в пространство LMS, близкое к откликами колбочек.
- Сжатие откликов колбочек.

Тонкий момент в применении модели к задаче тональной компрессии заключается в необходимости обеспечить корректное восстановление хро-

матических компонент после осуществления преобразования над яркостью (шаг 4 базового алгоритма). В исходном алгоритме это осуществлялось благодаря формуле Шлика [5]. Однако, эта формула неверно работает для выбранной цветовой модели. Вместо неё существует другой, однозначный способ восстановления. Если вынести в совокупные коэффициенты сложные параметры, то яркость рассчитывается по формуле:

$$Q = k * (aL + bM + cS)^\alpha,$$

где  $L, M, S$  – результат сжатия откликов колбочек. Предположим, что яркость преобразовалась в новое состояние  $Q'$ . Мы можем увидеть, что новые значения для колбочек можно рассчитать по следующим формулам:

$$L' = L * \left(\frac{I}{Q}\right)^\alpha, M' = M * \left(\frac{I}{Q}\right)^\alpha, S' = S * \left(\frac{I}{Q}\right)^\alpha$$

Здесь коэффициент  $\alpha$  определяется исходя из параметров модели и не требует специального задания пользователем. Затем, произведем дополнительное выравнивание полученных откликов колбочек относительно исходных с помощью уравнивания значений яркости по гистограмме. Уравнивание проводится по отсечкам 0.5% относительно верхнего и нижнего пределов яркости в обоих изображениях. Это делается для того, чтобы избежать возможных ошибок в результате получения слишком больших значений после применения нелинейного коэффициента. После этого необходимо выполнить обратные преобразования хроматических компонент LMS в компоненты RGB. После этого нужно применить нормализацию по гистограмме, семплирование и гамма-коррекцию.

### 4.3 Расчет воспринимаемой яркости

Предлагается рассчитывать воспринимаемую яркость по модели BIWaM (Brightness Induction Wavelet Model) [7]. Затем эту воспринимаемую яркость можно использовать в качестве обратной карты рекомендаций  $C$ .

Модель описывает эффекты яркостного контраста, основываясь на известных свойствах зрительной системы. При применении модели моделируются многие иллюзии восприятия яркости. На Рис.3 приведены примеры иллюзий контраста. На примере слева иллюзия заключается в том, что серый квадрат выглядит более светлым на темном фоне и более темным на светлом. На примере справа серый ромб контрастирует с фоном, а не с примыкающей к нему решеткой.

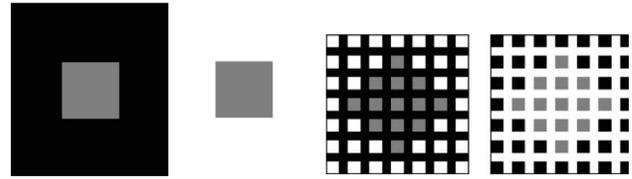


Рис. 3: Примеры иллюзий контраста.

В отличие от других аналогов, её применение не приносит сильных искажений в итоговое изображение, что делает её применимой в рассматриваемой задаче. У модели 2 параметра – базовая выраженность контраста и целевой масштаб, отражающий условия просмотра изображения. Каждый из этих параметров интуитивно понятен.

Эта модель осуществляет комбинацию яркости и контраста, самостоятельно устанавливая соотношения между ними на основе функции контрастной чувствительности. Модель использует декомпозицию изображения на набор частотно-ориентированных компонент.

Алгоритм, реализующий эту модель, принимает на вход одноканальное изображение, лежащее в диапазоне  $[0,1]$ . На выходе он возвращает другое изображение, вообще говоря, не ограниченное этим диапазоном. Поэтому для применения алгоритма необходимо нормировать полученную на первом шаге карту излучательности, а после применения алгоритма отрицательные значения можно ограничить небольшим положительным порогом, чтобы избежать деления на 0.

$$C = \max\left(\varepsilon, \text{BIWaM}\left(\frac{Q}{\max(Q)}\right)\right)^{-1}$$

$Q$  – карта излучательности

$\varepsilon = 0.03$  – небольшой коэффициент

Внутри себя алгоритм BIWaM работает следующим образом:

1. Декомпозиция входного изображения  $Q$  на набор  $3 * N + 1$  частотных компонент, где  $N = \log_2(\max(W, H))$  – максимально доступный уровень детализации ( $W, H$  – ширина и высота изображения  $Q$ ). На всех уровнях, кроме последнего присутствует по 3 разно-ориентированных компоненты, по смыслу отражающие контраст изображения. А на последнем уровне единственная компонента отражает общую яркость изображения.
2. На каждом уровне детализации (кроме последнего), для каждой ориентации, производится коррекция коэффициентов с использованием компоненты и двух нормирующих кривых:

$$w_{out}(s, o, p) = w_{in}(s, o, p) * (CSF_1(s) * Z_{s,o}(p) + CSF_2(s))$$

$w_{in}, w_{out}$  – вход и выход шага соответственно  
 $s, o$  – текущий масштаб и ориентация

$p$  – пиксель изображения

$Z_{s,o}(p)$  – контраст для текущей компоненты

$CSF_1(s), CSF_2(s)$  – кривые взвешивания для текущего масштаба

У кривых можно регулировать положение их пика (связан с целевым масштабом просмотра) и преобладание кривой 1 над кривой 2.

Контраст для текущей компоненты рассчитывается по формуле:

$$Z_{s,o}(p) = \frac{\sigma_{cen}^2(p)}{\sigma_{cen}^2(p) + \sigma_{sur}^2(p)},$$

$p$  – пиксель изображения

$Z_{s,o}(p)$  – контраст для текущей компоненты

$\sigma_{cen}^2$  – дисперсия в окне  $5 \times 5$

$\sigma_{sur}^2$  – дисперсия в окне  $13 \times 13$ , из которого удалена сердцевина размером  $5 \times 5$

3. Модифицированные компоненты собираются в результирующее изображение.

Использование алгоритма VIWaM позволяет сохранить на изображении детали, важные с точки зрения человеческого восприятия, и сжать детали, не важные с точки зрения человеческого восприятия.

## 5. Результаты алгоритма

Благодаря произведенным модификациям удалось улучшить качество алгоритма. На Рис.4 видно, что значительно повысилась детализация в области маски и платья. На Рис.5 видно, что исходный алгоритм потерял большую часть детализации в темных областях, а предложенный сохранил.

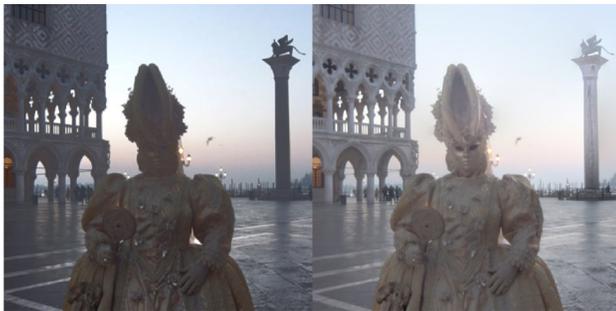


Рис. 4: Пример результатов исходного (слева) и модифицированного (справа) алгоритмов.



Рис. 5: Пример результатов исходного (слева) и модифицированного (справа) алгоритмов.

На Рис.6 витраж сохранил одновременно и высокую детализацию, и яркость относительно окружающих объектов. На Рис.7, наконец, показан результат на проблемном изображении, приведенном в начале статьи. Видно, что кроме отсутствия искажений, цвет на изображении стал более естественным.



Рис. 6: Пример результатов исходного (слева) и модифицированного (справа) алгоритмов.



Рис. 7: Пример результата модифицированного алгоритма.

## 6. Заключение

В статье предложена модификация алгоритма тональной компрессии на основе глобальной оптимизации линейных окон. Использование модели восприятия яркости позволяет отказаться от трудоемкой настройки параметров для каждого изображения. Модифицированный алгоритм демонстрирует улучшение с точки зрения количества различных деталей.

## Литература

- [1] Shan Q., Jia J., Brown M.S. Globally Optimized Linear Windowed Tone Mapping. IEEE Transactions On Visualization And Computer Graphics, Vol. 16, No. 4, July/August 2010.
- [2] Debevec P., Jitendra Malik. Recovering high dynamic range radiance maps from photographs. // Proceeding SIGGRAPH '08 ACM SIGGRAPH 2008 classes Article No. 31. 2008.

- [3] Devlin K. A Review of Tone Reproduction Techniques, // Technical Report CSTR-02-005, Dept. of Computer Science, Univ. of Bristol. 2002.
- [4] Reinhard E., Ward G., Pattanaik S., and Debevec P., High Dynamic Range Imaging. Morgan Kaufman, 2005.
- [5] Schlick C. A Customizable Reflectance Model for Everyday Rendering, //Proc. Fourth Eurographics Workshop Rendering, pp. 73-83, 1993.
- [6] Moroney N. The CIECAM02 color appearance model. // Color and Imaging Conference, 10th Color and Imaging Conference Final Program and Proceedings, pp. 23-27(5)
- [7] Otazu X., Vanrell M., Párraga A.C. Multiresolution wavelet framework models brightness induction effects. // Vision Res. 2008 Feb; 48(5):733-51.

### Об авторах

Кристина Зипа – аспирант Московского Государственного Университета, Факультет Вычислительной Математики и Кибернетики. Её контактные данные [kzipa@graphics.cs.msu.ru](mailto:kzipa@graphics.cs.msu.ru)

Алексей Игнатенко – к. ф.-м. н., с.н.с. Лаборатории Компьютерной Графики и Мультимедиа, Факультет Вычислительной Математики и Кибернетики. Его контактные данные [ignatenko@graphics.cs.msu.ru](mailto:ignatenko@graphics.cs.msu.ru)