

*Государственное образовательное учреждение высшего профессионального
образования*

**«Московский государственный технический университет
имени Н. Э. Баумана»
(МГТУ им. Н.Э. Баумана)**

ФАКУЛЬТЕТ

«Информатика и системы управления»

КАФЕДРА

«Программное обеспечение ЭВМ и информационные технологии»

Р А С Ч Ё Т Н О - П О Я С Н И Т Е Л Ь Н А Я З А П И С К А
к курсовой работе на тему:

Разработка программы по записыванию видео в расширенном
динамическом диапазоне

Студент

Киселев А.М.

(Подпись, дата)

Руководитель курсового проекта

Оленев А.А.

(Подпись, дата)

Москва 2018

Содержание

Введение	3
1 Аналитический раздел	4
1.1 Различия LDR и HDR изображений	4
1.2 Процесс получения HDR изображения	5
1.2.1 Метод получения HDR изображения на камере со слабым сенсором	6
1.2.2 Получение HDR видео на камере со слабым сенсором в режиме реального времени	6
1.2.2.1 Способы получения разной экспозиции в видео	6
1.3 Выравнивание изображений	7
1.3.1 Алгоритм MTB(Median Threshold Bitmap)	7
1.4 Удаление артефактов изображения	10
1.4.1 Алгоритм BMD(Bitmap Movement Detection)	10
1.5 Получение HDR изображения	12
1.5.1 Слияние экспозиций	12
2 Конструкторский раздел	15
2.1 Описание метода	15
2.2 Декомпозиция разрабатываемого метода	15
3 Технологический раздел	19
3.1 Выбор инструментов для разработки	19
4 Исследовательский раздел	20
Список использованных источников	21

Введение

В 20 веке был принят стандарт CIE, который охарактеризовывал то, как должна отображаться картинка на восьмибитных мониторах. Технологии того времени позволяли передавать только сильно ограниченный диапазон цветовой информации, поэтому картинка, предоставляемая монитором, была далека от более сочных и ярких цветов, которые человек видит в повседневной жизни. Такое явление может быть охарактеризовано следующими терминами: LDR(low dynamic range) или SDR(standart dynamic range) image – маленький динамический диапазон или стандартный динамический диапазон изображения. Мониторы LDR и SDR не могли передавать широкий спектр цветов. К тому же не все сенсоры цифровых камер могут позволить себе захватить широкий диапазон оттенков. Из-за этого возникает проблема засвеченных и затемненных участков на фотографии или видео.

С развитием технологий, стали появляться так называемые HDR мониторы, которые позволяют передавать цветовую информацию с глубиной в 10 битов на канал. По причине того, что LDR и SDR мониторы обладают маленькой глубиной цвета, они не в состоянии корректно отображать картинку с широким диапазоном цветов.

Из-за того, что не все сенсоры могут запечатлить широкий спектр цветов, были придуманы специальные алгоритмы и методы получения HDR изображений с обычных и самых распространенных сенсоров(цифровые камеры, мобильные телефоны, планшеты, веб-камеры и т.п.). Полученные HDR снимки или видео с камер, сенсор которых рассчитан на LDR и SDR изображения, могут быть отображены на HDR мониторах.

Получение HDR кадра - нетривиальная задача, которая делится на несколько этапов: получения серии снимков с разной длиной экспозиции, выравнивание кадров по отношению друг к другу, реконструкция и удаление движущихся объектов полученных изображений, объединение изображений(merging), проведение цветовой коррекции изображения.

1 Аналитический раздел

SDR(Standart Dynamic Range) изображение[1] – изображение, пиксели которого содержат цвета и яркость, максимальная глубина которых 8 бит.

LDR(Low Dynamic Range) изображение[1] – изображение, пиксели которого хранят ограниченный диапазон цветов и яркости, предназначенное для отображения на мониторах с глубиной цвета 8 бит или меньше.

HDR(High Dynamic Range) изображение[2] – изображение, пиксели которого содержат более широкие значения цвета и яркости в сравнении с изображениями стандартного диапазона(SDR), глубина цвета которых больше 8-ми бит.

Глубина цвета – количество бит, приходящихся на один пиксель

1.1 Различия LDR и HDR изображений

Несмотря на то, что технологии за последние несколько лет быстро развиваются и качество полученных кадров и устройств, их отображающих, увеличивается, получение картинки, сопоставимой с реальным окружением, остается нелегкой задачей. Яркость, диапазон цветов, которые видит человек в повседневной жизни, невозможно отобразить на большинстве мониторов.

Хотя уже начинают появляться так называемые HDR мониторы и существуют камеры, сенсоры которых позволяют запечатлить широкий спектр цветов, яркостей и деталей, цена таких устройств очень большая, поэтому большинство мониторов остаются SDR формата и не в состоянии передать картинку с большим количеством цветовой информации.

Глубина цвета на большинстве мониторов составляет 8 бит. Глубина цвета самого распространенного формата изображений JPEG также составляет 8 бит, в котором используется цветовое пространство YC_rC_b [1]. Это цветовое пространство позволяет использовать лишь малую часть видимых человеком цветов. Так же это цветовое пространство не способно передать большую часть яркостей, которых способен распознать человеческий глаз.

В противовес формату JPEG существует так называемый RAW формат. В отличии от JPEG он способен содержать гораздо больше информации. Глубина цвета такого формата может достигать 12-16 бит(Значение может варьироваться в зависимости от возможностей сенсора камеры), который доступен в большинстве современных камер. Чаще всего этот формат автоматически переводят в JPEG во время съемки, что приводит к потере многих деталей без возможности восстановления. Однако, используя специальные инструменты RAW изображение в дальнейшем можно преобразовать в LDR изображение, получив на выходе кадр без потери нужной информации(проводи tonemapping)[1].

На цветовом пространстве 1.1 наглядно показан охват возможных отображаемых цветов в RGB пространстве или на тех же LDR мониторах и в HDR с глубиной цвета хотя бы в 12 бит.

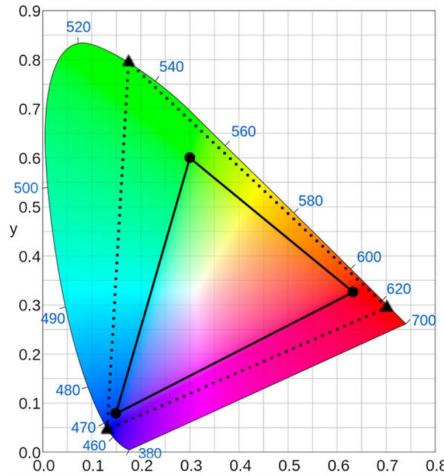


Рисунок 1.1 — Охват видимых человеком цветов цветовым пространством RGB и HDR.

Главное различие HDR и традиционных LDR изображений[1] в том, что HDR картинка не зависит от какого-либо устройства и содержит максимально большое количество информации. Количество отображаемой информации на устройстве уже зависит непосредственно от его собственных характеристик и возможностей. LDR изображение же зависит от устройства, на котором оно будет отображаться и спроектировано специально под мониторы или экраны, которые отображают определенное количество цветовой информации.

1.2 Процесс получения HDR изображения

HDR изображение можно получить несколькими способами[1]:

- с помощью объединения снимков с разной длинной экспозиции,
- с помощью камеры, сенсор которой позволяет захватить широкий объем данных,
- при помощи перевода LDR изображения в HDR специальными алгоритмами

Первый метод является более распространенным, так как устройства, которые больше всего распространены в повседневной жизни(телефоны, планшеты, вебкамеры) не обладают достаточно мощными сенсорами, для того, чтобы захватить широкий диапазон цветов. Последний метод не получил распространения, потому что задача перевода LDR или SDR изображения в HDR возможна только при помощи преминения алгоритмов реконструкции, связанных на нейронных сетях, появившихся достаточно недавно.

1.2.1 Метод получения HDR изображения на камере со слабым сенсором

- Выравнивание кадров,
- Реконструкция и удаление движущихся объектов(артефактов),
- Слияние кадров с разной экспозицией,
- Преобразование HDR к LDR(tonemapping) если это требуется

Наглядно процесс получения HDR кадра с возможностью перевода его в LDR[3] показан на рисунке 1.2

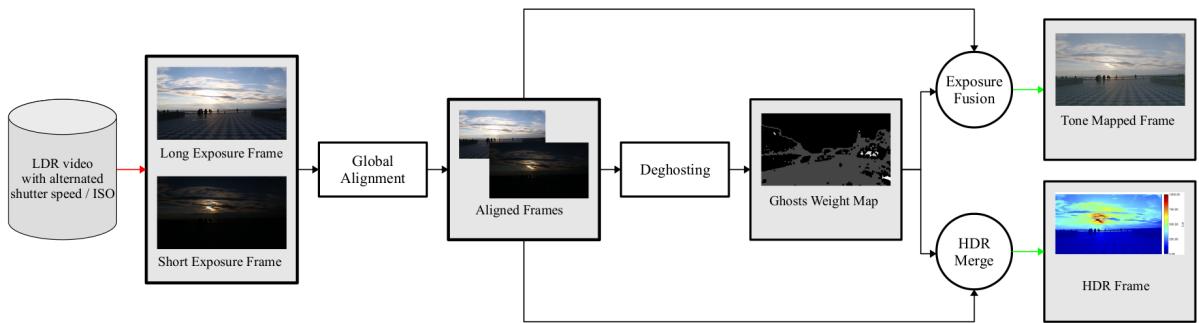


Рисунок 1.2 — Процесс получения HDR изображения путем слияния кадров с разной экспозицией

1.2.2 Получение HDR видео на камере со слабым сенсором в режиме реального времени

Данная задача подразумевает собой быстрое создание HDR изображений из 2-3 кадров с изменяющейся экспозицией, которые получаются, путем изменения экспозиции через каждый последующий кадр. Таким образом, можно быстро получить два кадра, которые будут содержать всю информацию о самых темных и самых светлых участках сцены. Так, как эти кадры следуют друг за другом, различия в них, предполагается, будут незначительными.

При записи видео в таком формате очень важна скорость алгоритмов, использующихся при получении HDR кадра. Операции выравнивания, удаления артефактов, слияния кадров и преобразование полученного кадра в LDR должны протекать быстро.

1.2.2.1 Способы получения разной экспозиции в видео

Получение разной экспозиции кадров можно добиться несколькими способами[1]:

- а) Использовать несколько камер, каждая из которых будет вести запись с соответствующей экспозицией;

- б) Использовать камеру с нескольким сенсорами, каждый из которых будет вести запись с соответствующей экспозицией;
- в) Использовать одну камеру, меняя параметр выдержки или ISO, которые влияют на полученную экспозицию;

Первые два способа позволяют сохранить количество кадров в секунду и улучшить качество получаемой картинки, но они требуют предварительной калибровки. Так же несколько камер и камеры с дополнительными сенсорами не самые распространенные решения в телефонах и других гаджетах.

Третий способ позволяет получать кадры с разной экспозицией на любом устройстве, в котором есть камера, что делает его более универсальным, но это снижает количество кадров в секунду почти в два раза.

1.3 Выравнивание изображений

Первым шагом в получении HDR изображения является глобальное выравнивание последовательности кадров с разной экспозицией. Предполагается, что 2 соседних кадра не меняются слишком быстро.

Но при съемке с рук или при съемке в движении, кадры все равно меняются, положение записывающего устройства тоже может быть нестабильным. В дальнейших шагах получения HDR снимка очень важно, чтобы кадры с разной экспозицией были максимально выровнены по отношению друг к другу.

Задача выравнивания усложняется тем, что экспозиции изображений сильно разнятся. Для решения этой задачи используется алгоритм MTB(Median Threshold Bitmap)[4] пороговое отсечение по медианному значению.

1.3.1 Алгоритм MTB(Median Threshold Bitmap)

На вход алгоритму подается N 8-битных монохромных изображений, которые можно получить путем использования зеленого канала изображения, либо путем перевода каждого пикселя 24-битного sRGB изображения в серый цвет с помощью выражения 1.1

$$grey = \frac{(54 * red + 183 * green + 19 * blue)}{256} \quad (1.1)$$

Одно из изображений выбирается базовым. Остальные N-1 при выравнивании будут опираться на базовое.

Далее требуется для каждого из изображений:

- а) Найти медианное значение(8-битное значение) из гистограммы низкого разрешения пикселей монохромного изображения.

б) Создать растровое изображения, в котором 0 будут отмечаться пиксели, которые меньше или равны медианному значению и 1 будут отмечаться пиксели, которые строго больше этого медианного значения.

Результаты этих шагов представлены на примере двух изображений с разной экспозицией 1.3

После этого каждое N-1 изображение подгоняется под базовое. Вначале они сравниваются. Сравнение происходит путем побитовой операции XOR – каждый пиксель изображения высчитывается с каждым соответствующим пикселям базового изображения операцией XOR. Результаты сохраняются и на выходе получается очередное растровое изображение, в котором 0 - пиксели, информация которых совпала с базовым изображением и 1 - пиксели, информация которых не совпала.

Для устранения шумов, которые могут получиться при съемке в темных, плохо освещенных помещениях, используется *исключающее растровое изображение*. Т.к. шумы возникают близко к медианному значению, пиксели, лежащие на расстояние ближе чем 4 пикселя к медианному значению не учитываются.

Так генерируется исключающее растровое изображение, в котором 0 – отмечается пиксель, лежащий близко к медианному значению, иначе 1.

Для выравнивания находится значение сдвига (x, y): x - смещение изображения по горизонтали, y - сдвиг изображения по вертикали. Самый оптимальный сдвиг будет тот, у которого при сравнении изображения с базовым, количество пикселей 1 будет минимальное. Искать его можно несколькими способами. Самый очевидный - простой перебор всевозможных сдвигов (x, y), но используется более оптимальный метод. С помощью пирамidalной сегментации изображения[5] можно добиться наименьшего количества сравнений двух растровых побитово размеченных изображений.

Каждое монохромное изображение переводится на следующий уровень путем уменьшения его разрешения в два раза. Таким образом получаем несколько уровней, которые содержат монохромные изображения. Начиная с самого последнего уровня(изображения которых содержат наименьшее разрешение) нужно найти для них исключающее растровое изображение по медианному значению, найти для них растровое изображение по медианному значению и с помощью операции AND получить растровое изображение, которое должно будет содержать минимальное число 1. Далее найти сдвиг перебором по значениям +1 пиксель в каждом направлении(по горизонтали и вертикали)

На следующем уровне "пирамиды" нужно домножить полученный сдвиг на 2(значение, на которое менялось разрешение на каждом уровне пирамиды) и снова вычислить наилучший сдвиг на расстоянии +1 пиксель от текущего сдвига в разные направления x и y .



Рисунок 1.3 — Получение растровых изображений, использующихся для дальнейшего сравнения(справа), оригиналы изображений(слева)

Эти действия продолжаются до последнего(изначального) уровня пирамиды, на котором получается окончательный требуемый сдвиг (x, y) для изображения, чтобы быть выравненным по отношению к базовому.

1.4 Удаление артефактов изображения

При записи видео с камеры, объект, который ведет съемку, или окружение вокруг него могут находиться в движении. По этой причине могут появляться так называемые "призраки". На одном из этапов создания HDR изображения полученные кадры сливаются. Но если предварительно не позаботиться о движущихся объектах, на результирующем изображении можно наблюдать подобные артефакты, пример на изображении 1.4



Рисунок 1.4 — Изображение с удаленными артефактами движущегося объекта(справа), артефакты на необработанном изображении(слева)

Во время съемки видео такие артефакты могут значительно повлиять на отображаемую информацию исходной картины. Чтобы этого не допустить требуется позаботиться об удалении избыточной информации перед слиянием кадров с разной экспозицией.

В статье [6] рассматриваются разные решения удаления артефактов при получении HDR изображения. По результатам их исследований алгоритм BMD является наиболее оптимальным в получении изображения с динамически расширенным диапазоном.

1.4.1 Алгоритм BMD(Bitmap Movement Detection)

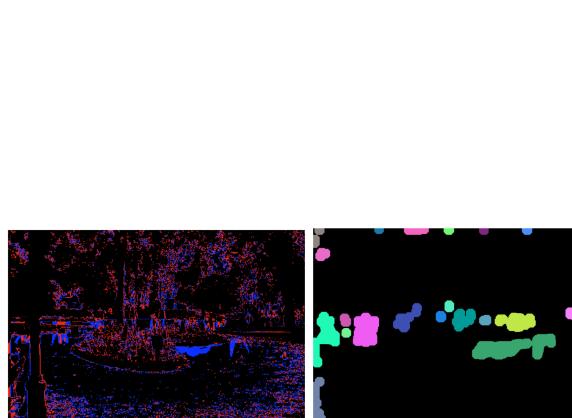
Алгоритм BMD[6] основывается на использовании MTB[4] растрового изображения. Использование побитовых операций позволяет добиться быстрой скорости и точности вычислений. В качестве результата алгоритм предоставляет *Размеченную матрицу движений*. С помощью этой матрицы при склеивании кадров с разной экспозицией можно будет удалить артефакты.

Первым делом требуется найти для каждого изображения МТВ растровое изображение. Пусть B_i – массив этих растровых изображений. На неподвижной, статичной сцене предполагается, что каждый пиксель сохраняет свое значение на всех растровых изображениях, хранящихся в B_i . Очевидно, что если значение пикселя поменяется, в области пикселя произошло движение.

Суммируя все растровые изображения из B_i между собой, получается M^* , в которой любые значения кроме 0 или N (количество исходных изображений) рассматриваются как совершенное движение в районе пикселя.

M^* может содержать некоторое количество шумов, которые могут повлечь за собой неприятные последствия – некорректное определение области движения. Использование размытия по Гауссу позволяет значительно уменьшить шумы. Таким образом M^* переводится в M – матрицу движений 1.5.

Затем матрица движений переводится в кластерное отображение, которое получается с помощью Connected Component labeling, где у каждого кластера есть своя метка. Отсюда получается размеченная карта движений L_M с размеченными кластерными зонами Ω_i , которые содержат пиксели, влияющие на появление артефактов.



(а) Представление матрицы движений(справа), и матрицы движений до преобразований(из-за большого количества шума сложно понять где и что передвигается).



(б) Исходная последовательно изображений с разными экспозициями.

Рисунок 1.5 — Пример нахождения движений на последовательности изображений с разными экспозициями.

1.5 Получение HDR изображения

Целью данной работы является получение HDR изображение на LDR мониторах, которые не в состоянии отобразить широкий диапазон яркостей. В связи с этим наиболее подходящим алгоритмом в получении такой картинки является *слияние экспозиций*[7]. Данный метод позволяет получить сразу LDR изображение с динамически расширенным диапазоном без применения tonemapping.

1.5.1 Слияние экспозиций

С помощью слияния экспозиций изображение выстраивается из "лучших" частей входных изображений с разными экспозициями. Оно рассчитывается с помощью матрицы весов. Исходные изображения объединяются, опираясь на нее(изображение 1.6а). Каждое значение матрицы весов основывается на трех главных параметрах: контрасте, насыщенности и на экспозиционном значении.

a) **Контраст:** каждое входное изображение переводится в монохромное и на него накладывается фильтр Лапласса. Абсолютное значение каждого элемента полученной матрицы будет являться контрастностью C соответствующего пикселя.

б) **Насыщенность:** Насыщенностью S будет являться различие красного, зеленого и синого каналов каждого пикселя.

в) **Значение экспозиции:** Значения интенсивностей каналов показывает на сколько хорошо выстроена экспозиция пикселя. Интенсивности должны лежать между нулем(засвеченное изображение) и единицей(слишком темное изображение). Каждая интенсивность канала рассчитывается с помощью уравнения 1.2, в котором определяется на сколько близко значение лежит к 0.5 с помощью кривой Гаусса.

$$E = \exp \left(-\frac{(i - 0.5)^2}{2 * \sigma^2} \right) \quad (1.2)$$



(a) Входные изображения с соответствующими матрицами весов



(b) Изображение после слияния кадров

Рисунок 1.6 — Пример построение изображения путем слияния кадров.

, где i - интенсивность канала, σ - константа, равняющаяся 0.2.

Вычисления выше используются для построения матрицы весов и вычисляются для каждого пикселя путем перемножения всех трех параметров: контрастности, насыщенности и значения экспозиции. Таким образом, в одном значении матрицы весов учитываются все эти параметры. Вычисление значений матрицы весов представлено в 1.3

$$W_{ij,k} = (C_{ij,k})^{\omega_C} \times (S_{ij,k})^{\omega_S} \times (E_{ij,k})^{\omega_E} \quad (1.3)$$

, где С, S и E - контраст, насыщенность и значение экспозиции соответственно. ω_C , ω_S и ω_E - параметры, с помощью которых регулируется приоритет того или иного значения. ij , k - (i , j) пиксель k -того изображения. Если значение ω равняется 0, то соответствующее значение не учитывается. Полученный вес пикселя $W_{ij,k}$ будет использоваться при слиянии кадров.

Для слияния N изображений вначале высчитывается среднее значение веса каждого пикселя. Вычисление среднего значения веса представлено в выражении 1.4

$$\hat{W}_{ij,k} = \left[\sum_{k'=1}^N W_{ij,k'} \right]^{-1} W_{ij,k} \quad (1.4)$$

Для дальнейшего слияния изображений используется алгоритм пирамидальной декомпозиции изображения(пирамида Гаусса и пирамида Лапласса)[8], который позволяет получить отфильтрованные изображения в разных разрешениях. Затем производится слияние кадров на каждом из уровней пирамиды.

Получить результирующее изображение с произведенным слиянием можно с помощью выражения 1.5

$$L\{R\}_{ij}^l = \sum_{k=1}^N G\{\hat{W}\}_{ij,k}^l L\{I\}_{ij,k}^l \quad (1.5)$$

Где l - уровень пирамиды декомпозированного изображения, а $L\{A\}^l$ - декомпозиция изображения A на l уровне пирамиды Лапласса и GB^l - декомпозиция изображения B на уровне l пирамиды Гаусса. Пример такой декомпозиции представлен на изображении 1.7

Самый нижний уровень пирамиды $L\{R\}$ и будет являться результатом слияния экспозиций - R .

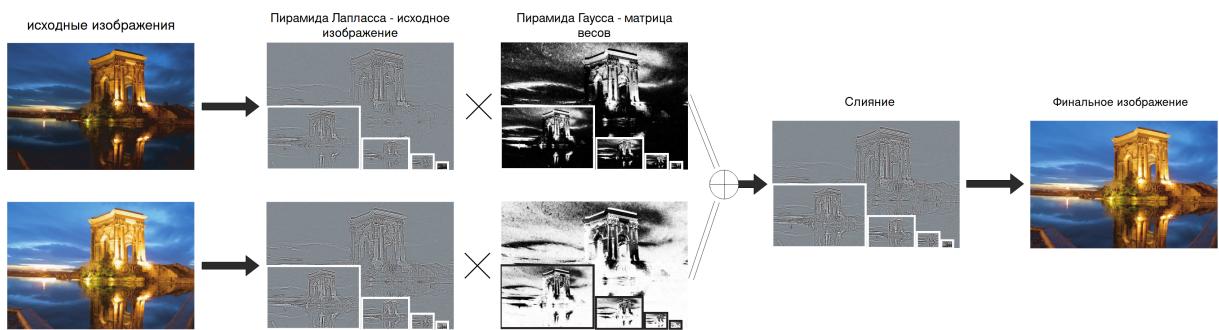


Рисунок 1.7 — Представление процесса слияния экспозиций

2 Констукторский раздел

Рассмотрим задачу получения видео с динамически расширенным диапазоном из поданного потока данных.

2.1 Описание метода

Особенностью решения данной задачи является то, что исходное видео, подающееся на вход, может быть получено путем ручной съемки. Так как видео материал это не что иное, как последовательность кадров, они могут быть не выровнены по отношению друг к другу, а так же объекты, запечатленные на камеру, могут быть не статичны, что усложняет получение HDR изображения, а следовательно и самого видео. В качестве алгоритма выравнивания(регистрации изображения) был выбран алгоритм MTB[4]. Алгоритм для удаления движущихся объектов на последовательности кадров - MBD[6]. Так как стоит задача отображения HDR видео на LDR мониторе, наиболее подходящим алгоритмом слияния кадров является алгоритм слияния экспозиций[7]. Так же требуется выбирать кадры с разной экспозицией из последовательности кадров(видео). Для алгоритма выделяются основные этапы:

- Выбор нужных кадров из видео материала
- Выравнивание кадров по отношению друг к другу(MTB алгоритм)
- Нахождение артефактов(возникших из-за движущихся объектов)
- Получение LDR изображения с расширенным динамическим диапазоном
- Объединение полученных кадров в видео ряд и сохранение на устройстве

В итоге, вся задача сводится к работе с изображениями.

2.2 Декомпозиция разрабатываемого метода

Представим задачу получения HDR видео ряда на LDR мониторе в виде диаграммы "IDEF0"на Рисунке 2.1, Рисунке 2.2, Рисунке 2.3, Рисунке 2.4, Рисунке 2.5, Рисунке 2.6

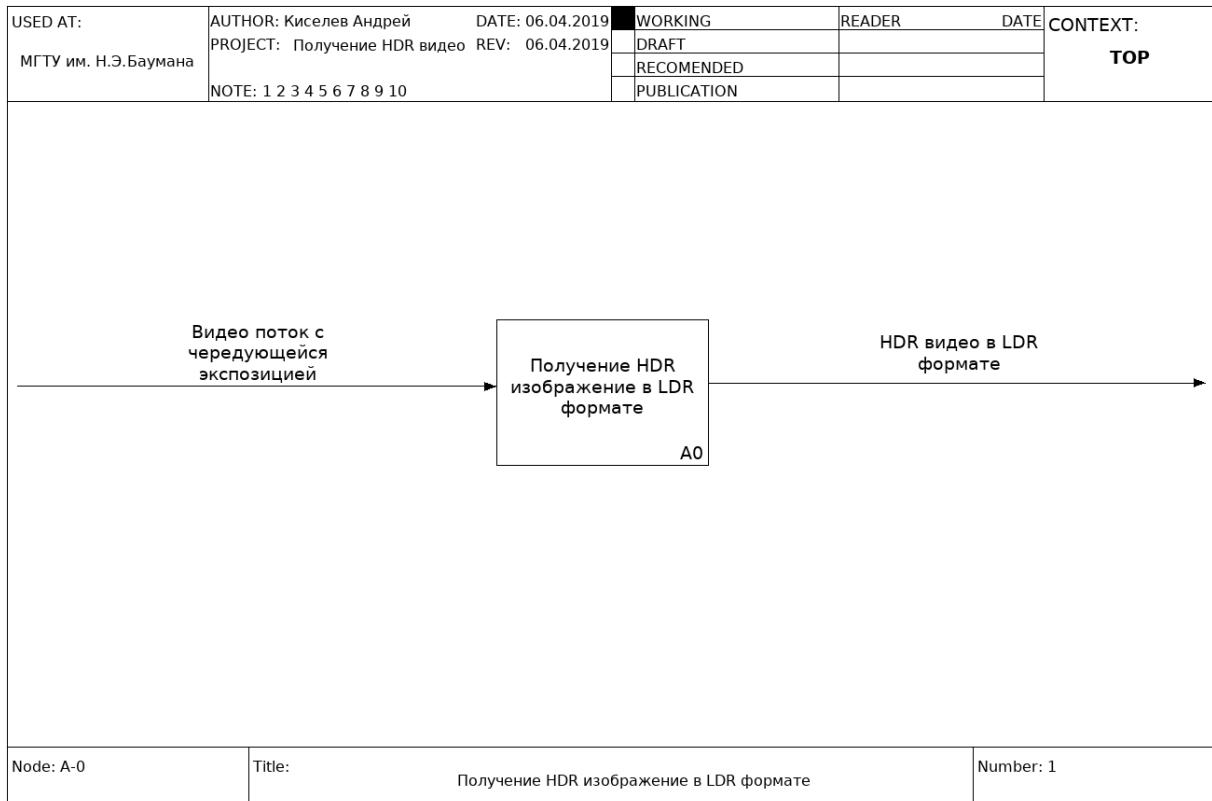


Рисунок 2.1 — IDEF0 диаграмма метода(A)

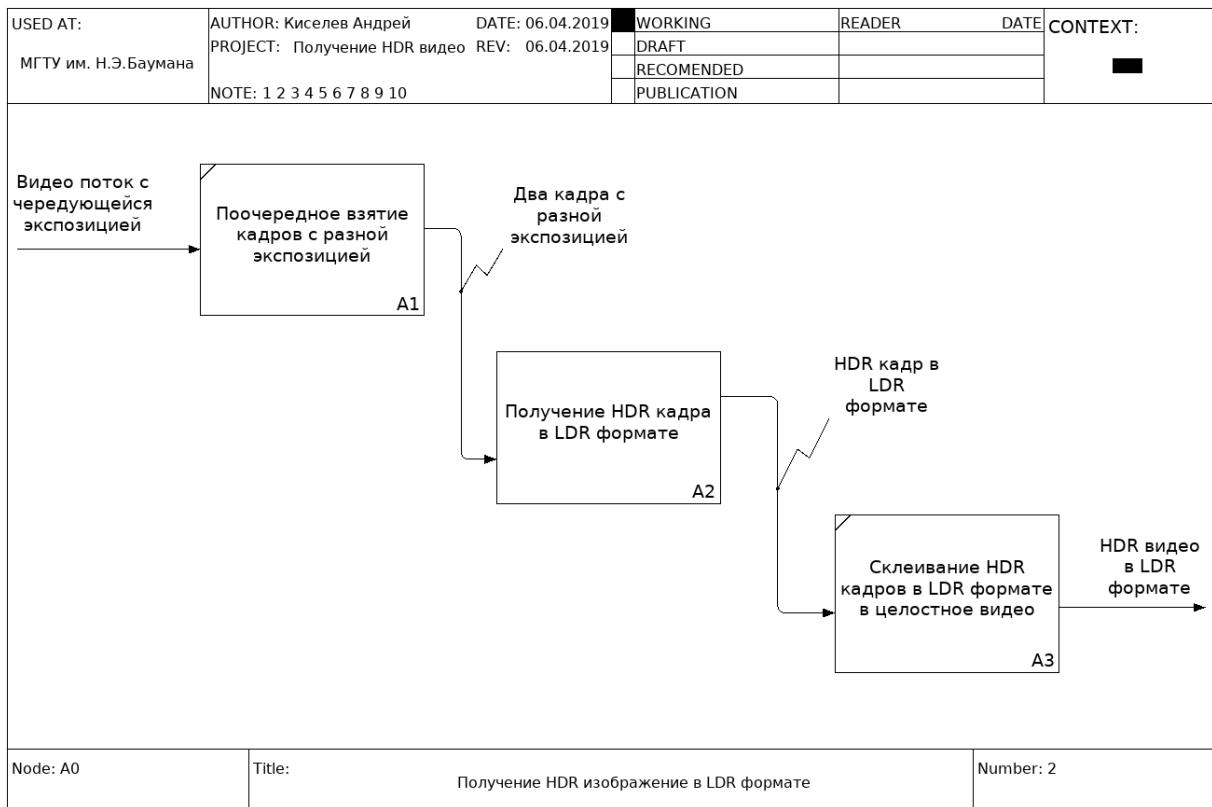


Рисунок 2.2 — IDEF0 диаграмма метода(A0)

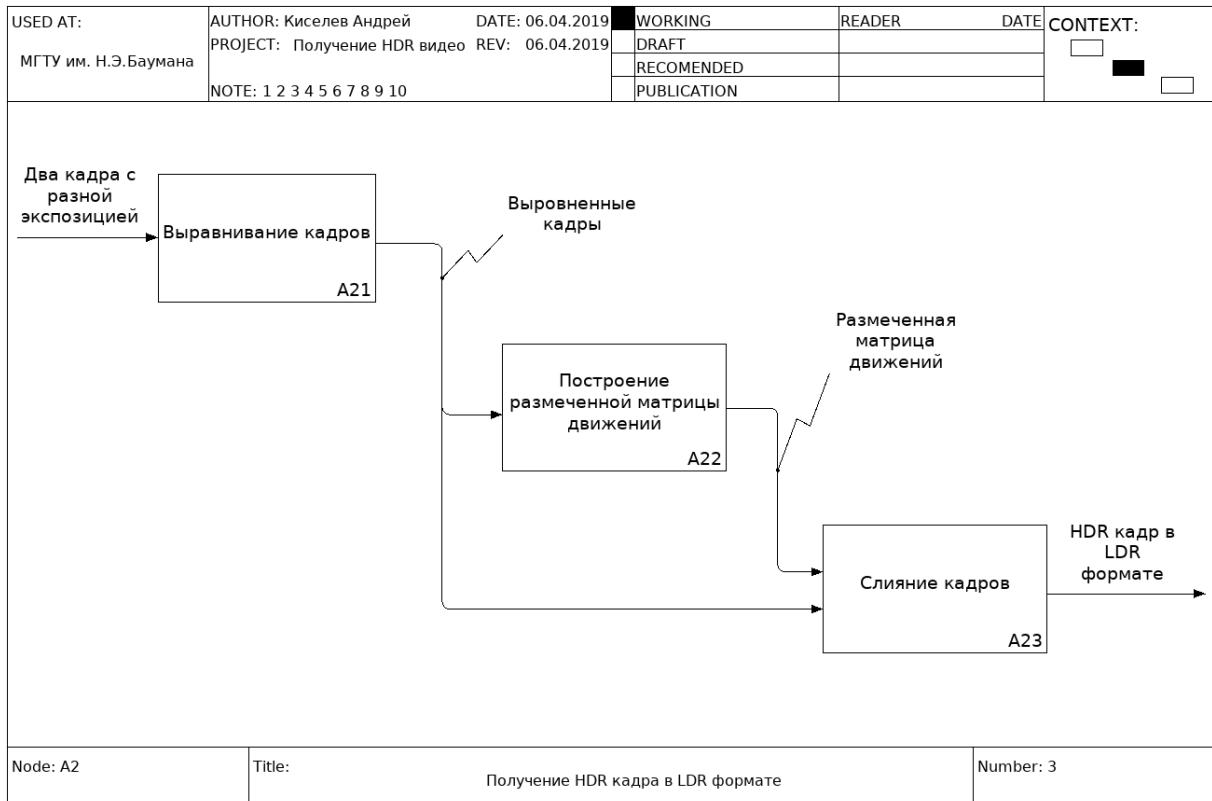


Рисунок 2.3 – IDEF0 диаграмма метода(A2)

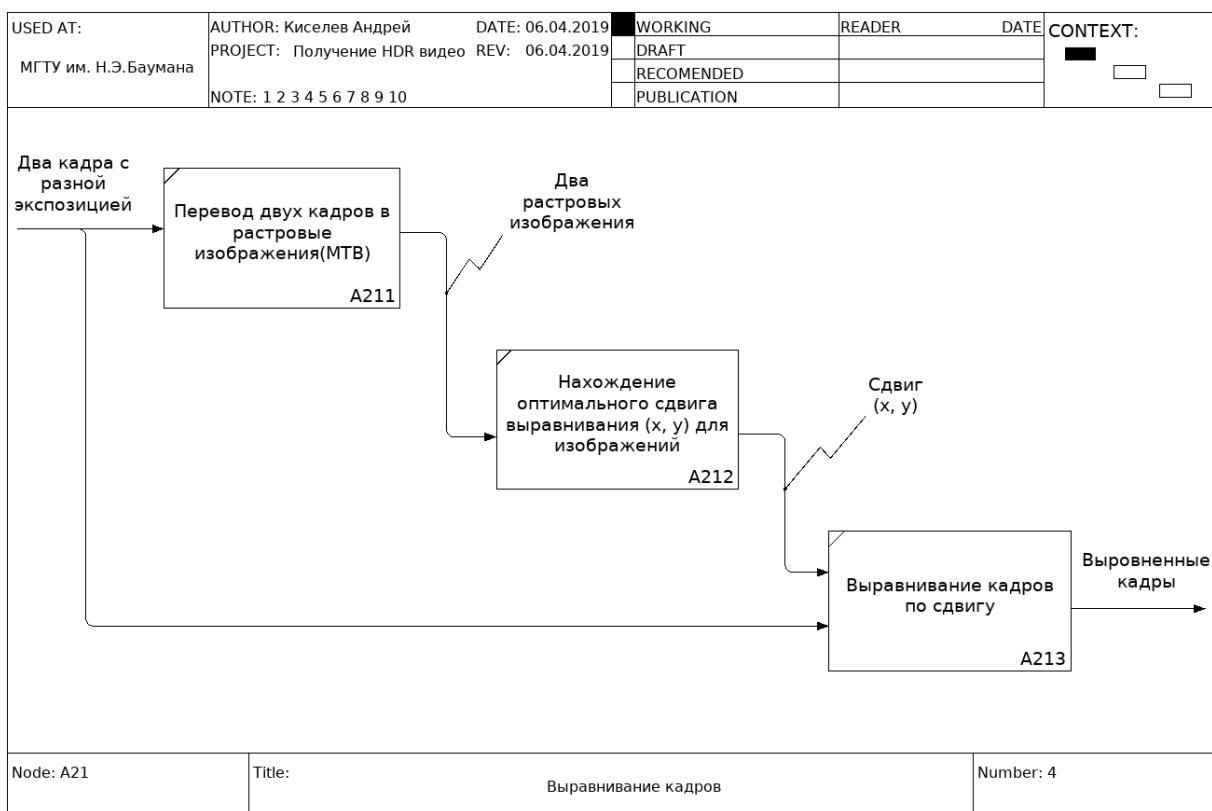


Рисунок 2.4 – IDEF0 диаграмма метода(A21)

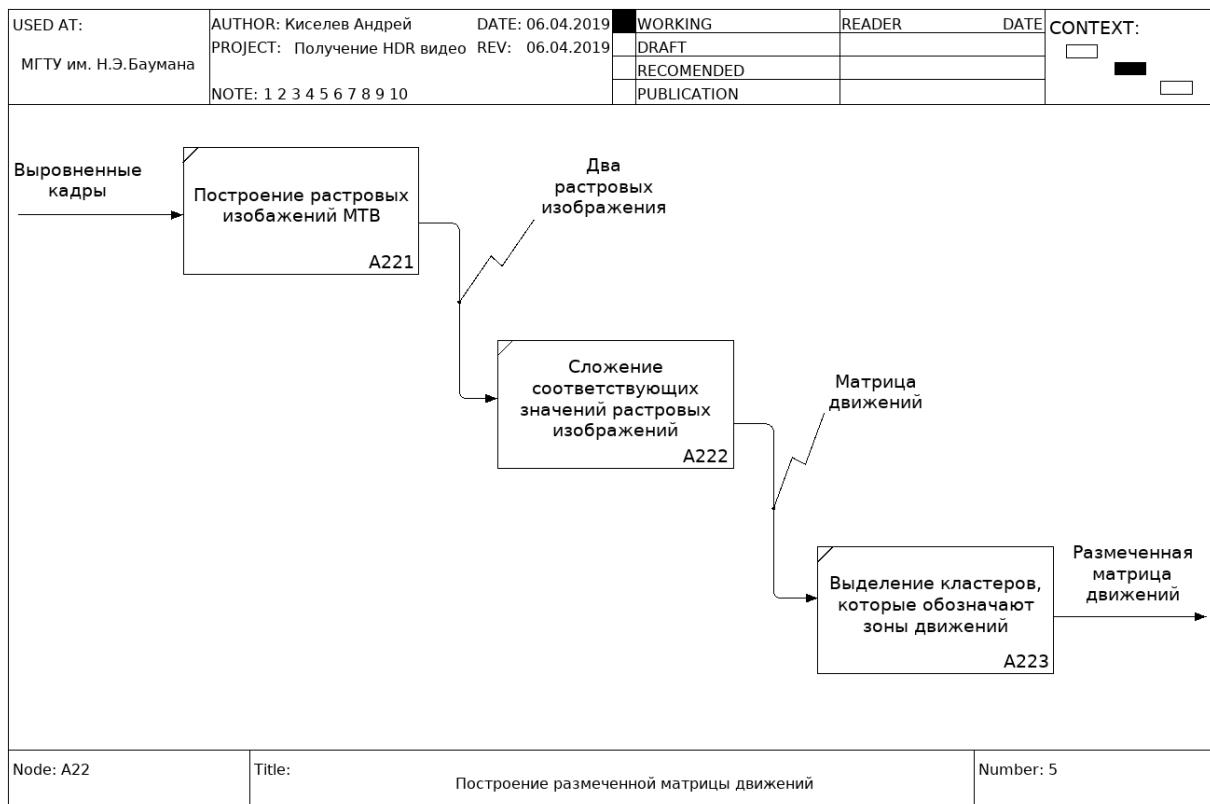


Рисунок 2.5 – IDEF0 диаграмма метода(A22)

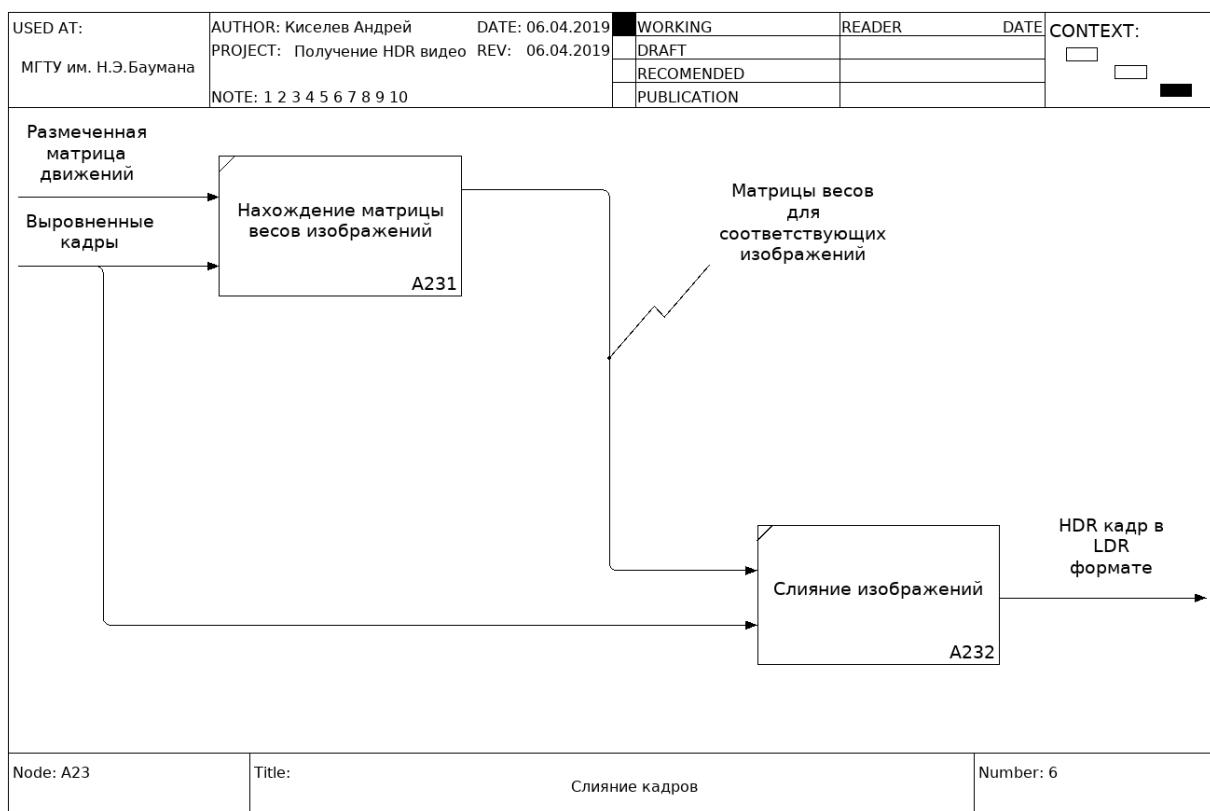


Рисунок 2.6 – IDEF0 диаграмма метода(A23)

3 Технологический раздел

В данном разделе описаны инструменты, использованные при реализации приложения, демонстрирующего работу метода составления HDR видео, отображаемом на LDR мониторе.

3.1 Выбор инструментов для разработки

Для разработки приложения был выбран язык C# с использованием библиотеки OpenCV. Данная библиотека ориентирована на работу с изображениями и с камерами и предоставляет удобный интерфес для работы с видео потоками, а так же универсальные структуры данных, которые позволяют работать с многоканальными изображениями.

Графическим интерфейсом и средой разработки был выбран Qt и Qtcreator. В Qtcreator входят удобные инструменты по работе с библиотекой Qt для профилирования работы и разработки графического интерфейса.

4 Исследовательский раздел

Список использованных источников

1. *Mantiuk, R. K.* High Dynamic Range Imaging. / R. K. Mantiuk, K. Myszkowski, Hans-Peter Seidel. — Wiley Encyclopedia of Electrical and Electronics Engineering, 2016.
2. *Eilertsen, Gabriel.* The high dynamic range imaging pipeline. Tone-mapping, distribution, and single-exposure reconstruction. / Gabriel Eilertsen. — Devision of Media and Information Technology. Department of Science and Technology, 2018.
3. Practical-HDR: A Simple and Effective Method for Merging High Dynamic Range Videos. / E. Akcora, F. Banterle, M. Corsini et al. — ACM press, 2016.
4. *Ward, Greg.* Fast, Robust Image Registration for Compositing High Dynamic Range Photographs from Handheld Exposure / Greg Ward. — 2012.
5. *Thevenaz, Philippe.* A Pyramid Approach to Sub-Pixel Registration Based on Intensity / Philippe Thevenaz, Urs E.Ruttmann, Michael Unser. — IEEE Transactions on Image Processing, 1998.
6. *Pece, Fabrizio.* Bitmap Movement Detection: HDR for Dynamic Scenes. / Fabrizio Pece, Jan Kautz. — Conference on Visual Media Production(CVMP), 2010.
7. *Mertens, Tom.* Exposure Fusion / Tom Mertens, Jan Kautz, Frank Van Reeth. — Pacific Graphics, 2007.
8. *Burt, P.* The Laplacian Pyramid as a Compact Image Code / P. Burt, T.Adelson. — IEEE Transactions on Communication, 1983.