

Сравнение алгоритмов нейросетевой и круговой фильтрации синтезированных RGBD изображений

И.Е. Кинёв¹, Г.В. Гебель¹, Д.Д. Жданов¹, А.Д. Жданов¹

¹ Университет ИТМО, Кронверский проспект д.49, литера А, Санкт-Петербург, 197101, Россия

Аннотация

Проведено исследование причин возникновения конфликта вергенции-аккомодации человеческого зрения в системах виртуальной и смешанной реальностей. Рассмотрены технические и алгоритмические подходы по снижению и устранению конфликта вергенции-аккомодации в системах виртуальной реальности. В качестве технического решения выбран подход, обеспечивающий адаптивную фокусировку окуляра системы виртуальной реальности к точке конвергенции глаз человека, определяемой системой трекинга его зрачков. Рассматриваются возможные алгоритмические решения, обеспечивающие фокусировку изображения виртуальной реальности в соответствии с ожидаемой аккомодацией глаз человека. В качестве основных решений рассматриваются классическое решение фильтрации изображения в соответствии с дефокусировкой, вызванной естественной аккомодацией на заданную дистанцию, и решение, в котором соответствующая фильтрация выполняется с помощью нейросетевых технологий. Рассматриваются преимущества и недостатки предложенных решений. В качестве критерия корректности использовалось визуальное сравнение результатов дефокусировки изображения с решением, полученным методом физически корректного рендеринга с использованием модели человеческого глаза. В качестве основы физически корректного рендеринга использовался метод двунаправленной стохастической трассировки лучей с обратными фотонными картами. В работе представлен анализ преимуществ и недостатков предложенных решений.

Ключевые слова

Конфликт вергенции-аккомодации, виртуальная реальность, нейронные сети, система трекинга глаз, микродисплеи.

Comparison of Neural Network and Circular Filtering Algorithms for Synthesized RGB Images

I.E. Kinev¹, G.V. Gebel¹, D.D. Zhdanov¹, A.D. Zhdanov¹

¹ ITMO University, Kronversky prospect 49, liter A, Saint-Petersburg, 197101, Russia

Abstract

A study of the causes of the conflict of vergence-accommodation of human vision in virtual and mixed reality systems has been conducted. Technical and algorithmic approaches to reduce and eliminate the conflict of vergence-accommodation in virtual reality systems are considered. As a technical solution, an approach was chosen that provides adaptive focusing of the eyepiece of a virtual reality system to the convergence point of a person's eyes, determined by the tracking system of his pupils. Possible algorithmic solutions providing focusing of the virtual reality image in accordance with the expected accommodation of human eyes are considered. As the main solutions, we consider the classical solution of image filtering in accordance with defocusing caused by

ГрафиКон 2022: 32-я Международная конференция по компьютерной графике и машинному зрению, 19-22 сентября 2022 г., Рязанский государственный радиотехнический университет им. В.Ф. Уткина, Рязань, Россия
EMAIL: igorkinevitmo@gmail.com (И.Е. Кинёв); gebel.german@yandex.ru (Г.В. Гебель); ddzhdanov@mail.ru (Д.Д. Жданов); andrew.gtx@gmail.com (А.Д. Жданов)
ORCID: 0000-0003-2929-1203 (И.Е. Кинёв); 0000-0002-5399-4366 (Г.В. Гебель); 0000-0001-7346-8155 (Д.Д. Жданов); 0000-0002-2569-1982 (А.Д. Жданов)



© 2022 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

natural accommodation at a given distance, and a solution in which the corresponding filtering is performed using neural network technologies. The advantages and disadvantages of the proposed solutions are considered. As a criterion of correctness, a visual comparison of the results of image defocusing with the solution obtained by physically correct rendering using a human eye model was used. The method of bidirectional stochastic ray tracing using backward photon maps was used as the basis for physically correct rendering. The paper presents an analysis of the advantages and disadvantages of the proposed solutions.

Keywords

Conflict of convergence-accommodation, virtual reality, neural networks, eye tracking system, microdisplays.

1. Введение

Человек обладает стереоскопическим зрением и воспринимает окружающий мир объемным. Для восприятия объема существуют два основных стимула человеческого зрения: конвергенция, сведение глаз на точку наблюдения, что приводит к соответствующему диспаритету между изображениями левого и правого глаз, и аккомодация, фокусировка взгляда на точке конвергенции. Оба эти стимула взаимосвязаны и находятся в соответствии, то есть при изменении направления взгляда происходит аккомодация на точку наблюдения и уточнение конвергенции глаз на точку наблюдения. При нарушении соответствия между вергенцией и аккомодацией человеческого зрения возникает дискомфорт и утомление, причиной которого является попытка мозга устранить данный конфликт и адаптироваться к некорректным условиям наблюдения.

Большинство современных систем виртуальной и смешанной реальностей имеют в качестве источника изображения виртуального мира плоский экран, что приводит к формированию изображения плоского мира в глазах пользователя данных устройств. Длительное использование устройств виртуальной и смешанной реальностей может вызывать дискомфорт зрительного восприятия, проявляющийся в виде рези в глазах, спазма мышц, двоения изображения, тошноты, дезориентации и т.п. Никакие программные решения не могут устранить данное рассогласование и для решения данной проблемы необходимо совместное применение программно-аппаратных решений: трекинг направления взгляда, адаптация оптики окуляра на точку конвергенции и формирование пары стереоизображений с корректной конвергенцией и аккомодацией на точку конвергенции.

Проблеме согласования вергенции и аккомодации посвящено большое число исследований. В работе [1] автор описал определенные проблемы, которые возникают при использовании систем виртуальной реальности. Первая заключается в том, что при формировании виртуального изображения частично, а не полностью, учитывается движение зрачков глаз, которые смещаются при изменении направления взгляда. На рисунке 1 (а), представленном в работе [1], можно видеть, что видимое положение точек А и В, находящихся на разных объектах сцены, зависит от ориентации центров зрачков С и С'. Данные точки могут проецироваться в одну точку на сетчатке глаза или на две. Однако на рисунке 1 (б) показано, что если задана статическая плоскость носителя, то точки А и В проецируются не в две точки на сетчатке глаза, а в одну, что вводит дополнительный дискомфорт при просмотре изображения на плоском экране.

Второй проблемой является описанный выше конфликт вергенции-аккомодации. При фокусировке на объекте сцены или на точке в пространстве существует зависимость между углом конвергенции глаз человека и их аккомодацией. Эта зависимость проявляется в том, что при изменении угла вергенции происходит изменение аккомодации, которая подстраивается под новые параметры фокусировки.

Однако, если данная зависимость нарушается, возникает дискомфорт, называемый конфликтом вергенции-аккомодации. На рисунке 2 показан пример, демонстрирующий природу возникновения данной проблемы: при смещении виртуальной плоскости аккомодации, она уже не соответствует конвергенции глаз наблюдателя.

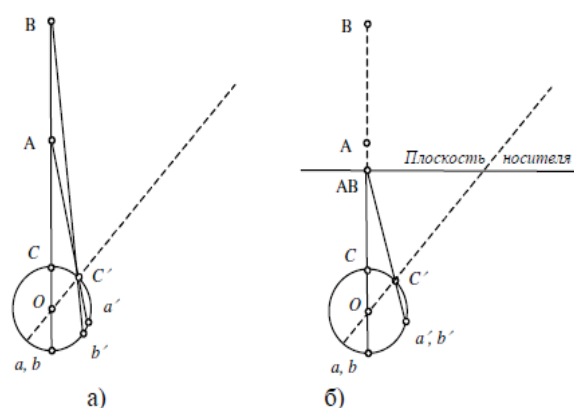


Рисунок 1 – Влияние оптического центра глаза на видимость предметных точек в пространстве и на проекции (рисунок заимствован из работы [1])

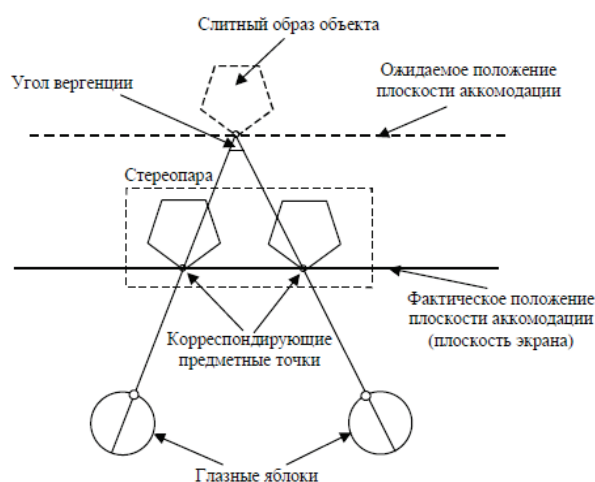


Рисунок 2 – Несоответствие угла вергенции и положения плоскости аккомодации (рисунок заимствован из работы [1])

Комфортным изображением для глаз будет ситуация, когда размытие и двоение ближних объектов происходит при наблюдении дальних объектов, а размытие и двоение дальних объектов происходит при наблюдении близлежащих объектов. При этом диспарат и размытие левого и правого изображений должны быть согласованы с естественной вергенцией-аккомодацией наблюдателя. Данный конфликт и способ его устранения представлен на рисунке 3.

В большинстве случаев для смешанной и виртуальной реальностей используются близкие к глазу дисплеи (near-eye displays - NEDs), которые формируют два изображения, для левого и правого глаза соответственно. С помощью соответствующих окуляров человек воспринимает изображения как одно и видит реалистичную и объемную сцену. На данный момент существуют прототипы устройств виртуальной реальности с соответствующими аппаратными возможностями, которые обеспечивают автоматическую фокусировку окуляров в соответствии с состоянием вергенции глаз человека [2].

В статьях [3-5] авторы предлагают использовать ретинальные дисплеи (дисплеи с прямой проекцией на сетчатку глаза) для решения конфликта вергенции-аккомодации. Преимущество данных дисплеев заключается в том, что они имеют возможность непосредственно проецировать трехмерную сцену на сетчатку глаза и автоматически устраняют конфликт вергенции-аккомодации. Однако данные устройства достаточно сложны и требуют специальной системы рендеринга для создания реалистичного изображения на экране дисплея.

В статье [6] авторы предложили компактный голографический дисплей для устройств виртуальной и смешанной реальностей, основанный на фазовой голографической проекции.

Кроме того, нововведением являются томографические дисплеи [7, 8]. Результаты работы последних показан на рисунке 4.

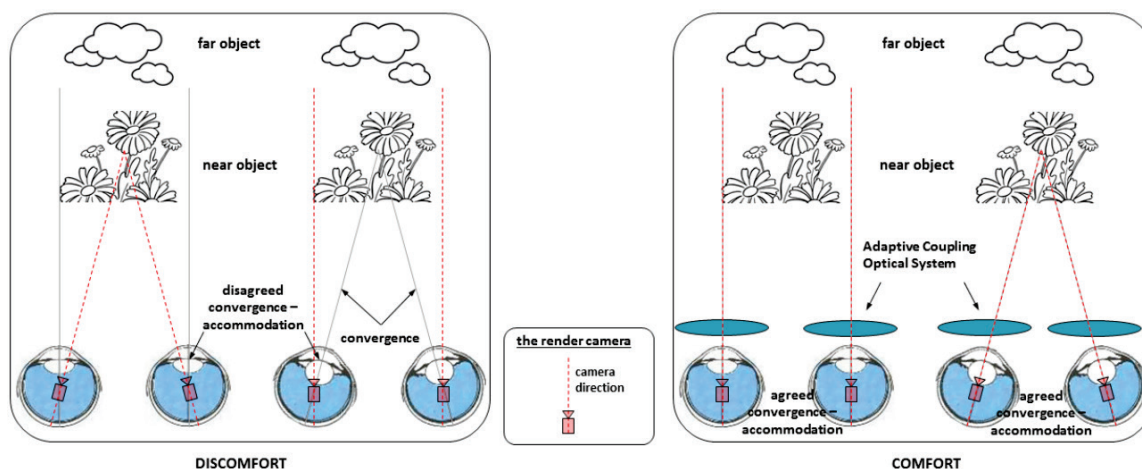


Рисунок 3 – Условия дискомфорта и комфорта зрительного восприятия в системе виртуальной реальности

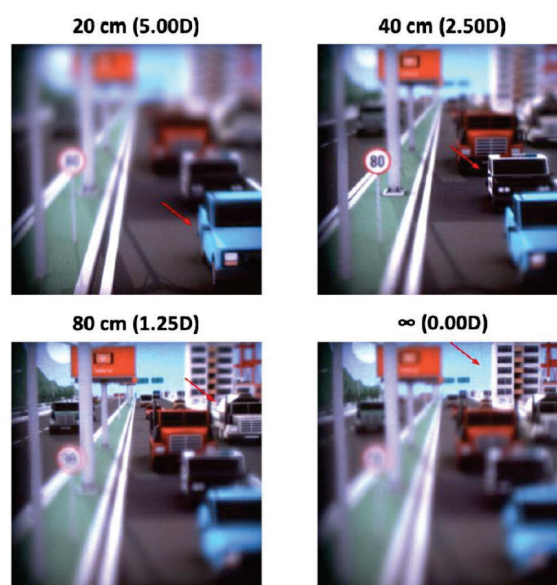


Рисунок 4 – Результат работы томографических дисплеев

Видно, что операция дефокусировки изображения работает недостаточно корректно. Основные проблемы, видимые на рисунке 4: на левом верхнем изображении фокусировка производится на синюю машину, но в плоскости аккомодации синей машины также находится знак дорожного движения, ограничивающий максимальную скорость передвижения, который размыт, и становится четким на правом верхнем изображении, когда фокусировка осуществляется на полицейскую машину. Также автомобили, находящиеся на заднем плане, размываются сильнее при фокусировке на бесконечность.

Большинство методов, представленных в работах [3-8] находится еще на стадиях разработки, поэтому проблему конфликта вергенции-аккомодации предполагается решать для более традиционных устройств, основанных на технологии плоских LCD дисплеев. Для определения параметров конвергенции предполагается использовать систему трекинга зрачков, которая должна формировать соответствующие сообщения адаптивному окуляру для фокусировки на плоскость конвергенции и системе рендеринга для формирования пары изображений, сформированных для камер, направленных в точку конвергенции с дефокусировкой, соответствующей данной конвергенции и размеру зрачков глаз человека. В данной работе

основное внимание уделяется классическим и нейросетевым методам формирования дефокусированного изображения для заданной точки конвергенции и размеру зрачков глаз человека.

Классические методы дефокусировки изображения описаны в работе [9]. Наиболее физически корректным решением является расчет дефокусировки с использованием модели линзовой камеры и трассировки лучей. Основным недостатком данного решения является его трудоёмкость и невозможность реализации в режиме реального времени (даже в условиях распределённых вычислений), что необходимо для работы VR/AR систем. Для формирования физически корректной глубины резкости в режиме реального времени можно использовать метод буфера накопления, который обычно используется для накопления нескольких изображений при рендеринге в реальном времени. Каждое отдельное изображение формируется лучами, проходящими через разные точки апертуры линзовой камеры. После слияния изображений формируется физически корректная дефокусировка. Однако, для качественной дефокусировки требуется значительное число кадров, что затрудняет работу алгоритма в реальном времени. Следующим методом является метод многослойной глубины резкости, который применяется для простых двухполовинной мерных сцен и при редактировании фотографий. Идея этого метода заключается в том, чтобы разделить 2D сцену на слои, которые не перекрываются по глубине, и произвести размытие каждого слоя на основе некоторой репрезентативной глубины в пределах диапазона слоя. Далее размытые слои комбинируются в окончательное изображение и создается эффект глубины резкости. К недостаткам данного метода можно отнести проблему правильного разделения сцены на слои, что приводит к появлению артефактов на изображении, и отсутствие физически корректной основы данного метода.

Наиболее естественное решение – это использование прямого метода фильтрации с ядром фильтра, соответствующего функции рассеивания точки для дефокусировки. Для идеальной линзы данная функция рассеивания точки является кругом, радиус которого зависит от величины дефокусировки. Карта глубин сохраняется вместе с изображением и радиус размытия каждого пикселя определяется по величине дефокусировки из карты глубин. Данный метод не является абсолютно корректным, поскольку он не учитывает апертуру линзовой камеры и то, что яркость пикселя может формироваться на разных глубинах сцены. Данный метод имеет ряд модификаций и может работать как с изображением и картой глубин, так и быть интегрирован в модель рендеринга, которая вместе с изображением формирует карту глубин. Таким образом, ни один метод на данный момент не может удовлетворить всем требованиям качества и скорости работы.

В работе [10] авторы предложили метод вычисления глубины резкости с использованием уравнения теплопроводности для расчета корректного эффекта дефокусировки изображения в режиме реального времени. Авторы утверждают, что данный метод обеспечивает одновременно высокое качество и работу в реальном времени.

В рамках данного исследования предполагается сравнить два метода дефокусировки изображений, которые могут быть применены для систем виртуальной реальности, свободных от конфликта вергенции-аккомодации человеческого зрения.

2. Принцип построения системы виртуальной реальности, свободной от конфликта вергенции-аккомодации

При использовании алгоритмического или нейросетевого подхода для создания условий естественной аккомодации система виртуальной реальности приобретает вид, представленный на рисунке 5.

Данная система должна уметь определять направление взгляда и точку, на которую смотрят глаза наблюдателя. Эта точка необходима для определения значения аккомодации зрения. Важным моментом является определение размера зрачка, поскольку он является апертурной диафрагмой оптической системы глаза и определяет величину дефокусировки изображения. Далее параметры (карта глубин, диаметр зрачка, фокусное расстояние оптической системы глаза и его задний отрезок) поступают в алгоритм фильтрации изображения или нейронную сеть для

получения изображения с корректной аккомодацией. Затем оптическая система очков системы виртуальной реальности должна выполнить фокусировку на значение аккомодации глаз человека. Для фокусировки могут использоваться различные решения от простейшего, например, смещения LCD дисплея, до сложных, связанных с применением динамических фазовых модуляторов или жидкокристаллической градиентной оптики. В результате дефокусированное изображение на LCD дисплее оказывается в фокусе глаз наблюдателя и воспринимается без дефокусировки. Это приводит к естественному восприятию сцены, поскольку то, что находится в фокусе глаз, сфокусировано, а ближние и дальние планы по отношению к точке фокусировки соответствующим образом размыты.

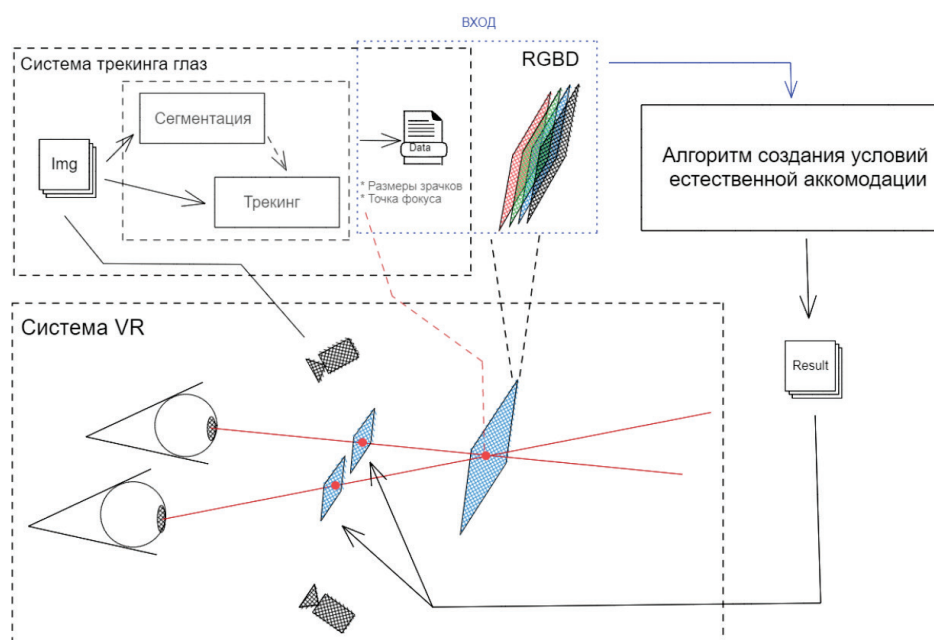


Рисунок 5 – Система для создания условий естественной аккомодации

Первая задача заключается в отслеживании направления взгляда, т.е. в определении, на какую точку изображения направлен взгляд, и сегментации различных частей глаза, таких как радужка, зрачок, склера, поскольку их размеры важны для реалистичной аккомодации. Решением данной задачи занимались исследовательские группы и корпорации, производящие устройства виртуальной реальности. Например, для решения этой задачи компания META создала соревновательный конкурс «OpenEDS Semantic Eye Segmentation Challenge», где был предоставлен достаточно большой набор данных для сегментации различных областей глаза, пример которых показан на рисунке 6. В результате было предложено решение данной задачи с помощью нейронной сети EyeNet, которая позволяет достичь высокую точность сегментации элементов глаз [11] (рисунок 6).

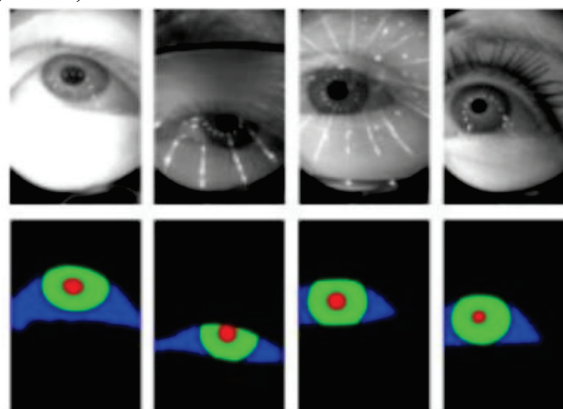


Рисунок 6 – Датасет ОунокпенEDS Semantic Eye Segmentation Challenge

Основные характеристики точности данного сетевого решения: основная метрика pixel wise ассигасы (РА) находится на уровне 99.3%. Точность сегментирования отдельных участков, например зрачка, достигает 97%. Сегментация частей глаза является важной задачей, поскольку она позволяет определить не только направление взгляда, но и размер зрачка, который также определяет радиус размытия изображения.

Для демонстрации зависимости радиуса размытия изображения от размера человеческого зрачка была построена модель сцены, представляющая собой комнату с кубом, имеющим текстуру в виде шахматной доски. Глаз сфокусирован на переднюю грань куба. Рассмотрены два случая: диаметр зрачка глаза 3 и 6 мм. Пример физически корректного рендеринга, выполненного для двух диаметров зрачка глаза, показан на рисунке 7 и демонстрирует зависимость размытия изображения от размера зрачка.

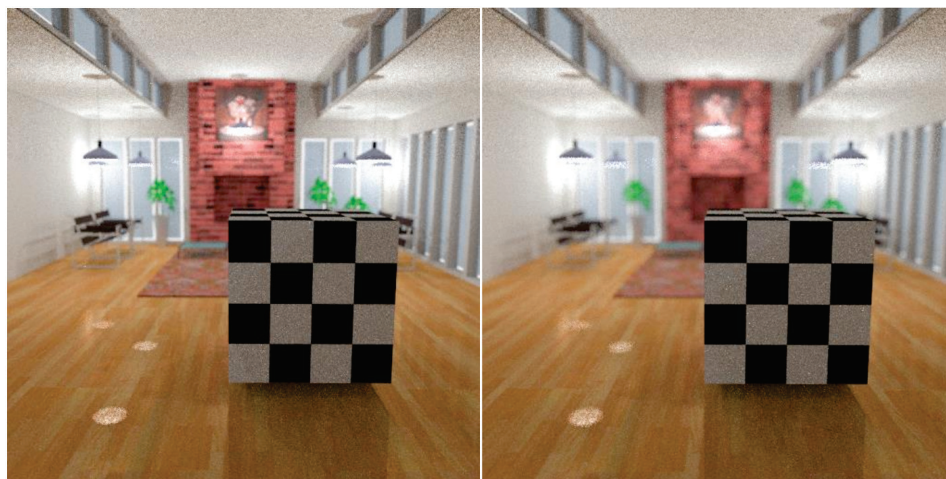


Рисунок 7 – Зависимость размытия изображения от диаметра зрачка глаза, на левом изображении диаметр зрачка глаза равен 3 мм, на правом он равен 6 мм

Определение размера зрачка это только одна задача, более актуальная проблема – определение направления взгляда. Этой проблеме посвящена работа [12]. Идея достаточно проста. Глаз человека имеет отражательную способность, и если его осветить небольшим инфракрасным диодом (не воспринимаемым человеческим зрением), то можно построить модель глаза и определить направление его взгляда по ключевым точкам. Это позволяет восстановить все параметры, необходимые для корректной фокусировки на заданную точку сцены: направление взгляда и размер зрачка. Проблемы с миопией человеческого зрения могут быть решены с помощью соответствующей подвижки окуляров.

3. Алгоритм круговой фильтрации

Для реализации данного алгоритма необходимы следующие данные:

- RGB изображение объектов виртуального мира (формируется программой рендеринга системы виртуальной реальности).
- Карта глубин сцены (формируется программой рендеринга системы виртуальной реальности).
- Фокусное расстояние глаза (внутренний параметр системы виртуальной реальности).
- Задний отрезок глаза (внутренний параметр системы виртуальной реальности).
- Точка фокусировки (формируется системой трекинга зрачка).
- Диаметр зрачка (формируется системой трекинга зрачка).

Данных параметров достаточно для реализации фильтра дефокусировки. Фильтр дефокусировки представляет собой круговой фильтр с постоянным значением ядра внутри круга рассеивания. Радиус круга рассеивания определяется простыми геометрическими

соотношениями по отклонению точки объекта от плоскости фокусировки. Рисунок 8 иллюстрирует принцип определения радиуса рассеивания.

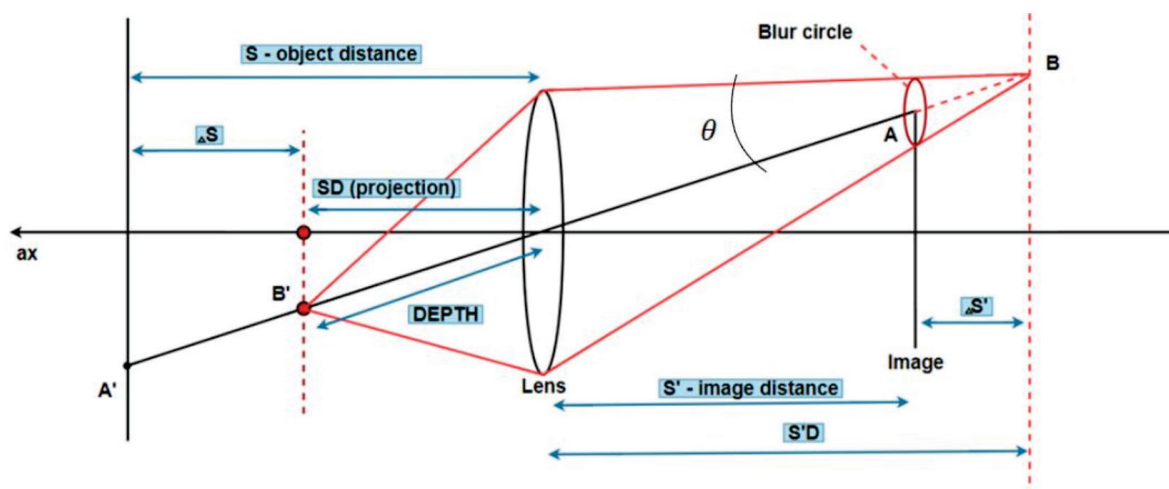


Рисунок 8 – Принцип формирования пятна рассеивания при возникновении дефокусировки

Основная проблема физически корректной и эффективной фильтрации — это постоянство ядра фильтра внутри круга, а не внутри ячейки, что приводит к необходимости вычисления ядра фильтра для каждой точки дефокусируемого изображения. Рисунок 9 иллюстрирует данную проблему.

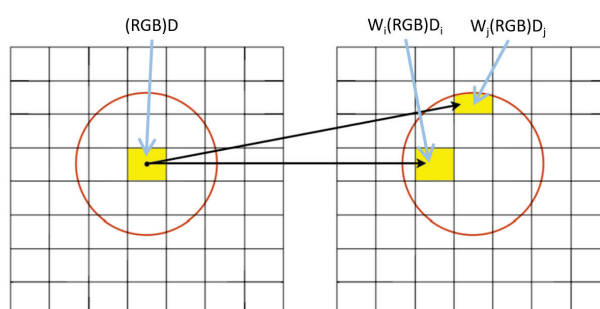


Рисунок 9 – Адаптивный расчет ядра фильтра дефокусировки

Поскольку ядро фильтра осесимметрично, то можно вычислять только одну четверть ядра фильтра, что позволяет значительно ускорить процесс фильтрации изображения. Кроме того, для ускорения фильтрации можно предварительно вычислить и сохранить ядра фильтра для ряда значений радиусов фильтрации, например с шагом 0.1 размера пикселя до радиуса 10 пикселей и далее с шагом 1 пиксель (для фильтрации с радиусом более 10 пикселей нет смысла в точном расчете ядра). Важным моментом ускорения и повышения качества фильтра дефокусировки является анализ точек, на которые допустимо рассеивание. Если точка, на которую осуществляется рассеивание, лежит перед точкой, создающей рассеивание, то рассеивание на эту точку не осуществляется. Это в определенной мере соответствует реальной природе дефокусировки, дальний фон не может размываться на ближний. Например, если на рисунке 9 $D_i < D * k$, где k некоторая константа, определяющая допуск на близость объектов, то рассеивание на точку i осуществляться не будет. Все это позволяет значительно ускорить процесс фильтрации и приблизить качество фильтруемого изображения к качеству изображения, реально воспринимаемому наблюдателем.

Используя алгоритм фильтрации, удалось достичь качества работы фильтра практически неотличимого от физически корректного рендеринга с моделью человеческого глаза. Рисунок 10 иллюстрирует полученный результат. Фокусировка осуществлялась на крышку переднего чайника, находящегося на расстоянии 570 мм от глаз наблюдателя. Поле зрения наблюдателя 10 градусов. Левое изображение показывает то, что видит наблюдатель. Это изображение было получено программой физически корректного рендеринга Lumicept [14], при визуализации

использовалась модель глаза со зрачком 3 мм и фокусным расстоянием 23 мм. Правое изображение – это изображение, видимое через реальный окуляр оптической системы виртуальной реальности. То есть виртуальное изображение было дефокусировано, спроецировано (виртуально) через окуляр на LCD, а затем наблюдалось через этот окуляр.

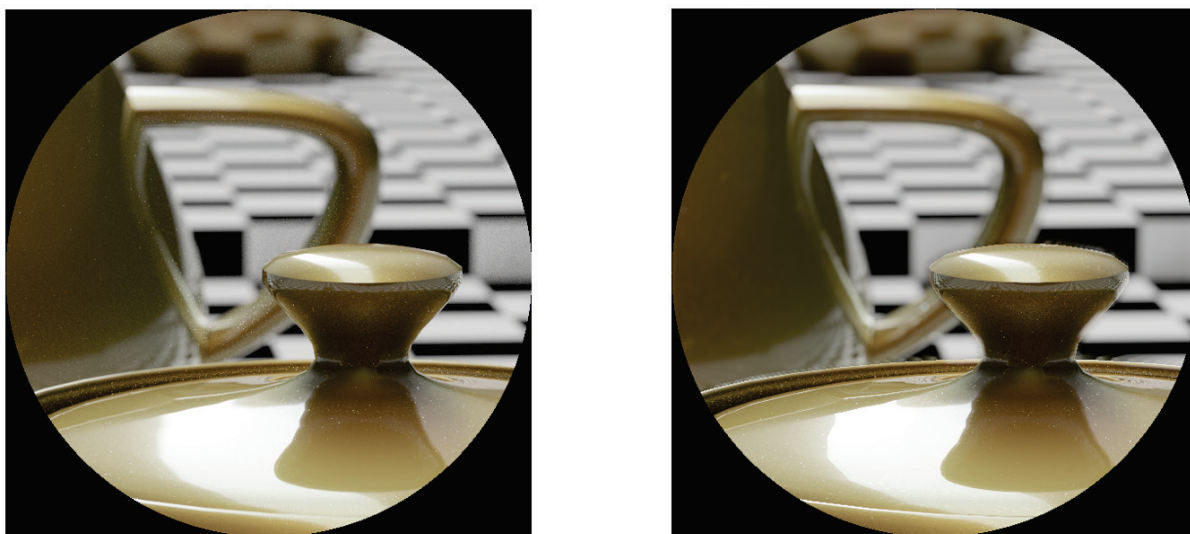


Рисунок 10 – Ожидаемое изображение (слева), реально наблюдаемое изображение в системе виртуальной реальности (справа)

Необходимо отметить, что алгоритм предполагает размытие яркости на заданном расстоянии, что требует корректного определения расстояния до объекта дефокусировки. Например, если мы видим изображение в зеркале или через стекло, то расстояние дефокусировки должно быть вычислено не до зеркала (стекла), а до объекта, формирующего реальную яркость. Такое решение не представляет проблем для систем виртуальной реальности, поскольку сцена виртуальна и любое расстояние может быть вычислено корректно. Однако для систем смешанной реальности данное решение представляет проблему, поскольку RGB-D датчик или нейросеть могут не разобрать оптические характеристики наблюдаемых объектов, что приведет к ошибочной дефокусировке.

Другая проблема, напрямую относящаяся к системам виртуальной реальности, – это множественные источники яркости для одного направления взгляда. Такая ситуация возникает, например, при отражении от стекла. Часть света отражается от стекла, а часть проходит непосредственно через стекло. Это формирует два источника яркости, находящихся на разных расстояниях от наблюдателя, что требует применения двух фильтров в одной точке для двух компонент яркости. Данное решение возможно, однако достаточно трудоемко и требует дополнительных ресурсов, что сказывается на производительности работы системы дефокусировки. Поэтому в данной реализации многозначность компонент яркости сводится к суммарной яркости и дистанции до ее компоненты с максимальным значением.

4. Алгоритм нейросетевой фильтрации

Одну из самых значимых работ по нейросетевым методам фильтрации проводила компания многопользовательской социальной сети. В работе [13] авторы предлагают архитектуру нейронной сети DeerFocus, показанную на рисунке 11.

Данная нейронная сеть принимает на вход изображение с картой глубин, с помощью которой определяется расстояние до объектов сцены. Далее по заданной точке на изображении и ее глубине формируется эффект размытия, который создает условия, близкие к естественной аккомодации человеческого глаза. Пример работы данной сети представлен на рисунке 12. Левый рисунок – исходное изображение с картой глубин, следующие два изображения получены

с фокусировкой на дальнюю дистанцию (0.1 D) и ближнюю дистанцию (3.0 D). Размер зрачка глаза не определен (предположительно 3 мм).

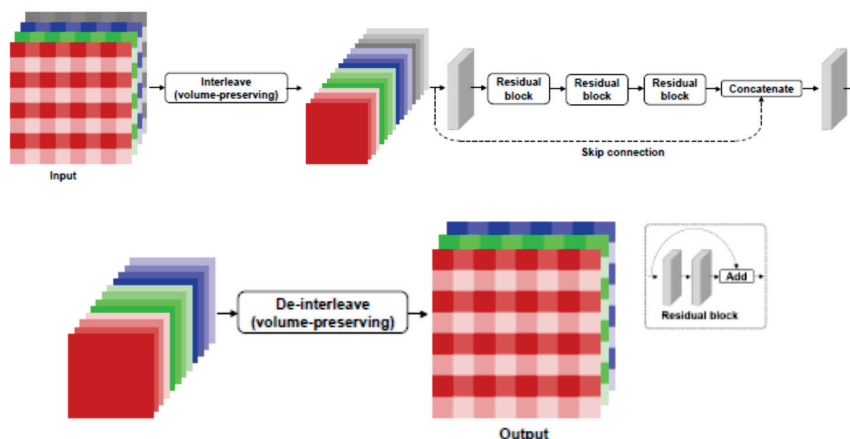


Рисунок 11 – Архитектура нейронной сети DeepFocus

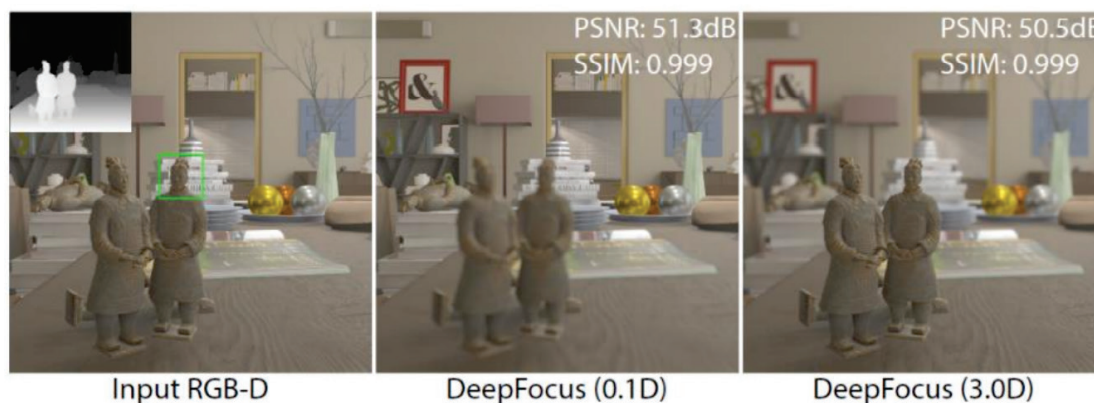


Рисунок 12 – Создание условий естественной аккомодации с помощью нейронной сети Deep Focus

В работе [14] авторы показали, что унаследованная ими данная архитектура глубокой нейронной сети очень сложна и дорогостояща для обучения. Ими было предложено использовать сети с так называемыми связными блоками и при этом не увеличивать глубину нейронной сети. При запуске работы нейронной сети запускается виртуальное окружение, которое подгружает модель, с помощью которой будут выполняться преобразования, и далее нейронная сеть рассчитывает размытие. Эксперименты с данной сетью показали, что время запуска виртуального окружения 15 секунд, время размытия изображения с разрешением 1024x1024 равна примерно 1.2 секунды, хотя авторы заявляют о времени работы в пределах 10 мс. Такое замедление может быть связано с низкой производительностью графического ускорителя, который использовался в работе, и полученные результаты говорят о том, что на данном оборудовании нейронная сеть не может работать в режиме реального времени.

В нейросетевом обучении существует специальная практика – первоначальное обучение на большом наборе данных, а затем дообучение (fine tuning) [15] на объеме данных меньшего размера, но более высокого качества. Большой набор данных формирует базу знаний, с помощью которой модель, обученная на них, способна извлекать более общие функции, параметры изображения, в нашем случае различные зависимости, которые помогут более точно создавать условия естественной аккомодации.

Для создания набора данных дообучения использовался программный комплекс Lumiscept [16]. Используя его API, была построена анимационная модель рендеринга, в которой камера

двигалась по сцене и формировала изображения для разных ориентаций и параметров камеры (диаметр зрачка и дистанция фокусировки). Данный подход проиллюстрирован на рисунке 13.

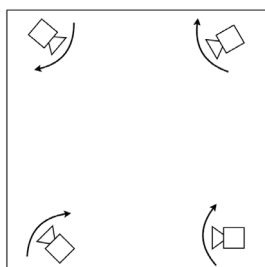


Рисунок 13 – Метод увеличения набора данных

В результате был сформирован дополнительный набор данных, состоящий из 1800 изображений.

5. Результаты

В данном исследовании было произведено качественное сравнение результатов работы двух алгоритмов фильтрации: кругового алгоритма фильтрации с адаптивным радиусом и нейросетевого алгоритма. Сравнение происходило на синтезированных сценах, поскольку данные алгоритмы должны применяться в системах виртуальной реальности, где сцены и их изображения всегда являются синтезированными. Тестирование на изображениях реальных сцен имеет смысл проводить для видео-прозрачных систем смешанной реальности, где изображение реального мира и карта его глубин формируются 3D камерами этих систем. Кроме того, сравнение алгоритмов происходило на качественном уровне, где основными факторами сравнения являлось то, что, во-первых, изображение в плоскости фокусировки должно быть четким, во-вторых, чем дальше объект находится от плоскости фокусировки, тем сильнее его изображение размыто. В качестве эталонной величины размытия изображения предполагается использовать значение размытия алгоритма круговой фильтрации, поскольку на простых сценах, имеющих один видимый источник яркости, данный алгоритм показал физически корректное размытие изображения, представленное, например на рисунке 10. Поэтому исследование в большей степени было ориентировано на оценку качества работы нейросетевого фильтра, т. е. сможет ли этот фильтр обеспечить корректную дефокусировку изображения.

Для создания нейросетевого фильтра была реализована и запущена нейронная сеть DeepFocus [17]. Для ее запуска был произведен рефакторинг ее модели на фреймворк Tensorflow последней версии. Для исследования качества работы нейросетевого фильтра было выполнено сравнение разработанного алгоритма круговой фильтрации изображения с нейросетевым решением. Для тестирования использовалось синтезированное изображение сцены с роботами, содержащее карту глубин. Разрешение изображения 1024x1024. В соответствии с картой глубин ближний робот находился на расстоянии 1.4 м от камеры, дальний робот – на 6.5 м от камеры. Взгляд наблюдателя был сфокусирован на ближнем роботе. Для оценки методов фильтрации были применены оба метода с максимально близкими настроечными параметрами. В методе круговой фильтрации размер зрачка был выбран равным 3мм, а фокусное расстояние глаза 23 мм. В нейросетевом методе размер зрачка установить не удалось (предположительно был 3 мм). Алгоритмы были реализованы на компьютере с процессором Ryzen 5 5600h и видеокартой Nvidia gtx 3050, RAM : 16 GB. Результат работы обоих методов представлен на рисунке 14.

Оба фильтра показывают похожую фокусировку. Ближний робот находится в фокусе и его изображение практически не размыто. Увеличенные изображения данной области представлены на рисунке 15.

Видно, что нейросетевое решение создает незначительное размытие изображения, однако с точки зрения зрительного восприятия это выглядит более естественно, поскольку позволяет убрать алиасинг изображения в методе круговой фильтрации. Нейросетевое решение показывает более сильную дефокусировку, что соответствует либо большему размеру зрачка

глаза, либо большему фокусному расстоянию оптической системы глаза. Визуально дефокусировка в нейросетевом методе выглядит неестественно сильной, что в принципе может вызывать дискомфорт зрительного восприятия. Увеличенные изображения области с сильной дефокусировкой представлены на рисунке 16. Данная область находится за плоскостью фокусировки.

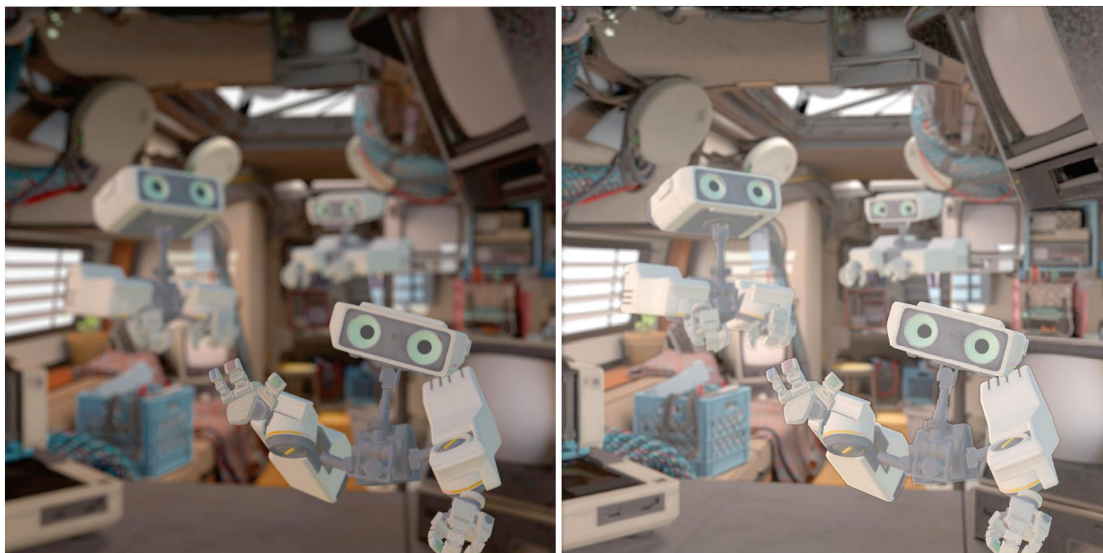


Рисунок 14 – Результат работы нейросетевого фильтра (слева) и адаптивного кругового фильтра (справа)



Рисунок 15 – Область изображения вблизи плоскости фокусировки, результат работы нейросетевого фильтра (слева) и адаптивного кругового фильтра (справа)



Рисунок 16 – Область изображения за плоскостью фокусировки, результат работы нейросетевого фильтра (слева) и адаптивного кругового фильтра (справа)

Поскольку метод круговой фильтрации обеспечивает физически корректное размытие изображения (по крайней мере для такой простой сцены), можно заключить, что именно нейросетевое решение не дает корректного размытия и его достаточно сложно настроить на текущие параметры глаза наблюдателя.

В данном примере фокусировка осуществлялась на ближний объект сцены. При фокусировке на дальний объект сцены происходит размытие ближних объектов, что представлено на рисунке 17.



Рисунок 17 – Область изображения перед плоскостью фокусировки, результат работы нейросетевого фильтра (слева) и адаптивного кругового фильтра (справа)

Эффект «заглядывания» за дефокусированный объект отсутствует в алгоритме круговой фильтрации. Некоторое присутствие этого эффекта можно наблюдать в нейросетевом алгоритме, хотя его причина скорее в размытии дальнего объекта (более сфокусированного) на ближний объект. С точки зрения оптики (эффект дифракции не рассматривается), «заглядывание» за размытый объект на сфокусированный невозможно. Визуально эффект «заглядывания» может объясняться только сильным размытием объекта, находящегося вдали от плоскости фокусировки.

Стоит отметить, что скорость нейросетевого решения на данном примере составила 1.2 с на кадр, что значительно меньше, чем скорость работы алгоритма круговой фильтрации, которая составила 3.8 с на кадр. Однако, перевод нейросетевого решения на высокопроизводительное GPU позволит значительно ускорить процесс фильтрации и достичь 10 мс, которые декларировались в работе [12]. С другой стороны, реализация алгоритма круговой фильтрации на высокопроизводительном GPU также имеет шанс перевести работу данного алгоритма в режим реального времени. Поэтому при выборе метода фильтрации необходимо делать выбор между высокой производительностью с возможностью потери естественной аккомодации и более низкой производительностью, но с корректной аккомодацией.

6. Заключение

Проведенное исследование дает представление о том, что вызывает конфликт вергенции-аккомодации и какие методы существуют для решения данной проблемы. Были рассмотрены методы трекинга (сегментации) глаз наблюдателя, которые позволяют определить точку фокусировки и размер зрачка глаза. Был разработан адаптивный фильтр круговой фильтрации изображения и рассмотрено нейросетевое решение для реализации фильтра дефокусировки изображения. Исследование показало, что адаптивный фильтр круговой фильтрации обеспечивает качество дефокусировки изображения, сопоставимое с ожидаемым. Однако время работы кругового фильтра дефокусировки несколько выше, чем нейросетевого. С другой стороны, нейросетевой фильтр не так хорошо адаптируется к размеру зрачка глаза человека, что может привести к дискомфорту зрительного восприятия. Кроме того, в ряде случаев нейросетевой фильтр некорректно обрабатывает карту глубин, что также создает определенный дискомфорт зрительного восприятия. В качестве продолжения данной работы планируется адаптация нейросетевого фильтра к размеру зрачка человека и поиск причин некорректной обработки карт глубин изображения. Кроме того, планируется реализация алгоритма адаптивной круговой фильтрации на GPU.

7. Благодарности

Работа была выполнена при финансовой поддержке Российского Научного Фонда, грант № 22-11-00145.

8. Список источников

- [1] 1. В. О. Афанасьев. Развитие модели формирования бинокулярного изображения виртуальной 3D-среды. // Программные продукты и системы. 2004. №4. С. 27–32.
- [2] Byoungcho Lee, Youngjin Jo, Dongheon Yoo, Juhyun Lee. Recent progresses of near eye display for AR and VR // SPIE Optical Metrology. 2021. V. 11785. P. 1–6.
- [3] Jeong, J., Lee, J., Yoo, C., Moon, S., Lee, B., Lee, B. Holographically customized optical combiner for eye-box extended near-eye display. // Opt. Express. 2019. V. 27. P. 38006–38018.
- [4] Jo, Y., Yoo, C., Bang, K., Lee, B., Lee, B. Eye-box extended retinal projection type near-eye display with multiple independent viewpoints. // Appl. Opt. 2021. V. 60. P. 268–276.
- [5] Jang, C., Bang, K., Li, G., Lee, B. Holographic near-eye display with expanded eye-box. // ACM Trans. Graph. 2018. V. 37. P. 1–14.
- [6] Maimone, A., Georgiou, A., Kollin, J. S. Holographic near-eye displays for virtual and augmented reality. // ACM Trans. Graph. 2017. V. 36. P. 1– 16.
- [7] Lee, S., Jo, Y., Yoo, D., Cho, J., Lee, D., Lee, B. Tomographic neareye displays. // Nature Comm. 2019. V. 10. P. 2497–2507.
- [8] Dongheon Yoo, Seungjae Lee, Youngjin Jo, Jaebum Cho, Suyeon Choi, Byoungcho Lee, 15 focal planes head-mounted display using LED array backlight. // Proc. SPIE, Optical Design Challenge. 2019.
- [9] Hubert Nguyen. GPU Gems 3 // Addison-Wesley Professional. 2007. Chapter 23. Depth of Field: A Survey of Techniques.
- [10] Kass, Michael & Lefohn, Aaron & Owens, John. (2006). Interactive Depth of Field Using Simulated Diffusion on a GPU. Pixar Animation Studios Tech Report. 2.
- [11] P. Kansal and S. Devanathan, EyeNet: Attention Based Convolutional Encoder-Decoder Network for Eye Region Segmentation, // IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). 2019. P. 3688-3693.
- [12] Zhengyang Wu, Srivignesh Rajendran, Tarrence van As, Joelle Zimmermann, Vijay Badrinarayanan, Andrew Rabinovich. EyeNet: A Multi-Task Network for Off-Axis Eye Gaze Estimation and User Understanding. – URL: <https://arxiv.org/abs/1908.09060> (дата обращения: 05.05.2022).
- [13] DeepFocus: Learned Image Synthesis for Computational Display. – URL: <https://clck.ru/oBFwW> (дата обращения: 05.05.2022).
- [14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun; Deep Residual Learning for Image Recognition. // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 770-778.
- [15] Fine-Tuning. - URL: http://d2l.ai/chapter_computer-vision/finetuning.html (дата обращения: 05.05.2022).
- [16] Lumiccept Integra [Электронный ресурс]. Режим доступа – URL: <https://integra.jp/en/products/lumiccept>.
- [17] Ковалев В.В., Сергеев Н.Е. Реализация сверточных нейронных сетей на встраиваемых устройствах с ограниченным вычислительным ресурсом // Известия ЮФУ. Технические науки. 2021. №6.