

Федеральное государственное автономное образовательное
учреждение высшего образования
**«МОСКОВСКИЙ ФИЗИКО-ТЕХНИЧЕСКИЙ ИНСТИТУТ
(НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)»**

На правах рукописи

Таамазян Ваге Арамаисович

**ИССЛЕДОВАНИЕ И РАЗРАБОТКА МЕТОДОВ ТРЕХМЕРНОГО
СКАНИРОВАНИЯ И СЕГМЕНТАЦИИ ОБЪЕКТОВ С
ИСПОЛЬЗОВАНИЕМ ПОЛЯРИЗАЦИИ СВЕТА**

Специальность 05.13.01 –
«Системный анализ, управление и обработка информации
(информационные и технические системы)»

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата технических наук

Научный руководитель:
доктор технических наук, г.н.с.
Аведьян Эдуард Дзиронович

Москва – 2019

Работа прошла апробацию на кафедре «Интеллектуальные информационные системы и технологии» Федерального государственного автономного образовательного учреждения высшего образования «Московский физико-технический институт (национальный исследовательский университет)»

Научный руководитель:

доктор технических наук,
главный научный сотрудник,
Аведьян Эдуард Дзержонич

Ведущая организация: Федеральное государственное унитарное предприятие «Всероссийский научно-исследовательский институт оптико-физических измерений».

Защита состоится 25.12.2019 в 13:00 на заседании диссертационного совета ФРКТ.05.13.01.009 по адресу 141701, Московская область, г. Долгопрудный, Институтский переулок, д. 9.

С диссертацией можно ознакомиться в библиотеке и на сайте Московского физико-технического института (национального исследовательского университета)
<https://mipt.ru/education/post-graduate/soiskateli-tekhnicheskie-nauki.php>

Работа представлена «15» октября 2019 г. в Аттестационную комиссию федерального государственного автономного образовательного учреждения высшего образования «Московский физико-технический институт (национальный исследовательский университет)» для рассмотрения советом по защите диссертаций на соискание ученой степени кандидата наук, доктора наук в соответствии с п. 3.1 ст. 4 Федерального закона «О науке и государственной научно-технической политике»

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы диссертации. Поляризация – неотъемлемое свойство любых электромагнитных волн, включая также видимый свет. При этом, это свойство незаметно глазу: органы чувств человека, как и большинства других живых существ (за редким исключением типа креветок, осьминогов и ряда других морских существ) не способны улавливать наличие или отсутствие поляризации у поступающего света, не говоря уже о степени поляризации или угле поляризации света. Вероятно, именно поэтому долгое время поляризация крайне неактивно использовалась в исследованиях по вычислительному имаджингу (computational imaging) и компьютерному зрению (computer vision).

Тем не менее, в последние годы интерес к использованию информации о поляризации света в решении различных задач компьютерного зрения постоянно растет. Поляризация помогает в решении множества задач, таких как трехмерное сканирование лица и прочих объектов, включая также восстановление формы поверхности океана, для дехэйзинга изображений, разделении отраженной и рассеянной компонент изображения, и в прочих исследованиях.

Наибольшее число публикаций и исследований относится именно к области трехмерного сканирования объектов, потому что поляризация света несет в себе информацию о форме объекта, от которого свет отразился. Влияние формы объекта на поляризацию отраженного света было известно веками и сформулирована в уравнениях Френеля. Этот принцип лежит в основе методов Формы из Поляризации (ФиП), целью которой является реконструкция информации о трехмерной форме объекта на основе данных о поляризации света, извлеченных из трех и более фотографий объекта под разными углами поляризатора. Одной из целей данной работы было продемонстрировать использование алгоритмов ФиП в сочетании с неточной картой глубин для ее последующего уточнения, а также показать способ использования поляризации для более точного разделения рассеянной и отраженной компонент.

Еще одной областью компьютерного зрения, очень активно развивающейся в последние 5 лет, является сегментация изображений. Задачу сегментации можно разделить на 2 основных направления:

1. Семантическая сегментация - определение всех пикселей для объектов определённого класса или фона на изображении. Если несколько объектов одного класса пересекаются, их пиксели никак не отделяются друг от друга.
2. Сегментация объектов - определение пикселей, принадлежащих каждому объекту каждого класса по отдельности.

С развитием глубинного обучения и сверточных нейронных сетей, полноценное решение этой задачи стало ближе. Однако тем не менее, все еще существует ряд типов объектов, с которыми современные алгоритмы сегментации справляются плохо или не справляются вовсе. Одним из показательных примеров является сегментация прозрачных объектов: стаканов, бутылок, игрушек и так далее. Использование поляризации света для более точного решения этой задачи также рассмотрено в данном диссертационном исследовании, а также приведены результаты, демонстрирующие качественное и количественное улучшение результатов, что также определяет актуальность данной диссертации.

Степень разработанности темы. 3D сканирование на основе поляризации (Shape from Polarization, SFP, Форма из поляризации, ФиП) света базируется на фундаментальном физическом принципе, связывающем ориентацию нормали на поверхности объекта с поляризацией света, отраженного или рассеянного от данной точки объекта. Различные части данного метода были впервые описаны в публикациях Lawrence B. Wolff, Terrance E. Boult, Megumi Saito et al., Stefan Rahmann, Daisuke Miyazaki и других в 1990х и начале 2000х годов.

Как правило, для оценки параметров поляризации отраженного от объекта света используется камера с закрепленным на объективе поляризатором. С помощью этой камеры делается 3 фотографии с разными углами поляризатора, которые затем обрабатываются для реконструкции карты нормалей. Существует 2 основных направления ФиП: на основе отраженного зеркально света, а также на

основе рассеянного света. Первый направление активно используется для оценки формы металлов, прозрачных объектов, темных объектов и даже океанских волн. Второе направление базируется на анализе света, который претерпел рассеяние внутри объекта (возможно, в тонком слое у поверхности), из-за чего уравнения Френеля применяются в ином виде.

Существуют работы по использованию поляризованного света для детализированного трехмерного сканирования лиц. Однако в данных работах используется активная поляризованная подсветка, а также анализ ее отражения от поверхности с нескольких ракурсов. Также ряд работ демонстрирует использование круговой поляризации для трехмерной реконструкции объекта, исправление эффекта задымления в изображении, мультиплексирование освещения, создание панорам, подводное рассеяние света, а также трехмерные дисплеи.

Также в последние годы крайне активно развивается область детектирования и сегментации объектов, в особенности благодаря развитию глубоких сверточных нейронных сетей. Так, в течение последних 3-4 лет вышли публикации Fast/Faster R-CNN, Mask R-CNN, Fully Convolutional Network (FCN), YOLO, SSD, U-Net, а также ряд других статей. Отдельное внимание исследователей обращено на сегментирование и детектирование прозрачных объектов, как более сложный частный случай задачи сегментирования.

Целью исследования является улучшение методов трехмерного сканирования объектов, а также сегментации объектов с помощью поляризации света.

Объектом исследования диссертации являются системы трехмерного сканирования объектов с использованием поляризации, а также системы сегментации объектов на изображениях.

Предметом исследования диссертации являются модели, методы, алгоритмы и программы трехмерного сканирования объектов, а также сегментации объектов на изображении с использованием поляризации света и сверточных нейронных сетей.

Методы исследования. Для решения задач используются методы вычислительной математики, компьютерные методы обработки изображений, сверточные нейронные сети, а также методы разработки приложений на языках Matlab и Python.

Для достижения целей диссертации, а также в соответствии с объектом и предметом диссертации были сформулированы следующие **задачи**:

1. Провести исследование по улучшению ФиП метода трехмерного сканирования объектов.
2. Провести исследование и разработать методы по улучшению сегментации прозрачных объектов с использованием поляризации света.
3. Собрать датасет для последующего использования в исследованиях задачи сегментации другими научными группами.

Научная новизна результатов, полученных в диссертационной работе, состоит в следующем:

1. Предложены способы усовершенствования ФиП метода трехмерного сканирования объектов. Показано, как можно добиться уточнения формы объекта с помощью сшивки карты нормалей, полученных из анализа поляризации отраженного света и карты глубин низкого разрешения. Также предложен метод разделения зеркальной и рассеянной компонент отраженного света для более точной реконструкции в случае смешанного отражения.
2. Собран датасет прозрачных объектов с попиксельной сегментацией, а также информацией о поляризации отраженного от объекта света.
3. Предложен способ более точной и качественной сегментации прозрачных объектов с использованием информации о поляризации света и сверточных нейронных сетей. Показано, как можно работать с поляризационными данными с использованием сверточных нейронных сетей.

Практическая значимость. Результаты данной диссертационной работы могут быть применены в области трехмерного сканирования объектов, а также в

области, где требуется сегментация прозрачных объектов. Результаты диссертационного исследования внедрены в Федеральном государственном автономном научном учреждении "Центр информационных технологий и систем органов исполнительной власти" (ФГАНУ ЦИТиС), что подтверждено актом о внедрении. Кроме того, на результаты и программы исследования зарегистрирован патент США, а также они используются компанией Акаша Имаджинг (США) для разработки роботизированного манипулятора, захватывающего прозрачные объекты.

Обоснованность и достоверность результатов и выводов определяется следующим факторами: строгим доказательством полученных результатов, публикациями результатов исследования в рекомендованных Высшей аттестационной комиссией научных изданиях, практическим использованием результатов диссертационной работы, подтвержденным актом о внедрении.

Личный вклад соискателя состоит в непосредственном участии в разработке моделей, математических методов, алгоритмов и программного обеспечения, проведении исследовательских испытаний, апробации результатов исследования, обработке полученных данных, подготовке основных публикации по выполненной работе.

Апробация результатов работы. Материалы диссертационной работы были доложены и обсуждены на следующих всероссийских и международных конференциях: «Нейрокомпьютеры и их применение», 14 марта 2017, МГППУ, 60-ой Научной Конференции МФТИ, 25 ноября 2017, МФТИ, 61-ой Научной Конференции МФТИ, 25 ноября 2018, МФТИ, ITNT-2019, май 2019, Самара.

Публикации автора по теме диссертации. По теме диссертации опубликовано 6 работ в журналах и трудах конференций и еще 1 работа подана в печать, 3 из них находятся в списке научных изданий, зарегистрированных в Высшей аттестационной комиссии Министерства науки и высшего образования России (в том числе 3 статьи в журналах из перечня RSCI – Web of Science).

Структура и объем диссертации. Диссертационная работа состоит из титульного листа, оглавления, введения, трех глав, заключения, списка

литературы, приложений и акта о внедрении. Основная часть (без приложений) изложена на 88 страницах машинописного текста. Работа содержит 24 рисунка, 5 таблиц, список литературы, которой включает 95 наименований.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обосновывается актуальность темы диссертации, сформулированы цель и задачи, объект и предмет исследования, научная новизна и практическая ценность диссертации, определена степень изученности темы. Также представлены основные научные положения, которые выносятся на защиту, а также сведения об апробации полученных результатов и публикациях автора по теме диссертации.

В первой главе «Аналитический обзор методов и алгоритмов трехмерного сканирования и сегментации объектов с использованием поляризации света» описаны основные и наиболее значимые публикации в области использования поляризации в компьютерном зрении и вычислительном имаджинге, а также наиболее значимые публикации по сегментации объектов. В заключении этой главы делаются выводы по материалам обзора и формулируются задачи, которые должны быть решены в диссертации.

Во второй главе «Совершенствование технологии 3D сканирования с помощью ФиП» содержится:

Алгоритм увеличения точности и детализации карт глубин с помощью информации о поляризации отраженного света.

Согласно уравнениям Френеля, свет, отразившийся от поверхности объекта, частично поляризуется, то есть его можно разделить на 2 компоненты: не поляризованную и поляризованную (с явно выделенной осью колебаний электромагнитной волны). Известно, что свойства отраженного света зависят от:

1. свойств поляризации падающего света,
2. ориентации нормали в точке отражения поверхности,
3. свойств материала поверхности.

Таким образом, в предположении неполяризованного падающего света и известного до определенной точности коэффициента преломления материала отражающей поверхности можно восстановить ориентации нормали поверхности объекта. Отметим здесь 2 ключевых фактора:

1. При определении ориентации нормалей есть ряд неопределенностей, таких как неопределенность азимута на π и неопределенность зенита на $\pi/2$.
2. Ориентация поверхности не определяет абсолютные размеры объекта. Таким образом, с помощью данного метода можно восстановить форму объекта с точностью до масштаба, и нельзя восстановить расстояние до него.

Данные ограничения делают непригодным использование исключительно поляризационного метода для сканирования трехмерных моделей людей и прочих объектов. Однако карта нормалей может быть использована для уточнения неточных карт глубин, полученных с помощью альтернативных методов, что и показано в описанном ниже методе.

Интенсивность света, отраженного от определенной точки объекта и поступившая в пиксель матрицы камеры, пройдя предварительно через поляризатор, может быть выражена в следующем виде:

$$I(\phi_{pol}) = \frac{I_{max} + I_{min}}{2} + \frac{I_{max} - I_{min}}{2} \cos(2(\phi_{pol} - \phi))$$

где ϕ_{pol} – угол оси поляризатора, ϕ – фаза поляризованного света, I_{max} и I_{min} – максимальные и минимальные значения интенсивности света, прошедшего через поляризатор. Таким образом, так как для аппроксимации синусоиды необходимо три уравнения, измерив значение интенсивности при трех разных углах поляризатора, возможно оценить неизвестные величины ϕ , I_{max} и I_{min} . Решение данного уравнения производится в каждом пикселе сформированного на предыдущем шаге стека изображений линеаризацией с помощью замены переменных, а затем методом наименьших квадратов.

После вычисления угла азимута производится вычисление зенита. Оценка угла зенита полагается на величину степени поляризации света, вычисляемой следующим образом:

$$\rho = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}$$

Как и в случае с азимутом, тип отражения влияет на зависимость зенита от степени поляризации света. В случае зеркального отражения зависимость может быть записана следующим образом:

$$\rho = \frac{2 \sin \theta \tan \theta \sqrt{n^2 - \sin^2 \theta}}{n^2 - \sin^2 \theta + \tan^2 \theta}$$

где n - коэффициент преломления материала объекта, а θ - угол зенита нормали к поверхности. Если известны коэффициент преломления материала, а также степень поляризации отраженного света, то можно найти угол зенита, решив уравнение выше.

Для рассеянного отражения уравнение, связывающее данные величины, выглядит несколько иначе:

$$\rho = \frac{(n - \frac{1}{n})^2 \sin^2 \theta}{2 + 2n^2 - (n + \frac{1}{n})^2 \sin^2 \theta + 4 \cos \theta \sqrt{n^2 - \sin^2 \theta}}$$

В данном случае, выбор за основу модели отражения не играет большой роли, так как в дальнейшем производится коррекция угла зенита. Поэтому мы берем за основу модель рассеянного отражения, как более простую для обратного решения при вычислении зенита.

На последнем этапе алгоритма производится сшивка карты глубин, полученной с помощью стерео, активного стерео, времяпролетных методов и прочих методов захвата карт глубин с картой нормалей, полученной с помощью поляризации, с попутной коррекцией карты нормалей. Данная сшивка должна позволить увеличить детализацию карты глубин. Данный этап состоит из нескольких последовательных шагов. Сначала производится коррекция низкочастотной π -неопределенности азимута, с помощью сгенерированной

методом kD деревьев карты нормалей из карты глубин. Вторым шагом производится коррекция высокочастотной π -неопределенности азимута. Затем корректируется искажение карты нормалей поляризации, связанное с погрешностью выбора коэффициента преломления объекта, то есть таким образом выправляется зенит-компонента нормали. И после этого полученная карта нормалей интегрируется вместе с картой глубины в едином оптимизационном алгоритме.

Коррекция низкочастотной π -неопределенности азимута

Пусть $D \in \mathbb{R}^{M \times N}$ – карта глубин объекта. Предложенная схема использования поляризации для уточнения карты глубин оперирует в пространстве нормалей, поэтому из карты глубин вычисляется карта нормалей: $N_{x,y}^{depth} \in \mathbb{R}^{M \times N \times 3}$. Для карты глубин D производится преобразование в облако точек:

$$P_{x,y} = \begin{bmatrix} -\frac{u}{f_x} D_{x,y} & -\frac{v}{f_y} D_{x,y} & D_{x,y} \end{bmatrix}^T,$$

где u, v – координаты в пиксельном пространстве на карте глубин, f_x, f_y – фокусные расстояния в пикселях. Для каждой точки $P_{x,y}$ находится J точек, для которых Евклидово расстояние до точки $P_{x,y}$ меньше либо равно некоторого заданного расстояния. Таким образом, для каждой точки $P_{x,y}$ находится группа точек в окрестности $\{P_{x,y}, P_{x1,y1}, \dots, P_{xJ,yJ}\}$. Сформировав матрицу Q :

$$Q = \begin{bmatrix} P_{x,y}^T \\ P_{x1,y1}^T \\ \vdots \\ P_{xJ,yJ}^T \end{bmatrix},$$

Вычисление нормалей производится оптимизационным алгоритмом

$$N_{x,y}^{depth} = \underset{\vec{n}}{\operatorname{argmin}} \|(Q - \hat{Q})\vec{n}\|_2^2$$

где Q – матрица $J \times 3$ координат всех J точек, лежащих в заданном радиусе, а $\hat{Q}$ – матрица $J \times 3$, в каждом ряду которой находится центроида окрестности Q . Степень гладкости получившейся карты нормали контролируется с помощью радиуса окрестности для оптимизации.

Рассмотрим поляризационную карту нормалей N^{polar} . Задача ее коррекции эквивалентна задаче поиска следующего оператора $\hat{A}$:

$$\hat{A} = \underset{A}{\operatorname{argmin}} \|N^{depth} - A(N^{polar})\|_2^2 + \gamma \|\nabla A\|_1$$

при условии, что оператор $\hat{A}$ может принимать лишь бинарные значения 0 или 1. Фактически, данное ограничение на спектр принимаемых оператором значений используется для того, чтобы оператор исправлял неопределенность азимута поляризационных нормалей. То есть в каждой точке поляризационная нормаль либо претерпевает вращение, либо остается такой, как есть. Второе слагаемое накладывает ограничение к гладкости оператора. Это типичная total variation оптимизационная задача, которую можно решить методом graph cuts. Таким образом,

$$N^{corr} = \hat{A}(N^{polar})$$

Коррекция высокочастотной π -неопределенности азимута

Зашумленная карта глубин не сможет быть использована для коррекции азимутальной неопределенности на участках с небольшим характерным размером деталей (высокочастотных участках). Для того, чтобы разрешить неопределенность на этих участках, принимается, что участки должны быть замкнутыми и не содержать разрывов в итоговой карте глубин. Для этого нормали в группах пикселей корректируются таким образом, чтобы итоговая карта глубин, полученная интеграцией нормалей, не имела разрывов на границах высокочастотных участков. Алгоритм такой коррекции выглядит следующим образом: сначала из карты поляризационных нормалей вычитается карта нормалей от карты глубин; далее находятся точки, находящиеся на границе смены от 0 до ненулевого значения разницы – эти точки называются якорными; наконец жадным алгоритмом производится поворот участков между якорными точками, чтобы поверхность получилась замкнутой.

Коррекция зенита

Как уже упоминалось ранее, для вычисления корректного значения зенит-компоненты нормали к поверхности необходимо знать значение коэффициента

преломления поверхности объекта. Для диэлектриков значение коэффициента преломления варьируется не сильно, поэтому ошибка в определении этого значения не вызовет большой погрешности. Однако, для недиелектриков спектр принимаемых значений сильно шире, и некорректное его определение может вызвать существенную ошибку в итоговой 3D поверхности.

Для коррекции зенита сначала отмечаются регионы карты клбыин, которые могут дать грубое представление о форме объекта. Для этого определим бинарную маску M как:

$$M = \begin{cases} 1, & \nabla^T N^{depth} < \epsilon \text{ and } \nabla^T N^{corr} < \epsilon \\ 0, & o.w. \end{cases}$$

где ϵ – порог, задающий необходимую гладкость градиента. Смысл данной маски в том, чтобы выделить гладкие участки карт нормали, так как скорее всего на таких участках нет никаких высокочастотных деталей и по ним мы можем определить свойства материала объекта, а также взять карту нормалей из $N_{x,y}^{depth}$. Далее карта нормалей разбивается на патчи, в рамках каждого из которых производится операция поворота:

$$N^{corr} := \hat{R}(N^{corr})$$

где $\hat{R}$ – оператор поворота:

$$\hat{R} = \underset{R}{\operatorname{argmin}} \|\theta^{depth} - R(\theta^{corr})\|_2^2$$

где θ^{depth} и θ^{corr} – соответствующие значения зенита.

Объединение скорректированной карты нормалей и карты глубин

Имея корректную карту нормалей, можно получить трехмерную форму объекта, проинтегрировав данную карту нормалей. Однако, таким способом невозможно получить корректные разрывы в карте глубин. То есть, например, невозможно понять, есть ли расстояние между объектом и фоном. Поэтому для реконструкции трехмерной формы объекта нами предложен способ интеграции скорректированной карты нормалей N^{corr} , совместно с картой глубин D .

Стандартный способ интеграции карты нормалей в карту глубин – решение уравнения Пуассона:

$$\nabla^2 \hat{D} = \nabla^T N^{corr}$$

Однако для корректного объединения необходимо учесть, что не все нормали определены достаточно точно и гладко, поэтому вместо использования всех градиентов оптимальнее выбрать только наиболее надежные, сформировав тем самым дерево градиентов. Для формирования данного дерева используется сама физика поляризации – нормаль более шумная в тех точках, где степень поляризации ρ мала. Таким образом, формируется граф градиентов, по которому можно решать уравнение Пуассона:

$$\nabla_S^2 \hat{D} = \nabla_S^T N^{corr}$$

Вторым важным ограничением является корректность масштаба после интеграции карт нормалей. Другими словами, там, где карта глубин аккуратна, должно минимизироваться следующее ограничение

$$\|M \cdot (\hat{D} - D)\|_2^2$$

Таким образом, чтобы учесть упомянутые выше ограничения и корректно объединить карту глубин и карту нормалей, формируется и решается следующая оптимизационная задача:

$$\begin{pmatrix} \lambda M \cdot I \\ \nabla_S^2 \end{pmatrix} VEC(\hat{D}) = \begin{pmatrix} \lambda VEC(M \cdot D) \\ \nabla_S^T(N^{corr}) \end{pmatrix}$$

где VEC означает операцию векторизации, I – единичная матрица размера $MN \times MN$, λ – скалярный параметр, позволяющий варьировать степень влияния каждого из данных двух ограничений. Оптимизация производится с помощью QR факторизации.

Доказательства некоторых используемых утверждений.

В данном разделе второй главы диссертации приведены доказательства 2 ключевых утверждений:

Утверждение 1. Неопределенность по азимуту может принимать лишь 2 дискретных значения: неопределенность по π , либо по $\pi/2$, в зависимости от того, какая из компонент поступившего от объекта света больше в поляризации:

рассеянная или отраженная. Неопределенность не может принимать никаких промежуточных значений.

Утверждение 2. Неопределенность в определении зенита может быть скорректирована с помощью оператора поворота $\hat{R}$.

Анализ результатов работы алгоритма.

Приведены примеры работы алгоритма в различных условиях освещения и с различными материалами поверхности. Приведен сравнительный анализ с лазерным сканом объекта, а также численное сравнение точности предложенного метода относительно лазерного скана в сравнении с шейдинговым методом из [1].

Таблица 1. Численное сравнение точности предложенного метода, вычисленное как средняя абсолютная ошибка в миллиметрах относительно лазерного скана.

	Карта глубин, мм	Шейдинг, мм [1]	Предложенный, мм
Угол	5,39	4,78	3,63
Зеркальный шар	8,50	17,58	8,25
Белое лицо	18,58	18,30	18,28
Кофейный стакан	3,79	3,84	3,48

Алгоритм разделения отраженной и рассеянной компонент света с использованием поляризации света.

Большая часть объектов в реальном мире отражает свет не исключительно рассеяно или зеркально, а в виде композиции двух этих типов отражения. Таким образом, «смешанное отражение» случается, когда и рассеянная и зеркальная компоненты достигают камеры, затрудняя использование методов ФиП и добавляя дополнительный шум в карту нормалей. Был проделан анализ смешанного отражения и его влияния на существующие методы ФиП. В работе показано, что смешанное отражение действительно негативно влияет на полученный методом ФиП результат. В работе предложен метод анализа смешанного отражения на основе нескольких позиций камеры и показано сравнение с аналогичными

методами. Предложенный в работе вклад расширяет применимость ФиП, позволяя анализировать большой спектр возможных объектов с различными материалами.

Рассмотрим дихроматическую модель отражения, предполагающую, что интенсивность отраженного света содержит две ключевые компоненты:

$$I = I^d + I^s$$

где I^d и I^s относятся к интенсивности рассеянного и зеркально отраженного света соответственно. Предыдущие работы по ФиП вычисляли ρ из I_{max} и I_{min} , максимальной и минимальной интенсивности света для каждого пикселя при вращении поляризатора. Можно предположить, что вариации зеркально отраженной компоненты много больше вариаций рассеянной компоненты для поверхности со смешанным отражением, и поэтому измеренная степень поляризации света:

$$\rho = \frac{I_{max}^s - I_{min}^s}{I_{max}^s + I_{min}^s + I^d} = \tilde{\rho} \frac{I^s}{I} = \tilde{\rho} \frac{I - I^d}{I}$$

где I_{max}^s и I_{min}^s соответствуют максимуму и минимуму интенсивности зеркально отраженной компоненты света. Таким образом,

$$I^d = I(1 - \rho \frac{n^2 - \sin^2 \theta + \tan^2 \theta}{2 \sin \theta \tan \theta \sqrt{n^2 - \sin^2 \theta}})$$

В данном уравнении две неизвестных: интенсивность рассеянной компоненты света, I^d и коэффициент преломления n . В предположении Ламбертовского отражения, интенсивность рассеянной компоненты света не зависит от угла обзора. Коэффициент преломления, как свойство материала объекта, также не зависит от угла обзора. Таким образом, предлагаемый метод предполагает захват данных о поляризации отраженного света с нескольких ракурсов:

$$I^d = f(n, \theta_i, I_i, \rho_i) \triangleq I_i(1 - \rho_i \frac{n^2 - \sin^2 \theta_i + \tan^2 \theta_i}{2 \sin \theta_i \tan \theta_i \sqrt{n^2 - \sin^2 \theta_i}})$$

для i -го положения камеры среди N ракурсов. Для поиска I^d решается нелинейная оптимизационная задача в следующей форме:

$$\{\widehat{I^d}, \widehat{n},\} = \operatorname{argmin}_{I^d, n} \sum_{i=1}^N (I_d - f(n, \theta_i, I_i, \rho_i))^2.$$

Оптимизация производилась с помощью метода последовательного квадратичного программирования. Предложенный метод разделения рассеянной и отраженной зеркально компонент сравнен с методом, предложенным Nishino et al. [2].

В третьей главе «Сегментация прозрачных объектов на изображении с помощью поляризации отраженного света» приводится алгоритм сегментации прозрачных объектов на основе сверточной нейронной сети с использованием поляризации света.

С развитием сверточных нейронных сетей и глубокого обучения алгоритмы детектирования и сегментации объектов на изображении получили существенный прирост в точности при валидации на стандартных бенчмарках, таких как MSCOCO и Imagenet. Это достигается благодаря фундаментальному свойству глубоких сверточных нейронных сетей качественно обобщать большие датасеты изображений. Фактически, такое обобщение позволяет построить представление о визуальных характеристиках различных объектов, и это все базируется на предположении, что объект в валидационном датасете будет выглядеть до некоторой степени похоже на объект того же класса в тренировочном датасете. Именно это и позволяет глубоким сверточным нейронным сетям оставаться устойчивыми к сильным изменениям фона за объектом, условий освещения, и прочих характеристик. Действительно, эти условия корректны для большинства объектов: так, теннисный мяч будет выглядеть подобно теннисному мячу на различных изображениях. Однако, данное предположение не совсем выполнимо в случае прозрачных объектов. Прозрачные объекты по большей части пропускают свет сквозь себя, а не отражают его, поэтому их внешний вид существенно зависит от особенностей фона и окружающей среды, и поэтому его намного труднее обобщить. Из-за этого детектирование и сегментация прозрачных объектов - сложная задача для алгоритмов машинного обучения. Поляризация света позволяет

существенно упростить эту задачу, так как поляризация отраженного от прозрачного объекта света будет менее подвержена влиянию фона и будет более ярко выделять границы объекта.

Именно за счет этого физического свойства использование поляризации позволяет получить весомое преимущество при сегментации прозрачных объектов. В главе рассмотрены: описание архитектуры, примененной для задачи сверточной нейросети, описание собранного датасета и необходимой предобработки для подачи на вход в сверточную нейросеть, а также приведем ряд экспериментов, оценивающих результаты работы алгоритма в сравнении с сегментацией по обычным изображениям.

Архитектуры Mask R-CNN и Polar Mask R-CNN для сегментации изображений.

Наиболее актуальной архитектурой сверточных нейронных сетей для задачи сегментации объектов является Mask R-CNN, продолжившая цепочку архитектур R-CNN, Fast R-CNN и Faster R-CNN. Mask R-CNN производит одновременную сегментацию объектов, детектирование и классификацию. При этом, Mask R-CNN является по большому счету мета-архитектурой, то есть составные части нейросети можно составлять из других известных архитектур, таких как Resnet, FPN и FCN. По сути, нейросеть состоит из двух частей:

Backbone. Она также называется RPN, Region Proposal Network. Изображение целиком сначала проходит через эту часть нейросети. На выходе на изображении выделяются регионы, в которых с большей вероятностью есть какие-либо объекты. Далее, эти регионы по отдельности параллельно подаются в следующую часть сети - heads. В оригинальной статье авторами предлагалось несколько вариантов архитектуры для backbone-сети, в данной работе применяется архитектура Resnet101.

Heads. Каждый из поданных на вход регионов проходит через 3 параллельных ветви, работа которых не зависит друг от друга: первая ветвь занимается классификацией объекта в регионе, вторая - коррекцией охватывающей рамки, третья - сегментацией объектов в регионе. Причем, сегментация производится

сразу для всех возможных классов параллельно, и затем просто выбирается маска, соответствующая классу, выбранному параллельной веткой классификации. Функция потерь нейросети строится следующим образом:

$$L = L_{cls} + L_{box} + L_{mask},$$

где L_{cls} – функция потерь определения класса объекта (кроссэнтропия), L_{box} – функция потерь определения охватывающей рамки, L_{mask} – функция потерь определения маски, которая вычисляется как бинарная кроссэнтропия. Фактически, задача нейросети классифицировать корректно каждый пиксель на изображении, как относящийся к некоторому объекту, либо как относящийся к фону. Для сегментации используется архитектура FCN – Fully Convolutional Network.

Заметим, что в классической архитектуре Mask R-CNN предполагается, что изображение на входе будет обычным RGB изображением, в котором разные каналы имеют одинаковый физический смысл. Однако в случае с подаваемыми на вход поляризационными данными, их физический смысл и вид существенно отличается от обычного изображения в видимом свете. Поэтому, была произведена модификация архитектуры Mask R-CNN, чтобы учесть этот фактор, а именно, backbone сети был распараллелен в 3 независимых одинаковых ветви, по одному на каждый из каналов. Значения на выходе этих каналов усреднялись между собой. Таким образом, сверточная нейросеть сможет максимально адаптироваться к столь разным, по сути, данным как угол поляризации, степень поляризации и обычное изображение в видимом диапазоне. Мы назвали данную архитектуру Polar Mask R-CNN, ниже проанализируем прирост в точности, который такое изменение позволяет получить.

Датасет прозрачных объектов с поляризационными данными.

Для обучения нейросети был собран и размечен датасет изображений прозрачных объектов, для которых была также захвачена информация о поляризации света с помощью специальной поляризационной камеры. Поляризационная камера представляет из себя камеру машинного зрения, у которой на субпикселях расположены микрополяризационные фильтры. Группа из

четырёх субпикселей с разной ориентацией поляризационных фильтров образует один пиксель. В каждом таком пикселе можно определить степень поляризации и угол поляризации.

Всего было собрано более 700 изображений, из которых:

1. Класс 1: 586 изображений с прозрачными объектами 6 видов: шар, большой стеклянный стакан, елочная игрушка, небольшой пластиковый стаканчик, пластиковый лоток и стеклянная ваза. Все объекты лежали в картонной коробке в различном количестве и в разных комбинациях.
2. Класс 2: 130 изображений с прозрачными шарами, лежащими в различных местах вне картонной коробки – на столе, на ковре, и так далее.

Далее все эти изображения были размечены с помощью инструмента полигональной аннотации: на каждом был выделен контур прозрачного объекта, и был отмечен соответствующий класс. Поляризационные изображения были автоматически сконвертированы в карту углов азимута (с поляризационной неопределенностью) и карту степеней поляризации.

Дополнительной сложностью и новшеством в собранном датасете является наличие изображений с крайне большой плотностью объектов, из-за которой они фактически перекрываются друг другом. Это делает процедуру сегментации еще более сложной для алгоритма.

Заметим, что поляризационные изображения отличаются по своим свойствам от классических RGB/Grayscale изображений. Так, например, вычисленное изображение угла поляризации содержит ряд границ в классическом их понимании. Однако, на самом деле, это переходы от $-\pi$ к π и не являются границами в поляризационном понимании. Чтобы упростить задачу нейросети, можно добавить шаг препроцессинга и вычислить $\cos(\varphi)$, где φ – угол поляризации, чтобы сгладить такие границы. Также немаловажно правильно аугментировать поляризационные данные при подаче их на вход в обучение. Так, к примеру, вращение угла поляризации не будет оказывать влияния на результат, важно лишь

относительное его изменение. Поэтому из одного изображения можно вычислить несколько с повернутым азимутом и также подать их на вход. Данные действия аналогичны аугментации обычных изображений при обучении сверточных нейросетей, однако ранее такие подходы не предлагались, так как обучение на основе поляризационных данных не проводилось.

Анализ полученных результатов.

В главе приведен сравнительный анализ предложенного подхода сегментации и существующих на данный момент методов. Показано, что использование поляризации позволяет получить прирост в точности и надежности сегментации прозрачных объектов, а предложенная архитектура сверточной нейросети Polar Mask R-CNN позволяет дополнительно улучшить результат. Прирост в точности варьируется от 28% (с уменьшением ошибки на 79%) до 6.3%, в зависимости от условий эксперимента.

В **Заключении** подведены основные результаты диссертационного исследования.

Фокус данной диссертационной работы заключался в расширении спектра задач компьютерного зрения и вычислительного имаджинга, в которых дополнительную ценность принесет информация о поляризации поступающего в камеру света. Для этого было решено остановиться на двух задачах, в которых потенциальная польза от поляризации предполагалась наибольшей: уточнение трехмерной формы объекта с помощью методов ФиП, а также уточнение сегментации прозрачных объектов на изображениях. Разделение на главы диссертационной работы соответствует данному разделению на основные темы.

В ходе исследований получены следующие результаты:

1. Предложен алгоритм реконструкции трехмерной формы объекта с помощью совмещения информации о поляризации света и грубой карты глубин.
2. Проведены эксперименты на большом количестве объектов различной формы и свойств поверхности, а также при различных свойствах

освещения. Показано, что данный алгоритм превосходит по точности и надежности все существующие алгоритмы.

3. Проанализированы теоретические аспекты алгоритма, а также доказаны 2 ключевых утверждения.
4. Предложен метод разделения отраженной и рассеянной компонент света на основе свойств поляризации света. Данный метод имеет ценность сам по себе, так как является единственным методом разделения компонент, основанным на фундаментальных физических принципах каждой из компонент. Также, он может потенциально использоваться совместно с ФиП для уточнения поляризационной карты нормалей.
5. Предложен метод сегментации прозрачных объектов с использованием поляризации света.
6. Собран и размечен датасет изображений прозрачных объектов с информацией о поляризации света. Это крупнейший поляризационный датасет в мире, насколько нам известно, и единственный датасет, сфокусированный на прозрачных объектах.
7. Предложен способ предобработки поляризационных изображений перед использованием их для обучения сверточной нейросети, а также изменения архитектуры самой сверточной нейросети для поляризационных данных.
8. Показано, что разработанный метод сегментации изображений прозрачных объектов существенно лучше обобщает по сравнению с обычными изображениями в видимом спектре и более точен в задаче сегментации прозрачных объектов.

Таким образом, данная диссертационная работа приводит 2 новых способа применения поляризации света в задачах компьютерного зрения и показывает, что потенциал использования поляризации света далек от того, чтобы быть полностью раскрытым.

Приложения содержат программный код с реализацией алгоритма совмещения ФИП и карты глубин, а также акт о внедрении результатов разработки.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

В изданиях, входящих в перечень RSCI:

1. Kadambi A., Taamazyan, V., Shi, B., & Raskar, R. Polarized 3d: High-quality depth sensing with polarization cues //Proceedings of the IEEE International Conference on Computer Vision. – 2015. – С. 3370-3378.
2. Kadambi A., Taamazyan, V., Shi, B., & Raskar, R. Depth sensing using geometrically constrained polarization normals //International Journal of Computer Vision. – 2017. – Т. 125. – №. 1-3. – С. 34-51.
3. Таамазян В. А. Поляризация при смешанном отражении в задаче трехмерного сканирования объектов //Компьютерная оптика. – 2018. – Т. 42. – №. 5.

Патенты:

1. «Methods and Apparatus for Enhancing Depth Maps with Polarization Cues», Achuta Kadambi, Vage Taamazyan, Boxin Shi, Ramesh Raskar, US20160261844A1.

ЛИТЕРАТУРА

1. Wu C. et al. Real-time shading-based refinement for consumer depth cameras //ACM Transactions on Graphics (ToG). – 2014. – Т. 33. – №. 6. – С. 200.
2. Nishino, K., Zhang, Z., Ikeuchi, K.: Determining reflectance parameters and illumination distribution from a sparse set of images for view-dependent image synthesis. In: Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on. Volume 1., IEEE (2001) 599–606

Научное издание

Таамазян Ваге Арамаисович

**ИССЛЕДОВАНИЕ И РАЗРАБОТКА МЕТОДОВ ТРЕХМЕРНОГО
СКАНИРОВАНИЯ И СЕГМЕНТАЦИИ ОБЪЕКТОВ С
ИСПОЛЬЗОВАНИЕМ ПОЛЯРИЗАЦИИ СВЕТА**

АВТОРЕФЕРАТ

диссертации на соискание ученой степени

кандидата технических наук

Федеральное государственное автономное образовательное учреждение

высшего образования

Московский физико-технический институт

(национальный исследовательский университет)

141701,

Московская область, г. Долгопрудный

Институтский переулок, д. 9

<https://mipt.ru>