

На пути использования смартфона в качестве гиперспектральной камеры

Д.А. Реутский^{1,2}, А.В. Погадаев^{1,2}, Д.М. Владимиров^{1,2}, Е.И. Ершов^{1,2}

¹Институт проблем передачи информации им. А. А. Харкевича,
127051, г. Москва, Большой Каретный переулок, д. 19, стр. 1;

²Московский физико-технический институт (национальный исследовательский университет),
141701, Московская область, г. Долгопрудный, Институтский пер., д. 9

Аннотация

Спектральная реконструкция из RGB-изображений – одна из актуальных задач вычислительной фотографии, но на сегодняшний день даже самые современные алгоритмы не обеспечивают приемлемую точность. Любопытно, что в этой задаче современные мобильные устройства открывают новую возможность повышения качества реконструкции за счёт использования изображений сразу с нескольких камер. Это даёт почву для создания мобильной гиперспектральной камеры для широкого круга пользователей. В работе исследован вопрос достижимой точности при использовании одновременно нескольких одинаковых камер в сочетании с различными спектральными фильтрами. Предложено два алгоритма для поиска оптимальных фильтров для случаев обучения линейной регрессии и полносвязной нейронной сети в качестве алгоритма реконструкции. В результате вычислительных экспериментов было установлено, что в случае 4 камер и 4 фильтров точность реконструкции в три раза выше, чем при восстановлении из RGB-изображений.

Ключевые слова: спектр, гиперспектральные изображения, фильтр, восстановление, реконструкция.

Цитирование: Реутский, Д.А. На пути использования смартфона в качестве гиперспектральной камеры / Д.А. Реутский, А.В. Погадаев, Д.М. Владимиров, Е.И. Ершов // Компьютерная оптика. – 2025. – Т. 49, № 1. – С. 112-120. – DOI: 10.18287/2412-6179-CO-1315.

Citation: Reutsky DA, Pogadaev AV, Vladimirov DM, Ershov EI. Towards using smartphones as hyperspectral cameras. Computer Optics 2025; 49(1): 112-120. DOI: 10.18287/2412-6179-CO-1315.

Введение

Гиперспектральная съёмка широко используется в областях, где важно установить физические или химические свойства объектов. В медицине – для определения болезней кожи [1] или органов [2]; в сельском хозяйстве – для оценки состояния растительных культур [3], а также для отделения свежих фруктов и овощей от незрелых или испорченных [4]; в мусоропереработке – для сортировки отходов по химическому составу [5]; в цифровой печати – для калибровки и измерения цвета [6].

Однако камеры, способные снимать гиперспектральные изображения, очень дорогие для широкого применения, поэтому возникает идея использовать другие способы получения таких снимков. Наиболее актуальными в настоящий момент являются методы вычислительной фотографии, задача которых состоит в том, чтобы реконструировать спектры в видимом диапазоне из RGB-изображений [7]. Существует множество алгоритмов для такой спектральной реконструкции. Среди современных решений хорошую точность обеспечивают как классические алгоритмы [8], так и нейросетевые [9, 10].

Вдобавок к упомянутым алгоритмам существуют способы улучшить сами изображения, из которых будет производиться реконструкция, чтобы последняя удавалась точнее. Один из вариантов такого улучше-

ния – увеличение количества каналов изображения. Хеллинг и соавторы предложили использовать 7 цветных узкополосных фильтров [11], чтобы одноканальной камерой осуществлять мультиспектральную съёмку; таким образом, на выходе получалось 7 чёрно-белых фотографий одной сцены, которые объединялись в 7-канальное изображение. Вычислительные эксперименты показали, что средняя цветовая ошибка ΔE_{94} в данном случае составляет около 0,3 (меньше порога человеческого цветоразличения). В работе [12] Валеро и соавторы снимали одну и ту же сцену трёхканальной камерой сначала без фильтров, а затем поочерёдно с фильтрами из ограниченного набора. Таким образом, при использовании n фильтров получалось $(3n + 3)$ -канальное изображение. Целью работы был поиск наилучшей комбинации для спектральной реконструкции. С увеличением n качество реконструкции улучшалось, но на $n = 3$ авторы остановились, так как решили, что дальше оно существенно улучшаться не будет. В сравнении с реконструкцией по трём каналам у них получилось достичь втрое лучшей точности (по метрике RMSE, нормированной на максимум спектра).

Альтернативно выбору ограниченного набора фильтров из существующих в литературе рассмотрен вопрос проектирования спектральных фильтров «с нуля». Этому посвящена, например, работа Финлейсона и соавторов [13]: там ищется ровно один фильтр,

который будет наложен на трёхканальную камеру. Но ставится цель не реконструировать спектры, а сделать камеру более колориметрической, то есть такой, цветное преобразование координат которой позволит точнее предсказывать отклики стандартного наблюдателя (CIE 1931) на ту же сцену. Авторы показали, что точность предсказания с фильтром в среднем выше в 4 раза, чем без фильтра. Статья Ду-Йонга и соавторов [14] посвящена спектральной реконструкции при помощи съёмки трёхканальной камерой с наложением различных фильтров. Это та же идея, что и у Валеро [12], но функции пропускания фильтров имеют произвольную форму и ищутся математическим моделированием и оптимизацией. Полученные результаты используются для реконструкции некоторого ограниченного множества спектров, а именно спектр-стимулов зубов. Обе работы объединяют ограничения на функции пропускания фильтров – максимум второй производной и вложенность множества значений в отрезок $[0; 1]$. Первое ограничение авторы вводят как необходимое условие для физической реализуемости фильтров, а второе довольно естественно: фильтры не могут усиливать интенсивность света или делать её отрицательной.

В упомянутых работах есть недостаток: в них не рассматривается никаких методов спектральной реконструкции, кроме линейной регрессии. Этого недостатка лишена работа схожей тематики от Ни и соавторов [15], где при помощи глубокого обучения проектируются сами чувствительности мультиспектральной камеры, из снимков которой нейросетевым методом восстанавливаются гиперспектральные изображения.

В отличие от статьи [15] и сходно со статьями [12, 14], в данной работе рассматривается задача проектирования спектральных фильтров для уже существующего набора из 4 камер. Одновременно с этим выбирается алгоритм для наилучшей спектральной реконструкции. Эти два вопроса *вместе* не рассматривались ни в одной из приведённых статей.

Перспективная задача – сделать гиперспектральную съёмку доступной широкому кругу лиц. На данный момент самая распространённая категория смартфонов имеет четыре близко расположенные камеры: доля таких устройств на рынке превышает 30% [16]. Если изготовить чехол для смартфона, который «накроет» каждую камеру своим фильтром, чувствительности каждой из четырёх камер будут изменены по-своему, и вместо четырёх практически одинаковых 3-канальных сенсоров мы получим четыре *различных* 3-канальных сенсора. Или, иными словами, один 12-канальный. Это позволит получать 12-канальные изображения практически единомоментно. Если установить на телефон программное обеспечение для спектральной реконструкции и решить проблему параллакса (каждая камера видит сцену немно-

го под разными ракурсами), то мы получим мобильную гиперспектральную камеру (рис. 1).

Итак, цель данной работы – провести численное моделирование описанной 12-канальной съёмки (без параллакса) и при помощи методов оптимизации подобрать такие функции пропускания фильтров, чтобы восстановление гиперспектральных изображений удавалось наилучшим образом.

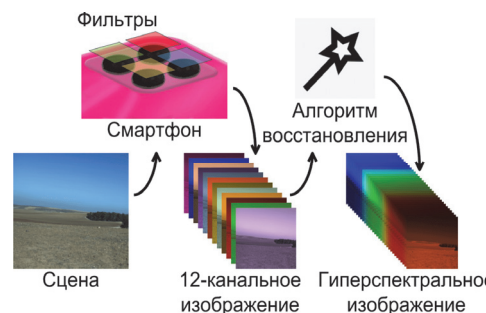


Рис. 1. Предлагаемая схема доступной спектральной реконструкции

Постановка задачи

Заранее отметим, что в данной работе мы будем иметь дело не с изображениями, а с набором пикселей, то есть не будем использовать пространственную информацию. Все спектры будут восстанавливаться независимо друг от друга.

Спектр излучения представим в упрощённом виде – в виде вектора – если зафиксировать диапазон длин волн и шаг дискретизации (например, 400, 410, 420, ..., 730 нм). Пусть $\{\mathbf{r}_i\}_{i=1}^n \subset \mathbb{R}^k$ – набор спектров в таком представлении. RGB-отклики камеры можно симулировать выражением $\mathbf{c}_i = S' \mathbf{r}_i$, где S' – матрица чувствительности камеры размером $3 \times k$. Но мы применяем 4 камеры и 4 различных фильтра, поэтому S' нужно заменить на $S \circ (F)_3$, где матрица $S: 12 \times k$ составлена из спектральных чувствительностей 4 камер, $F: 4 \times k$ – фильтры, $(\cdot)_3$ – матричный оператор, дублирующий каждую строку 3 раза, \circ – поэлементное умножение. Итак, отклики 12-канального сенсора даются выражением

$$\mathbf{c}_i = (S \circ (F)_3) \mathbf{r}_i. \tag{1}$$

Из откликов камеры требуется восстановить спектры. Задача некорректна, ведь каждому отклику, вообще говоря, соответствует бесконечное число спектров, которые могли его создать (метамеры). Однако практика показывает, что спектры сцен из окружающего мира покрывают отнюдь не всё k -мерное пространство. Фостер и соавторы [17] в своей работе показали, что спектры натуральных красителей при естественном освещении лежат на маломерном многообразии, а относительная частота метамеров (доля метамерических пар среди всевозможных пар спектров из датасета) обычно не превышает 10^{-4} .

Введём алгоритм восстановления спектра $\mathbf{g}_0: \mathbb{R}^{12} \rightarrow \mathbb{R}^k$, имеющий какие-то параметры

$(\theta_1, \theta_2, \dots)^T = \theta$. Тогда восстановленные спектры выражаются как

$$\hat{\mathbf{r}}_i = \mathbf{g}_\theta(\mathbf{c}_i) = \mathbf{g}_\theta((S \circ (F)_3) \mathbf{r}_i). \quad (2)$$

Ставится задача о минимизации

$$\sum_{i=1}^n \|\hat{\mathbf{r}}_i - \mathbf{r}_i\|_2^2$$

при различных θ и F . Аналогично предшествующим работам будем учитывать физические ограничения на множество всевозможных фильтров F :

- 1) *ограниченность* – диапазон значений элементов вложен в отрезок $[0; 1]$;
- 2) *гладкость* – ограниченность «второй производной». В качестве дискретного аналога последней используем разностную схему

$$f''(x) \approx \frac{1}{h^2} (f(x+h) + f(x-h) - 2f(x)) \quad (3)$$

и с помощью строкового преобразования

$$D = \begin{bmatrix} 1 & 0 & \vdots & 0 \\ -2 & 1 & \vdots & 0 \\ 1 & -2 & \vdots & 0 \\ 0 & 1 & \vdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \vdots & 1 \end{bmatrix} \quad (4)$$

получаем из матрицы фильтров матрицу вторых производных функций пропускания: FD . Значит, можно использовать регуляризационное слагаемое $\lambda \|FD\|_F^2$, которое поможет удовлетворить требованиям на гладкость функций пропускания.

В реальной жизни \mathbf{c}_i , подаваемые на вход алгоритму \mathbf{g}_θ , – это измерения спектрального сенсора (камеры), то есть данные с погрешностью, с шумом. Чтобы учесть это при оптимизации, добавим шум ε_i к цветовым откликам \mathbf{c}_i . Генерировать будем его из нормального распределения. Как будет видно в дальнейшем, такое упрощение поможет нам ускорить оптимизацию.

Итого задача оптимизации фильтров в общем случае ставится так:

$$\begin{aligned} \min_{\theta, F} \quad & \sum_{i=1}^n \|\mathbf{g}_\theta((S \circ (F)_3) \mathbf{r}_i + \varepsilon_i) - \mathbf{r}_i\|_2^2 + \lambda \|FD\|_F^2. \quad (5) \\ \text{s.t.} \quad & F \in [0; 1]^{4 \times k} \end{aligned}$$

Рис. 2 поможет разобраться с устройством первого слагаемого оптимизационного функционала.

Методы решения

Для решения поставленной задачи требуется рассмотреть два вопроса: создание метода поиска оптимального фильтра и создание методов спектральной реконструкции на основе 12-канального изображения.

Первый вопрос уже почти решён: в (5) все символы понятны, кроме \mathbf{g}_θ – он был введён как абстрактный параметризуемый метод спектральной реконструкции. Таким образом, остаётся только вопрос подбора методов спектральной реконструкции, которые, в свою очередь, разделились на два типа: параметризуемые и непараметризуемые. Для решения (5) годятся только параметризуемые, они позволяют одновременную оптимизацию формы фильтра и параметров алгоритма реконструкции. С рассмотрения параметризуемых методов и начнём. Мы попробовали два варианта.

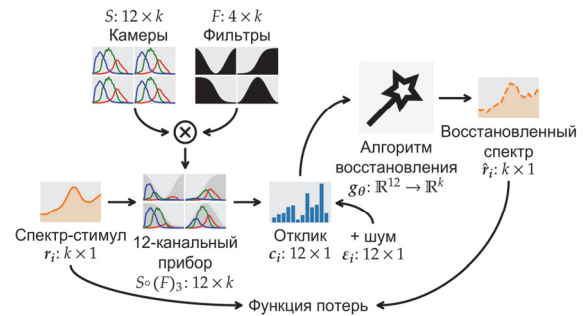


Рис. 2. Схема вычисления первого слагаемого из оптимизационного функционала (5)

1. Линейная регрессия. В задаче проектирования фильтров её часто используют в качестве baseline-решения. Зададим линейное преобразование из откликов в спектры матрицей $M: k \times 12$. Тогда $\hat{\mathbf{r}}_i = M(\mathbf{c}_i + \varepsilon_i)$. Нахождение в рамках простейшей линейной алгебры позволяет нам проделать несколько трюков.

Во-первых, перепишем первое слагаемое из (5) в вероятностном виде

$$\begin{aligned} \sum_{i=1}^n \|M((S \circ (F)_3) \mathbf{r}_i + \varepsilon_i) - \mathbf{r}_i\|_2^2 &\approx \\ &\approx n \mathbb{E}_{\mathbf{r}, \varepsilon} \|M((S \circ (F)_3) \mathbf{r} + \varepsilon) - \mathbf{r}\|_2^2, \quad (6) \end{aligned}$$

где случайный вектор \mathbf{r} равномерно принимает значения из $\{\mathbf{r}_i\}_{i=1}^n$, а ε – независимый от \mathbf{r} гауссовский случайный вектор с матожиданием 0 и матрицей ковариации $\sigma^2 I$.

Введя обозначение $M(S \circ (F)_3) - I = A$, проделаем преобразования:

$$\begin{aligned} \mathbb{E}_{\mathbf{r}, \varepsilon} \|M((S \circ (F)_3) \mathbf{r} + \varepsilon) - \mathbf{r}\|_2^2 &= \\ &= \mathbb{E}_{\mathbf{r}, \varepsilon} \|M(S \circ (F)_3) \mathbf{r} - \mathbf{r} + M \varepsilon\|_2^2 = \\ &= \mathbb{E}_{\mathbf{r}, \varepsilon} \|A \mathbf{r} + M \varepsilon\|_2^2 = \\ &= \mathbb{E}_{\mathbf{r}} \|A \mathbf{r}\|_2^2 + \mathbb{E}_{\varepsilon} \|M \varepsilon\|_2^2 + 2 \langle \mathbb{E}_{\mathbf{r}} A \mathbf{r}, \mathbb{E}_{\varepsilon} M \varepsilon \rangle = \\ &= \mathbb{E}_{\mathbf{r}} \|A \mathbf{r}\|_2^2 + \mathbb{E}_{\varepsilon} \text{tr}[\varepsilon^T M^T M \varepsilon] + 2 \langle \mathbb{E}_{\mathbf{r}} A \mathbf{r}, \mathbb{E}_{\varepsilon} M \varepsilon \rangle = \quad (7) \\ &= \mathbb{E}_{\mathbf{r}} \|A \mathbf{r}\|_2^2 + \text{tr}[M^T M \mathbb{E}_{\varepsilon} \varepsilon \varepsilon^T] + 2 \langle \mathbb{E}_{\mathbf{r}} A \mathbf{r}, M \mathbb{E}_{\varepsilon} \varepsilon \rangle = \\ &= \mathbb{E}_{\mathbf{r}} \|A \mathbf{r}\|_2^2 + \text{tr}[M^T M \text{var}_{\varepsilon} \varepsilon] = \\ &= \mathbb{E}_{\mathbf{r}} \|A \mathbf{r}\|_2^2 + \sigma^2 \text{tr}[M^T M] = \\ &= \mathbb{E}_{\mathbf{r}} \|A \mathbf{r}\|_2^2 + \sigma^2 \|M\|_F^2. \end{aligned}$$

Получается, *регуляризация гауссовским шумом на цветовых откликах эквивалентна L2-регуляризации на матрицу M .*

Кроме того, можно составить матрицу $R: k \times n$ из столбцов \mathbf{r}_i . Тогда задачу можно записать в следующей нотации:

$$\begin{aligned} \min_{M,F} \quad & \|M(S^\circ(F)_3)R - R\|_F^2 + \lambda \|FD\|_F^2 + \mu \|M\|_F^2, \\ \text{s.t.} \quad & F \in [0;1]^{4 \times k}, \end{aligned} \quad (8)$$

где $\mu = n\sigma^2$.

Первое слагаемое – самое вычислительно затратное из-за размеров R (n порядка нескольких миллионов, а $k \sim 30$). Используя сингулярное разложение $R = U\Sigma V^T$, а также тот факт, что ортогональное преобразование не меняет норму Фробениуса, получим

$$\|AR\|_F^2 = \|A U \Sigma V^T\|_F^2 = \|A U \Sigma\|_F^2, \quad (9)$$

где для сокращения записи вновь введено $A = M(S^\circ(F)_3) - I$.

Заметим теперь, что $U\Sigma$ имеет те же размеры, что и R , однако последние $n - k$ столбцов нулевые и не влияют на значение выражения. Значит, $U\Sigma$ можно обрезать до размеров $k \times k$. Теперь при подсчёте целевой функции и её градиентов не нужно проводить операции с большой матрицей R . Поскольку в процессе оптимизации R не меняется, сингулярное разложение можно предподсчитать. На практике такой подход позволил ускорить оптимизацию в 100 раз.

Сама оптимизация состоит из чередования двух этапов:

- шага градиентного спуска для F при фиксированной M ;
- нахождения оптимальной M при фиксированной F (линейная регрессия с L2-регуляризацией).

Скорость сходимости оптимизационного процесса можно видеть на рис. 3. Количество необходимых итераций определялось при слежке за изменением фильтров. Как только фильтры переставали изменяться, процесс останавливался.

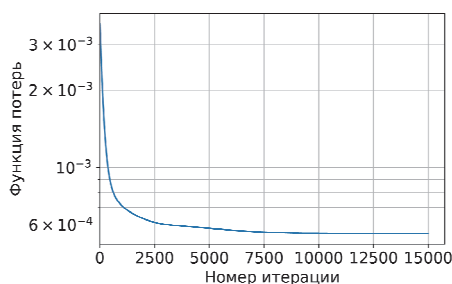


Рис. 3. Динамика изменения функции потерь из (8) при использовании оптимизатора Adam

Введение регуляризации на гладкость (параметр λ) было заимствовано из литературы [14]; однако в ней не даются какие-либо рекомендуемые значения. Поэтому пришлось обойтись подбором λ «на глаз». Слишком малые значения не оказывали достаточного влияния на гладкость. Слишком большие приводили к тому, что множество значений функций пропускания покрывало не весь отрезок $[0; 1]$. Все результаты в данной статье получены при $\lambda = n \cdot 0,0016$.

2. Нейронная сеть. В качестве \mathbf{g}_0 была опробована полносвязная нейронная сеть с тремя скрытыми слоями размерности 100 и функцией активации LeakyReLU. Вид задачи (5) не меняется, за исключением рассмотрения шума. Чтобы не подстраиваться под его конкретную реализацию, после каждого шага оптимизации ϵ_i регенерировался.

Для обоих вариантов \mathbf{g}_0 использовался оптимизатор Adam из пакета PyTorch, хорошо зарекомендовавший себя для невыпуклых задач. Он не поддерживает оптимизацию с ограничениями, поэтому после каждого шага отрицательные элементы матрицы F занулялись, а превосходящие единицу заменялись на 1.

Если зафиксировать F , то можно будет рассматривать не только параметризуемые алгоритмы восстановления. В данной работе в качестве такого фиксированного F детально рассматривается результат оптимизации фильтров при линейно-регрессионной \mathbf{g}_0 (нейросетевой вариант рассмотрен отдельно). А также добавлены некоторые baseline-решения для F :

1. Пусть у каждого фильтра будет три пика в красной, зелёной и синей зонах видимого спектра соответственно. Эти пики подберём вручную так, чтобы чувствительность каждого из 12 каналов полученного сенсора имела характерный пик в какой-то своей длине волны (см. рис. 4).
2. Сгенерируем элементы F из нормального распределения со средним 0,5 и дисперсией 0,1².
3. Все элементы F равны 1 (эквивалент отсутствия фильтров).

Каждый набор фильтров F будет иметь своё кодовое обозначение. Фильтры, полученные при помощи линейной регрессии, будут именоваться *Regression-trained*, а при помощи нейронной сети – *FCN-trained*. Baseline-фильтры из предыдущего параграфа назовём *Manual*, *Random* и *None* соответственно.

Сравнение «натренированных» фильтров с baseline-решениями поможет понять оправданность нашего вычислительного подхода к подбору фильтров.

Как только матрица F зафиксирована, постановка задачи становится классической, и выбор методов реконструкции практически не ограничен. Главное, чтобы метод можно было адаптировать под восстановление из 12-мерных цветовых откликов, а не из 3-мерных. В работе рассмотрены следующие алгоритмы восстановления:

1. **Линейная регрессия.** Уже была описана при постановке задачи (8).
2. **A+.** Алгоритм, предложенный Ву и соавторами [8], основанный на разреженном кодировании данных и представлении каждого исходного спектра линейной комбинацией из предпочитанного словаря.
3. **FCN.** Полносвязная нейронная сеть. Уже была описана сразу после линейной регрессии.
4. **Kernel FCN.** Данный метод основан на работе Финлэйсона и соавторов [18]. В ней предлагается

идея, позволяющая из алгоритмов спектральной реконструкции делать такие, которые предсказывают спектр с нулевой цветовой ошибкой. Идея состоит в следующем. Камеру можно рассмотреть как линейное преобразование из пространства спектров в пространство цветов. Разобьём спектральное пространство на ядро преобразования и дополнение к нему. Теперь восстановление спектра будет заключаться в восстановлении каждой из компонент из этих подпространств. Компонента не из ядра восстанавливается аналитически. А для восстановления компоненты из ядра будем использовать привычные алгоритмы спектральной реконструкции. В качестве такого мы взяли полносвязную сеть из предыдущего пункта.

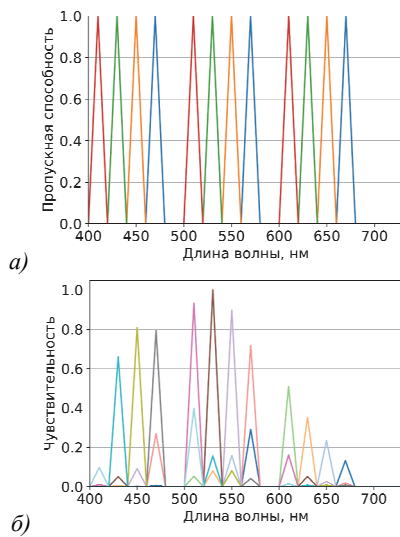


Рис. 4. Пропускные способности Матриал-фильтров (а) и чувствительности соответствующего 12-канального спектрального сенсора (б)

Оценка качества

После обучения алгоритмов нужно проверить их точность на тестовой части набора данных. Встаёт вопрос о выборе метрик качества. В исключительно прикладном исследовании конечная метрика должна быть проблемно-ориентированной: например, если целевая задача – угадывать цвет красителя, то метрика должна быть цветовой. В нашем случае целей для восстановления спектра множество, поэтому предстоит абстрактно оценивать отличие восстановленных спектров от исходных. В таком случае от метрики ожидаются следующие свойства:

1. Интерпретируемость. Возможность интуитивно понять, большая величина получилась или маленькая.
2. Практичность. Удобство в сравнении с другими исследованиями, а также в использовании для оценок точностей при решении разных конкретных задач. Например, если известно значение метрики восстановления спектра, какой будет метрика восстановления цвета?

С учётом сказанного мы ввели следующую метрику качества восстановления спектров:

$$NSE(R, \hat{R}) = \frac{1}{n} \sum_{i=1}^n \frac{\| \mathbf{r}_i - \hat{\mathbf{r}}_i \|_1}{\| \mathbf{r}_i \|_1}. \tag{10}$$

Данная метрика призвана решить проблему интерпретируемости. По сути, это площадь разности спектров, нормированная на площадь исходного. Такую величину можно измерять в процентах.

Также были выбраны ещё две метрики:

$$RMSE(R, \hat{R}) = \sqrt{\frac{1}{nk} \sum_{i=1}^n \| \mathbf{r}_i - \hat{\mathbf{r}}_i \|_2^2}, \tag{11}$$

$$MangE(R, \hat{R}) = \frac{1}{n} \sum_{i=1}^n \arccos \frac{\langle \mathbf{r}_i, \hat{\mathbf{r}}_i \rangle}{\| \mathbf{r}_i \|_2 \| \hat{\mathbf{r}}_i \|_2}.$$

RMSE является популярной метрикой качества восстановления спектра (например, ею оценивались решения крупнейшего соревнования по спектральному восстановлению [7]). По своей сути это среднеквадратичное отклонение, приходящееся на длину волны. По значению этой метрики можно приблизительно выразить, насколько два спектра могут отличаться между собой. Но интуитивно представить себе эту метрику сложно (допустим, получилось 0,0025 – это много или мало?).

Другая распространённая метрика — MangE. Она характеризуется тем, что оценивает отличие форм спектров, то есть не штрафует, если предсказатель ошибся только в интенсивности. Однако проблема интерпретируемости по-прежнему присутствует, поскольку речь идёт об углах в многомерном пространстве.

Эксперименты

Источники данных. Ли и соавторы [19] в 2021 году представили набор данных из 409 гиперспектральных indoor- и outdoor-изображений разрешения 512×512 с 34 спектральными каналами от 400 до 730 нм с шагом 10 нм. Он содержит не спектры-стимулы, а приблизительные отражательные способности объектов в сцене. Это позволяет симулировать новые гиперспектральные изображения тех же сцен, но под другими источниками света. Таким образом, мы из 409 изображений получили 1636, используя для каждой картинке 4 стандартных источника: A18, F2, D65, E [20] (рис. 5).

Все пиксели набора не уместятся в оперативную память, поэтому из каждой картинке пиксели брались с шагом 12 по горизонтали и вертикали. Разделение на тренировочную и тестовую части бралось в отношении 4:1.

Чтобы проверить полученные фильтры на «переобучённость», был добавлен второй тестировочный набор данных — NTIRE2020 [7]. Он содержит 450 outdoor-изображений разрешения 512×482 с 31 спек-

тральным каналом (от 400 до 700 нм). На этот раз даны именно спектральные освещённости сенсора. Таким образом, набор содержит более однообразные данные, и именно поэтому был выбран в качестве тестового, а не тренировочного. По аналогии с набором данных Li для экономии оперативной памяти пиксели брались из сетки с шагом 7.

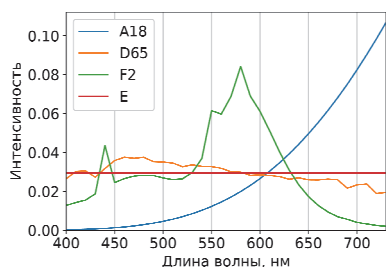


Рис. 5. Интенсивности использованных источников освещения

Для симуляции работы каждой из 4 камер использовались одни и те же чувствительности Canon 60D (рис. 6) [21].

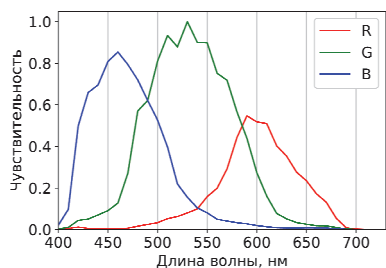


Рис. 6. Спектральные чувствительности камеры Canon 60D

Детали экспериментов. При подборе *Regression-trained*-фильтров шум на цветовых откликах предполагался гауссовским, поскольку это помогало ускорить оптимизацию. Во всех остальных случаях (подбор *FCN-trained*-фильтров, обучение алгоритмов восстановления) было решено прибегнуть к более реалистичному моделированию. Камера является по сути детектор фотонов, поэтому шум будет иметь две составляющие: независимую от уровня сигнала гауссовскую и зависимую от уровня сигнала пуассоновскую [22]. Вторую составляющую мы приближаем гауссовской случайной величиной с дисперсией, пропорциональной величине исходного сигнала.

Полностью все шаги численных экспериментов выглядят так:

1. Инициализация *Manual*-, *Random*- и *None*-фильтров.
2. Нахождение *Regression-trained*-фильтров путём решения (8).
3. Генерация откликов камеры для всех видов фильтров вместе с гауссовским шумом.
4. Обучение алгоритмов восстановления на полученных парах «отклик – спектр».

5. Оценка всех обученных алгоритмов на тестовой части датасета Li и на датасете NTIRE2020, подбор гиперпараметров.

Особняком стоял эксперимент с нахождением *FCN-trained*-фильтров путём обучения одновременно с сетью согласно задаче (5) и оценением полученной нейросети и фильтров на тех же датасетах.

Результаты

Фильтры, полученные в результате оптимизации (5) и (8), изображены на рис. 7 и 8.

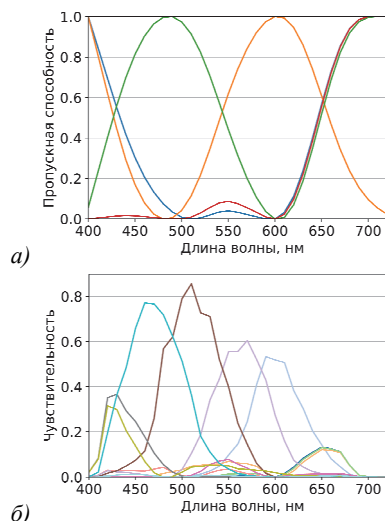


Рис. 7. Пропускные способности *Regression-trained*-фильтров (а) и чувствительности соответствующего 12-канального спектрального сенсора (б)

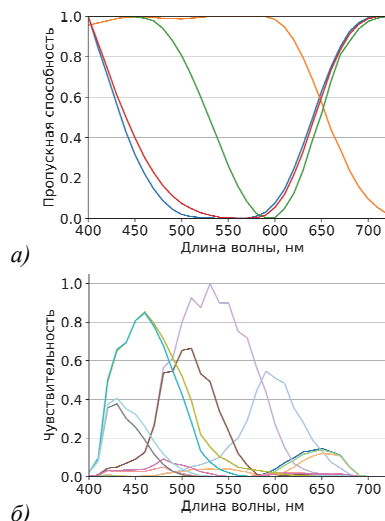


Рис. 8. Пропускные способности *FCN-trained*-фильтров (а) и чувствительности соответствующего 12-канального спектрального сенсора (б)

Результаты измерения всех трёх метрик на тестовых наборах данных представлены в табл. 1, 2 и 3.

Напомним, что *FCN-trained*-фильтры проверялись только в сочетании с алгоритмом восстановления FCN. Результат измерения метрик в этом случае представлен в табл. 4.

Табл. 1. Метрика NSE

		Regression-trained	Manual	Random	None
4*Li (тест)	Регрессия	23,3 %	35,6 %	34,3 %	31,4 %
	A+	19,8 %	29,3 %	28,5 %	26,5 %
	FCN	18,1 %	26,4 %	23,2 %	21,1 %
	FCN kernel	97,7 %	35,8 %	78,5 %	22,1 %
4*NTIRE	Регрессия	6,9 %	9,3 %	16,3 %	16,9 %
	A+	6,0 %	9,7 %	21,2 %	21,7 %
	FCN	6,4 %	10,1 %	25,1 %	28,2 %
	FCN kernel	90,2 %	10,9 %	50,6 %	26,1 %

Табл. 2. Метрика MangE

		Regression-trained	Manual	Random	None
4*Li (тест)	Регрессия	20,87	27,53	25,94	24,02
	A+	18,44	22,99	23,26	22,25
	FCN	16,14	20,27	18,79	17,54
	FCN kernel	64,63	31,41	50,75	17,96
4*NTIRE	Регрессия	6,77	8,82	13,37	13,71
	A+	6,01	8,24	15,82	16,20
	FCN	5,86	8,08	15,93	17,18
	FCN kernel	53,17	13,50	37,12	16,31

Табл. 3. Метрика RMSE

		Regression-trained	Manual	Random	None
4*Li (тест)	Регрессия	0,0049	0,0067	0,0071	0,0067
	A+	0,0047	0,0061	0,0064	0,0063
	FCN	0,0045	0,0056	0,0058	0,0056
	FCN kernel	0,1282	0,0074	0,0193	0,0056
4*NTIRE	Регрессия	0,0228	0,0279	0,0504	0,0532
	A+	0,0202	0,0297	0,0719	0,0729
	FCN	0,0210	0,0307	0,0870	0,0967
	FCN kernel	0,5776	0,0366	0,1335	0,0902

Табл. 4. Метрики для FCN-trained-фильтров

	MangE	NSE	RMSE
Li (тест)	19,23	22,1 %	0,0049
NTIRE	6,34	6,5 %	0,0220

Как видно, *Regression-trained*-фильтры показывают наилучшие результаты на каждом наборе данных. В сравнении с *None*-фильтрами, на датасете NTIRE удалось улучшить все метрики по меньшей мере в 3 раза. Наилучший алгоритм восстановления однозначно выбрать нельзя: на наборе данных Li лучше показывает себя FCN, а на NTIRE – алгоритм A+.

Любопытна форма чувствительностей полученного 12-канального сенсора (рис. 7). Она во многом следует принципу построения *Manual*-фильтров: каждая чувствительность имеет один пик, и все пики разные и почти равномерно распределены по видимому диапазону. Это значит, что при сочинении *Manual*-фильтров был найден правильный принцип и интуиция сработала хорошо.

Тем не менее, сами *Manual*-фильтры показывают в среднем самые худшие результаты. Это влияние шума: цветовые отклики при использовании *Manual*-фильтров оказываются малыми из-за маленькой интенсивности пропускания фильтров, а значит, они

больше подвержены негативному влиянию шума. Таким образом, неявно при поиске оптимальной формы фильтров с учетом зашумления учитывается и квантовая эффективность сенсора.

В целом можно сделать вывод, что на датасете Li фильтры расположились по качеству в порядке *Trained* > *None* > *Random* > *Manual*, а на датасете NTIRE: *Trained* > *Manual* > *Random* > *None*.

Заметим также, насколько неожиданно плохо работает алгоритм FCN-kernel. По своей задумке он должен был быть улучшенной версией FCN, однако показывает результаты намного хуже. Изучив, почему так происходит, мы обнаружили, что, если тестировать алгоритмы на бесшумных данных, результаты у FCN-kernel немного лучше, чем у FCN. Дело в том, что в случае FCN-kernel нейросеть учится восстанавливать составляющую спектра из подпространства – ядра преобразования 12-канального сенсора. Составляющая из дополнения к подпространству вычисляется аналитически: домножением на линейный оператор. Именно вторая составляющая и вносит негативный вклад в правильность восстановления спектра. Нейросеть была обучена на шумных данных, а линейный оператор – нет.

Интересно, что полносвязная сеть в сочетании с фильтрами, обученными под себя (*FCN-trained*), по-

казала результаты хуже, чем в сочетании с регрессионными фильтрами. По-видимому, при обучении сети одновременно с фильтрами мы имеем дело с задачей с большим количеством локальных минимумов, которые не дают нам найти глобальное решение. Также бросается в глаза чрезмерная гладкость *FCN-trained*-фильтров (рис. 8). Из-за того, что функция потерь и гладкая регуляризация – два слагаемых одного и того же оптимизационного функционала, оптимизатору Adam оказалось проще разгладить фильтры, чем уменьшать функцию потерь.

Заключение

Была рассмотрена задача восстановления гиперспектральных изображений из 12-канальных, которые, в свою очередь, были получены путём комбинирования 3-канальных изображений с 4 камер, на которые было наложено 4 разных фильтра. Математическая формулировка этой задачи позволила нам спроектировать фильтры и подобрать алгоритм восстановления так, что качество спектральной реконструкции улучшилось в три раза в сравнении с реконструкцией по обыкновенным 3-канальным изображениям. Процесс оптимизации фильтров был ускорен при помощи ряда математических приёмов. Было показано, что оптимизация фильтров позволяет добиться лучших результатов, чем проектирование фильтров интуитивно.

Также был выявлен недостаток идеи статьи Финлэйсона и соавторов [18]. Авторы работы не учитывали существенное влияние шума сигнала на точность восстановления спектра, что, как показывает практика, является важнейшим фактором. В связи с этим, мы предлагаем определять алгоритм восстановления спектра не как функцию, принимающую на вход цветовой отклик, а как функцию, принимающую на вход вероятностное распределение на пространстве цветовых откликов.

Дальнейшая работа будет нацелена на приближение условий вычислительного эксперимента к реальным. Требуется рассмотреть алгоритмы восстановления спектра, учитывающие пространственную информацию. Именно такие алгоритмы являются в настоящий момент state-of-the-art; к тому же они помогут нивелировать влияние шума в цветовых откликах, ведь соседние пиксели изображения будут иметь схожие средние значения. Планируется исследование вопроса прироста точности восстановления (в том числе спектров окрасок) от числа используемых камер, а также вопроса достижимой точности восстановления для различных уровней шума и битностей сенсоров. Отдельно большое внимание будет уделено задаче дальнейшей формализации требований к физически реализуемым фильтрам с учетом ограничений на точность восстановления сигнала и стоимость изготовления.

References

- [1] Leon R, Martinez-Vega B, Fabelo H, et al. Non-invasive skin cancer diagnosis using hyperspectral imaging for in-situ clinical support. *J Clin Med* 2020; 9(6): 1662. DOI: 10.3390/jcm9061662.
- [2] Fabelo H, Halicek M, Ortega S, et al. Deep learning-based framework for *in vivo* identification of glioblastoma tumor using hyperspectral images of human brain. *Sensors* 2019; 19(4): 920. DOI: 10.3390/s19040920.
- [3] Underwood E, Ustin S, DiPietro D. Mapping nonnative plants using hyperspectral imagery. *Remote Sens Environ* 2003; 86(2): 150-161. DOI: [https://doi.org/10.1016/S0034-4257\(03\)00096-8](https://doi.org/10.1016/S0034-4257(03)00096-8).
- [4] Chu X, Miao P, Zhang K, Wei H, Fu H, Liu H, Jiang H, Ma Z. Green banana maturity classification and quality evaluation using hyperspectral imaging. *Agriculture* 2022; 12(4): 530. DOI: 10.3390/agriculture12040530.
- [5] Specim, Spectral Imaging Ltd. Case study: Picvisa – machine vision solutions for waste treatment and recycling. 2020. Source: <https://www.specim.com/case-study-picvisa-machine-vision-solutions-for-waste-treatment-recycling>.
- [6] Specim, Spectral Imaging Ltd. Spectral imaging in machine vision solutions. 2022. Source: <https://www.specim.com/industry>.
- [7] Arad B, Timofte R, Ben-Shahar O, et al. NTIRE 2020 challenge on spectral reconstruction from an RGB image. 2020 IEEE/CVF Conf on Computer Vision and Pattern Recognition Workshops (CVPRW) 2020: 1806-1822. DOI: 10.1109/CVPRW50498.2020.00231.
- [8] Aeschbacher J, Wu J, Timofte R. In defense of shallow learned spectral reconstruction from RGB images. 2017 IEEE Int Conf on Computer Vision Workshops (ICCVW) 2017: 471-479. DOI: 10.1109/ICCVW.2017.63.
- [9] Shi Z, Chen C, Xiong Z, Liu D, Wu F. HSCNN+: Advanced CNN-based hyperspectral recovery from RGB images. 2018 IEEE/CVF Conf on Computer Vision and Pattern Recognition Workshops (CVPRW) 2018: 939-947. DOI: 10.1109/CVPRW.2018.00139.
- [10] Cai Y, Lin J, Lin Z, Wang H, Zhang Y, Pfister H, Timofte R, Van Gool L. MST++: Multi-stage spectral-wise transformer for efficient spectral reconstruction. 2022 IEEE/CVF Conf on Computer Vision and Pattern Recognition Workshops (CVPRW) 2022: 745-755. DOI: 10.1109/CVPRW56347.2022.00090.
- [11] Helling S, Seidel E, Biehlig W. Algorithms for spectral color stimulus reconstruction with a seven-channel multispectral camera. *CGIV 2004: The Second European Conf on Colour Graphics, Imaging and Vision 2004*; 2: art00051. DOI: 10.2352/CGIV.2004.2.1.art00051.
- [12] Valero EM, Nieves JL, Nascimento SMC, Amano K, Foster DH. Recovering spectral data from natural scenes with an RGB digital camera and colored filters. *Color Res Appl* 2007; 32(5): 352-360. DOI: 10.1002/col.20339.
- [13] Finlayson GD, Zhu Y. Designing color filters that make cameras more colorimetric. *IEEE Trans Image Process* 2020; 30: 853-867. DOI: 10.1109/TIP.2020.3038523.
- [14] Ng D-Y, Allebach JP. A subspace matching color filter design methodology for a multispectral imaging system. *IEEE Trans Image Process* 2006; 15(9): 2631-2643. DOI: 10.1109/TIP.2006.877384.
- [15] Nie S, Gu L, Zheng Y, Lam A, Ono N, Sato I. Deeply learned filter response functions for hyperspectral reconstruction. 2018 IEEE/CVF Conf on Computer Vision and Pattern Recognition 2018: 4767-4776. DOI: 10.1109/CVPR.2018.00501.

- [16] OMDIA. Smartphone oems shifting to higher camera resolution configurations with fewer lenses in response to continued semiconductor shortage, finds new Omdia research. 2021. Source: <<https://omdia.tech.informa.com/pr/2021-dec/smartphone-oems-shifting-to-higher-camera-resolution-configurations-with-fewer-lenses-in-response-to-continued-semiconductor-shortage-finds-new-omdia-research>>.
- [17] Foster DH, Amano K, Nascimento SMC, Foster MJ. Frequency of metamerism in natural scenes. *J Opt Soc Am A* 2006; 23(10): 2359-2372. DOI: 10.1364/JOSAA.23.002359.
- [18] Lin Y-T, Finlayson G. Physically plausible spectral reconstruction. *Sensors* 2020; 20(21): 6399. DOI: 10.3390/s20216399.
- [19] Li Y, Fu Q, Heidrich W. Multispectral illumination estimation using deep unrolling network. 2021 IEEE/CVF Int Conf on Computer Vision (ICCV) 2021: 1-8. DOI: 10.1109/ICCV48922.2021.00267.
- [20] Colorimetric illuminants. 2024. Source: <<https://cie.co.at/publications/colorimetric-illuminants>>.
- [21] Rochester Institute of Technology. Camera spectral sensitivity database. 2013. Source: <<http://www.gujinwei.org/research/camspec/db.html>>.
- [22] Shepelev DA, Bozhkova VP, Ershov EI, Nikolaev DP. Simulating shot noise of color underwater images. *Computer Optics* 2020; 44(4): 671-679. DOI: 10.18287/2412-6179-CO-754.

Сведения об авторах

Реутский Даниил Александрович, 1999 года рождения, студент 2-го курса магистратуры МФТИ.

Погадаев Артём Васильевич, 2001 года рождения, студент 4-го курса бакалавриата МФТИ.

Владимиров Даниил Михайлович, 2002 года рождения, студент 3-го курса бакалавриата МФТИ.

Ершов Егор Иванович, 1990 года рождения, старший научный сотрудник ИППИ РАН, кандидат физико-математических наук.

ГРНТИ: 28.23.15: Распознавание образов. Обработка изображений
Поступила в редакцию 07 апреля 2023 г. Окончательный вариант – 14 марта 2024 г.

Towards using smartphones as hyperspectral cameras

D. Reutsky^{1,2}, A. Pogadaev^{1,2}, D. Vladimirov^{1,2}, E. Ershov^{1,2}

*¹ Institute for Information Transmission Problems (Kharkevich Institute),
127051, Moscow, Bolshoy Karetny per. 19, build. 1,*

*² Moscow Institute of Physics and Technology (National Research University),
141701, Russia, Dolgoprudny, Moscow Region, Institutskiy per. 9*

Abstract

Spectral reconstruction (SR, recovering spectra from RGB measurements) is a vital problem of computational photography. However, even recent algorithms do not provide acceptable accuracy. As a matter of curiosity, modern mobile devices open a new opportunity to improve the quality of SR by utilizing images from several cameras at once. This leads to the idea of creating a mobile hyperspectral camera for the general public. In this paper we investigate the achievable accuracy when using several identical cameras simultaneously in combination with different spectral filters. To find optimal filters, two algorithms are proposed: one learns linear regression for SR and the other learns fully connected neural network. As a result of numerical experiments, it is found that in the case of 4 cameras and 4 filters, the SR accuracy is three times higher than in the case of SR from RGB images.

Keywords: spectrum, hyperspectral imaging, hyperspectral reconstruction.

Citation: Reutsky DA, Pogadaev AV, Vladimirov DM, Ershov EI. Towards using smartphones as hyperspectral cameras. *Computer Optics* 2025; 49(1): 112-120. DOI: 10.18287/2412-6179-CO-1315.

Authors' information

Daniil Reutsky, born 1999, 2nd year Master's student at MIPT.

Artem Pogadaev, born 2001, 4th year Bachelor's student at MIPT.

Daniil Vladimirov, born 2002, 3th year Bachelor's student at MIPT.

Egor Ershov, born 1990, Senior Researcher at IITP RAS, PhD.

Received April 07, 2023. The final version – March 14, 2024.
