

Разработка алгоритма обнаружения искусственных искажений цифровых изображений с использованием дескриптора FISH

А.А. Ланин¹, А.В. Кузнецов¹

¹Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086

Аннотация. Поддельные изображения становятся всё более распространёнными в современном мире. Влияя на наше мнение о человеке, они могут причинить немалый ущерб. Для определения таких изображений необходимы автоматические детекторы. В данной статье представлен метод автоматического обнаружения сплайсинга с помощью компьютерного зрения, основанный на сравнении параметров освещённости лиц на одном изображении.

1. Введение

Сплайсинг изображений лиц (face splicing) – это искусственное искажение изображения, при котором в оригинальное изображение производится вставка лица, вырезанного из другого изображения [1, 2]. В современном мире с широким распространением социальных сетей и средств обмена цифровыми изображениями такие подделки могут нанести вред репутации человека. Кроме того, благодаря прогрессу в области разработки программного обеспечения для обработки изображений малоквалифицированные пользователи могут без труда создать подделку, которую крайне сложно отличить от оригинала невооружённым глазом. Пример сплайсинга изображён на рисунке 1.



Рисунок 1. Пример сплайсинга (прямоугольником выделено лицо добавленного человека).

По этим причинам появилась необходимость в создании автоматических детекторов, которые позволили бы отличить поддельное изображение от настоящего. Исследователи из Флорентийского университета [3] представили модель одного из таких детекторов, основанного на использовании дескриптора FISH (Face intensity-shape histogram).

У данной модели есть ряд особенностей, отличающих её от других подобных алгоритмов:

- Использование гистограмм, доказавших свою эффективность во многих задачах компьютерного зрения.
- Зависимость только от локальных статистик изображения, без акцента на конкретную математическую модель, что делает алгоритм эффективнее других при работе с реальными изображениями.
- Более быстрые вычисления благодаря использованию гистограмм и возможности регулирования размера векторов признаков, которые вычисляются с использованием гистограмм.

На основе представленной модели можно построить автоматический детектор сплайсинга с использованием дескрипторов FISH, который можно разделить на 3 части:

- Детектирование лица.
- Построение 3D модели лица и вычисление векторов-нормалей.
- Извлечение дескрипторов FISH и их сравнение.

Структура детектора изображена на рисунке 2.



Рисунок 2. Структура автоматического детектора сплайсинга.

2. Детектирование лица

При реализации алгоритма для детектирования лица использовалась библиотека dlib из Python. На рисунке 3 показано исходное изображение. На первом этапе извлекаются области изображения, в которых были найдены лица (в данном случае их будет 2). Далее для каждого лица строится 68 опорных точек (Рисунки 4, 5), которые необходимы для построения 3D модели лица на следующем этапе.



Рисунок 3. Исходное изображение.



Рисунок 4. Извлечённое изображение лица 1 с опорными точками.



Рисунок 5. Извлечённое изображение лица 2 с опорными точками.

3. Построение 3D модели лица и вычисление векторов-нормалей

Для построения 3D модели лица и вычисления векторов-нормалей к поверхности лица использовалась 3DMM (3D Morphable Model). По 68 опорным точкам строится 3D модель лица (Рисунок 6), а по ней находятся нормальные векторы. Каждому нормальному вектору ставится в соответствие пиксель изображения. Матрицы пикселей и нормальных векторов передаются на следующий модуль.



Рисунок 6. 3D модель лица.

4. Извлечение дескрипторов FISH и их сравнение

При использовании дескрипторов FISH мы допускаем, что поверхность является выпуклой, с постоянной отражательной способностью, подчиняющейся закону Ламберта (дальнейшее улучшение разработанной версии связано с использованием других математических моделей освещенности, а именно модель Фонга и модель Блинна-Фонга, которые позволят применять), а источники света находятся далеко от сцены, тогда значения интенсивности пикселей изображения будут зависеть только от нормальных векторов к поверхности лица. Поэтому сначала из изображения удаляются области, не подходящие под данные допущения (шея, уши, губы, глаза и брови, а также слишком светлые и слишком тёмные области кожи). В оригинальной работе [3] авторы используют яркостный диапазон для обрезания нужных областей лица по порогу. В данной работе используется свёрточная нейронная сеть, позволяющая выполнять детектирование отдельных элементов лица [4]. Далее строится полусфера, аппроксимирующая полусферу, которая содержит 305 ячеек (Рисунок 7). Эти ячейки квантуют векторы-нормали 3D модели лица и для каждого цветового канала считаются через Гауссову ядерную оценку плотности по формуле:

$$I_i = L(n_i) \quad (1)$$

Пусть X_k - точка 3D модели, тогда нормальный вектор, соответствующий ей:

$$\hat{n}_k = N(X_k) \quad (2)$$

Если x_k — пиксель изображения, являющийся проекцией X_k , то значение интенсивности в этом пикселе:

$$\hat{I}_k = I(x_k) \quad (3)$$

Также рассчитаем распределение Гаусса:

$$z_{ik} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{\arccos(n_i \cdot \hat{n}_k)}{2\sigma}\right)^2} \quad (4)$$

Далее мы можем рассчитать веса с некоторым порогом τ_k :

$$\omega_{ik} = \begin{cases} z_{ik}, & \text{если } z_{ik} > \tau_k \\ 0, & \text{иначе} \end{cases} \quad (5)$$

$$\omega_i = \sum_k \omega_{ik} \quad (6)$$

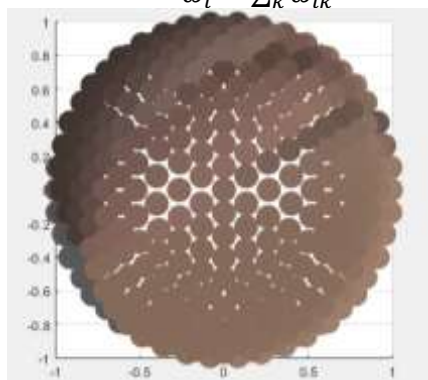


Рисунок 7. Полусфера с ячейками.

Теперь формулу (1) можно записать так:

$$I_i = \sum_k \frac{\omega_{ik}}{\omega_i} \hat{f}_k \quad (7)$$

Наконец, соединяя все три цветовых канала, получаем дескриптор FISH:

$$I = [L_R(n_i) \ L_G(n_i) \ L_B(n_i)]^T \quad (8)$$

Эти дескрипторы используются для сравнения двух лиц на изображении: чем меньше вычисленное FISH-расстояние, тем больше вероятность, что лица находятся под одинаковым освещением, а значит, выше вероятность, что изображение является подлинным. Сравнение происходит с помощью евклидовой нормы по следующей формуле:

$$D(a, b) = (\sum_i \|I_i^a - I_i^b\|)^{\frac{1}{2}} \quad (9)$$

Для того чтобы исключить влияние цвета кожи на значение FISH-расстояния, можно перед вычислением нормализовать дескрипторы по среднему RGB значению изображения лица:

$$I_i = (L(n_i)) / \mu \quad (10)$$

Можно также использовать второй способ нормализации, при котором учитываются средние RGB значения изображений обоих лиц:

$$I_i^{a \rightarrow b} = \min(255, L^a(n_i) \frac{\mu_b}{\mu_a}) \quad (11)$$

$$D(a, b) = \min(D(I_i^a, I_i^{b \rightarrow a}), D(I_i^b, I_i^{a \rightarrow b})) \quad (12)$$

После рассмотрения представленного алгоритма можно отметить его основные недостатки:

- Для обнаружения сплайсинга необходимо, чтобы на изображении было как минимум 2 лица.
- Алгоритм плохо работает, если сцена на изображении сильно отклоняется от сделанных допущений.
- Разрешение изображения должно быть достаточно высоким для построения 3D модели.

5. Тестирование алгоритма

Для тестирования были взяты 4 изображения (2 из них со сплайсингом). В таблице 1 приведены значения вычисленных для них «FISH-расстояний», а первые два примера рассмотрим более наглядно.

Таблица 1. Значения FISH-расстояний для отдельных примеров.

Способ нормализации	Пример 1	Пример 2 (сплайсинг)	Пример 3	Пример 4 (сплайсинг)
1	113.7991	187.7722	97.624	132.5672
2	44.0585	63.463	40.8798	73.9201

5.1. Пример 1 (без слайсинга)



Рисунок 8. Расчёт FISH-расстояния для примера 1.

Для данного примера FISH-расстояние с нормализацией способом 1 составляет 113.7991, FISH-расстояние с нормализацией способом 2: 44.0585

5.2. Пример 2 (со слайсингом)

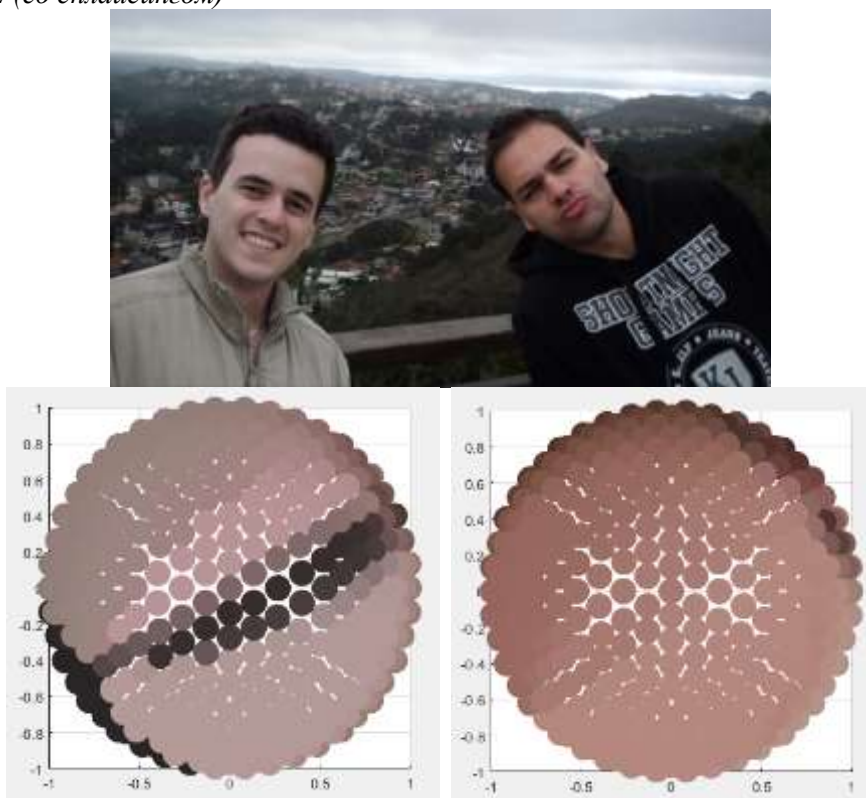


Рисунок 9. Расчёт FISH-расстояния для примера 2.

Для данного примера, содержащего искажения, FISH-расстояние с нормализацией способом 1 составляет 187.7722, FISH-расстояние с нормализацией способом 2 составляет 63.463.

6. Благодарности

Исследование выполнено при финансовой поддержке РФФИ в рамках научных проектов № 20-37-70053, 19-07-00138, 19-07-00474.

7. Литература

- [1] Piva, A. An overview on image forensics / A. Piva // *ISRN Signal Processing*, 2013. – P. 496701.
- [2] Mahdian, B. Using noise inconsistencies for blind image forensics / B. Mahdian, S. Saic // *Image Vis. Comput.* – 2009. – Vol. 27(10). – P. 1497-1503.
- [3] Fanfani, M. FISH: Face Intensity-Shape Histogram Representation for Automatic Face Splicing Detection / M. Fanfani, F. Bellavia, M. Iuliani, A. Piva, C. Colombo // *Journal of Visual Communication and Image Representation.* – 2019. – Vol. 63. – P. 102586.
- [4] Tanaka, M. Face Parts Detection // *MATLAB Central File Exchange*, 2020 [Electronic resource]. – Access mode: <https://www.mathworks.com/matlabcentral/fileexchange/36855-face-parts-detection>.

Development of an algorithm for detecting artificial distortion of digital images using the FISH descriptor

A.A. Lanin¹, A.V. Kuznetsov¹

¹Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086

Abstract. Fake images are becoming more common in the modern world. Influencing our opinion about a person, they may cause considerable damage. To detect such images, automatic detectors are needed. This article presents a method for automatic detecting splicing using computer vision, based on a comparison of the illumination parameters of faces in a single image.