

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
“КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ
СІКОРСЬКОГО”**

**ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ
КАФЕДРА ІНФОРМАЦІЙНОЇ БЕЗПЕКИ**

«До захисту допущено»

Завідувач кафедри

(підпис)
М. В. Грайворонський

“ _____ ” _____
(ініціали, прізвище) 2017 р.

Магістерська дисертація

за спеціальністю 8.04030101 «Прикладна математика»

на тему «Оптимізований метод реконструкції просторової конфігурації
людського обличчя в системах біометричної ідентифікації»

Виконав студент 6 курсу групи ФІ-51м

Кригін Валерій Михайлович

Керівник к.т.н., Барановський Олексій Миколайович

Консультант т. к. “Люксофт”, Сдобніков Віктор Юрійович

(підпис)

(підпис)

(підпис)

Засвідчую, що у цій дипломній роботі
немає запозичень з праць інших авторів
без відповідних посилань.

Студент _____

Київ — 2017 року

РЕФЕРАТ

Технический прогресс шагнул настолько далеко, что пространственную реконструкцию человеческого лица по снимку теперь возможно выполнить на обычном ноутбуке. Тем не менее, существующие методы можно улучшить как теоретически, так и практически.

Цель данной работы — постановка и решение задачи распознавания в терминах Байесовой теории распознавания, которая была предложена Теодором В. Андерсоном, и предоставить методы, которые позволят реализовать решение. Это даст возможность использовать корректно распознанные модели в системах биометрической идентификации.

Для достижения цели были использованы

- 1) Базелевская модель головы была использована как порождающая модель поверхности лица;
- 2) Байесовая теория распознавания образов была использована для правильной постановки и решения задачи и анализа существующих решений;
- 3) Численные методы были использованы для расчёта решения.

БАЙЕСОВАЯ СТРАТЕГИЯ, РАСПОЗНАВАНИЕ ОБРАЗОВ, ПРОСТРАНСТВЕННАЯ РЕКОНСТРУКЦИЯ, ПОРОЖДАЮЩАЯ МОДЕЛЬ, БИОМЕТРИЧЕСКАЯ ИДЕНТИФИКАЦИЯ

ABSTRACT

Technical progress has reached such performance, that inverse rendering of human face can be made on an average laptop for minutes. Though, existent methods can be improved theoretically and practically.

This study aims to formulate and solve the recognition problem in terms of Bayesian patterns recognition theory introduced by Theodore W. Anderson, and provide methods, which will help to implement the solution. This will allow to use correctly reconstructed models in biometric identification systems.

To perform the study

- 1) Basel morphable model was used as a generative model for the face surface;
- 2) Bayesian patterns recognition theory was used to formulate the problem correctly and analyse existent solutions;
- 3) Numerical methods were used to compute the solution.

BAYESIAN STRATEGY, PATTERNS RECOGNITION, INVERSE RENDERING, MORPHABLE MODEL, BIOMETRIC IDENTIFICATION

РЕФЕРАТ

Технічний прогрес досяг таких висот, що реконструювати просторову конфігурацію людського обличчя за знімком можливо на звичайному ноутбукі за кілька хвилин. Тим не менш, існуючі методи можна покращити як теоретично, так і практично.

Метою роботи є постановка та розв'язок задачі розпізнавання в термінах Баєсової теорії розпізнавання, що була запропонована Теодором В. Андерсоном, і вказати методи, які допоможуть реалізувати рішення програмно. Це допоможе використовувати коректно розпізнані моделі в системах біометричної ідентифікації.

Для досягнення мети було використано

- 1) Модель обличчя, зроблену в Базелівському університеті, в якості породжуючої моделі поверхні людського обличчя;
- 2) Баєсова теорія розпізнавання образів була використана для коректної постановки задачі і аналізу існуючих рішень;
- 3) Чисельні методи були використані для розрахунку рішення.

БАЄСОВА СТРАТЕГІЯ, РОЗПІЗНАВАННЯ ОБРАЗІВ, ПРОСТОРОВА РЕ-
КОНСТРУКЦІЯ, ПОРОДЖУЮЧА МОДЕЛЬ, БІОМЕТРИЧНА ІДЕНТИФІКА-
ЦІЯ

ЗМІСТ

Вступ	6
1 Попередні роботи	7
1.1 Приклади	7
1.1.1 Перша породжувальна модель обличчя	7
1.1.2 Модель Базелівського університету	8
1.1.3 Вітчизняна робота	10
1.1.4 Сучасні роботи	10
1.2 Постановка задачі	11
1.3 Питання щодо постановки задачі	13
1.4 Проблема у розв'язку задачі	13
1.5 Висновки	15
2 Теоретичні відомості	17
2.1 Породжувальна модель обличчя	17
2.1.1 Побудова	17
2.1.2 Використання	19
2.2 Задача	20
2.2.1 Початкові умови	20
2.2.2 Узагальнення на випадок кількох зображень	25
2.2.3 Баєсова задача розпізнавання	26
2.2.4 Бінарна функція витрат	30
2.2.5 Інтервальна функція витрат	31
2.2.6 Різниця моделей	32
2.2.7 Різниця параметрів	33
2.3 Розв'язок	35
2.3.1 Дискретизація простору параметрів	35
2.3.2 Інтервальна функція витрат	36

2.3.3	Різниця параметрів	39
2.3.4	Попередні результати в контексті Баєсової теорії розпізнавання . . .	41
2.4	Чисельні методи	42
2.4.1	Метод Монте-Карло	42
2.4.2	Методи мінімізації цільової функції	46
2.5	Висновки	48
3	Практичні результати	50
3.1	Портрет Чарльза Беббіджа	50
3.2	Висновки	52
	Висновки	53
	Перелік посилань	54

ВСТУП

Актуальність роботи. Примітивні системи біометричної ідентифікації особистості за фотознімком для коректної роботи потребують певних умов освітлення, положення обличчя на знімку, виразу обличчя та інших параметрів фотозйомки. Наразі ці обмеження вдається обходити за допомогою різних підходів, серед яких є використання насичених навчальних виборок при машинному навчанні, просторова реконструкція обличчя та розпізнавання опорних точок обличчя (ніс, рот, очі). Наступним етапом є подолання таких складно передбачуваних перешкод, як наявність косметики, аксесуарів (сережки, окуляри), волосся на обличчі (чубчик, борода, вуса). Деякі аксесуари можна зняти перед розпізнаванням, проте бороду та косметику складно прибирати кожен раз лише для того, щоб бути розпізнаним.

Об'єкт дослідження — обличчя.

Предмет дослідження — ідентифікація особистості за знімком обличчя.

Мета дослідження. Розробка оптимізованого методу реконструкції просторової конфігурації людського обличчя в системах біометричної ідентифікації.

Завдання наступні:

- 1) Дослідити існуючі методи реконструкції просторової конфігурації обличчя за одним або кількома зображеннями;
- 2) Проаналізувати існуючі методи з теоретичної точки зору.
- 3) Запропонувати кращий метод.

Практичне значення одержаних результатів.

1 ПОПЕРЕДНІ РОБОТИ

В даному розділі буде розглянуто коротку історію досліджень, що пов'язані з даною роботою. Потім слідує аналіз існуючих алгоритмів. В кінці розділу будуть висвітлені відкриті питання.

1.1 Приклади

1.1.1 Перша породжувальна модель обличчя

У 1972 році Фредерік Парке опублікував статтю, де було описано тривимірну модель обличчя [1]. Вона містила лише 250 полігонів та 400 вершин, проте могла імітувати різні людські емоції.

Модель з певним виразом обличчя створювалася за допомогою двох фотографій асистента: одне збоку і одне фронтальне. На обличчя асистента заздалегіть було нанесено необхідні полігони (рис. 1.1).

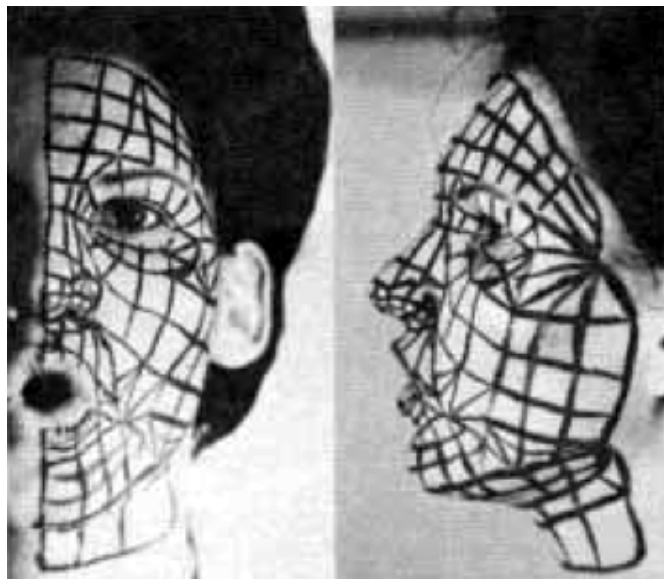


Рисунок 1.1 — Пара ортогональних фотографій асистента з нанесеними полігонами

Найпершою роботою зі створення тривимірної породжувальної моделі обличчя вважається дисертація Фредеріка Парке 1974 року [2]. Ця модель могла імітувати не тільки різні емоції, але й артикуляцію і різну форму обличчя (рис. 1.2).

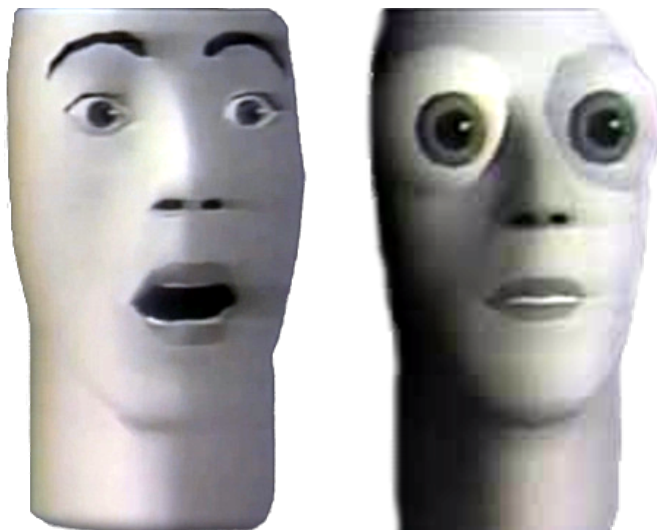


Рисунок 1.2 — Моделі з різною формою очей та виразами обличчя

Щоб оцінити приголомшливість цих результатів, пропоную згадати, коли з'явилися перші тривимірні відеоігри, персонажі яких мають анімовані об'ємні обличчя, а не лише анімовану текстуру на голові. Це не було можливо навіть у легендарній грі Quake, що вийшла у 1996 році, хоча технологія існувала вже понад 20 років.

1.1.2 Модель Базелівського університету

У даній роботі використовується породжувальна модель обличчя 3D Basel Face Model (BFM) розроблена командою Базельського університету [3]. Перший підхід, який дозволяє отримати тривимірну модель обличчя лише за одним фото, було запропоновано Томасом Феттером та Фолькером Бланцом [4]. Вони використовували BFM та стохастичний градієнтний спуск [5], що дозволяло досягти

добрих результатів за 40 хвилин з процесором Pentium III, 800MHz [6]. У статті продемонстровано результат реконструкції (рис. 1.3б) обличчя Тома Хенкса за допомогою одного кадру з кінофільму “Форрест Гамп” (рис. 1.3а).



Рисунок 1.3 — Демонстрація роботи алгоритму дослідників з Базелівського університету

Згодом ті ж автори розробили та використали у своїй роботі стохастичний алгоритм Ньютона [7], яким досягли реконструкції просторової конфігурації обличчя за 4.5 хвилини на Pentium 4, 2GHz.

Важливою деталлю цієї породжуючої моделі є те, що кожна вершина різних моделей має однакове семантичне значення. Краї очей, рота, кінчик носу, підборіддя та інші точки знаходяться в однакових комірках масивів вершин моделей. BFM розмічена опорними точками з наборів Farkas та MPEG4 FDP (рис. 1.4). Відповідність точок дає змогу створювати власні множини опорних точок.

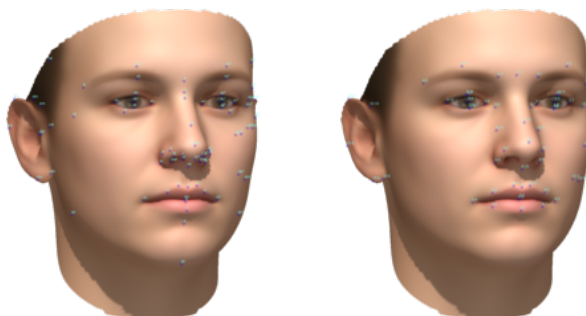


Рисунок 1.4 — Усереднені моделі облич з відміченими опорними точками

1.1.3 Вітчизняна робота

У 2011 році випускник Національного Технічного Університету України “Київський Політехнічний Інститут” запропонував та запатентував свій метод створення генеративної моделі обличчя та відновлення тривимірної поверхні обличчя за одним або кількома фото [8]. Основні відмінності полягають у тому, що задачу співставлення відповідних точок різних моделей було представлено та розв’язано як супермодулярну задачу розмітки [9], а нові моделі обличчя генеруються як зважене середнє. Проте було наявне спрощення: самозатінення не бралось до уваги, що дозволяло дуже швидко знаходити оптимальне освітлення за пласкою моделлю методом найменших квадратів.

1.1.4 Сучасні роботи

На момент написання дисертації одними з новітніх робіт, де було використано модель Базельського університету, є методи відстеження обличчя [10] та переносу виразу обличчя однієї людини іншій [11]. Докладніше про них буде сказано в наступному підрозділі.

1.2 Постановка задачі

Метою є мінімізація функції, що називається енергією. В першій роботі Бланца и Феттера [6] енергія складалася з суми квадратів різниць кольорів пікселів оригінального та синтезованого зображення, суми квадратів параметрів, що мають нормальний розподіл, та суми квадратів відхилень параметрів камери від початкових значень, що визначалися оператором:

$$E_N(x, \theta) = \sum_{i \in I' \subset I} \frac{\|y_i\|^2}{\sigma_N^2} + \sum_{j=1}^n x_j^2 + \sum_{k=1}^m \frac{\|\theta^k - \hat{\theta}^k\|^2}{\sigma_{\theta^k}^2}. \quad (1.1)$$

Дисперсія σ_N^2 відповідає шуму на оригінальному зображенні. Зауважимо, що беруться не всі пікселі зображення, а тільки ті, що відповідають пікселям, де намальовано модель. Щоб кількість доданків завжди була однаковою, на кожній ітерації алгоритму випадковим образом обирається певна кількість трикутників з ймовірністю, що пропорційна до їх площі.

Процедура мала наступний вигляд:

- 1) на перших кроках використовується модель з меншою кількістю трикутників та вершин;
- 2) спочатку використовуються лише параметри камери та найзначущі параметри моделі (перші головні компоненти), і поступово додаються нові компоненти;
- 3) спочатку σ_N обирається достатньо великим, аби збільшити апіорну ймовірність параметрів моделі, а з кожною ітерацією його зменшують, щоб отримувати все більш схоже зображення.

У новітніх роботах для кращих результатів реконструкції використовуються додаткові відомості про обличчя — опорні точки. Знаходження особливих

точок на зображенні голови являє собою окрему складну задачу. Одним з популярніших на даний час є метод, опублікований в статті “вирівнювання обличчя за одну мілісекунду”, що використовується в бібліотеці з відкритим кодом dlib [12]. Результати, що дає алгоритм, можна побачити на рис. 1.5.



Рисунок 1.5 — Приклади знаходження опорних точок обличчя

Окрім того, замість суми квадратів відхилень використовується зміщена оцінка дисперсії. Це дозволяє брати різну кількість точок на кожній ітерації. Нова цільова функція має вигляд

$$E(x, \theta) = \omega_c \cdot \sum_{i \in I' \subset I} \frac{y_i^2}{|I'|} + \omega_r \cdot \sum_{j=1}^n x_j^2 + \omega_l \cdot \sum_{l \in L} \frac{\Delta_l^2(x)}{|L|}, \quad (1.2)$$

де Δ_l — функція, яка перетворює параметри моделі на відхилення її опорних точок від реальних. Коефіцієнти ω_c , ω_r і ω_l обираються рівними 1, $2.5 \cdot 10^{-5}$ і 10 відповідно.

1.3 Питання щодо постановки задачі

Перше фундаментальне питання полягає в тому, навіщо потрібно мінімізувати саме такі функції, як було вказано вище. В роботі [6] вибір функції витрат (1.1) мотивується принципом оцінки апостеріорного максимуму. Проте функція (1.2) немає підґрунтя окрім того, що її мінімізація дає блискучі результати.

У формулі (1.2) є вагові коефіцієнти ω_c , ω_r і ω_l , походження яких не пояснюється. Також невідомо, чому значення мають бути самими такими, і чи є вони універсальними або залежать від вхідних даних.

У статтях зазначається, що енергія є складно влаштованою функцією, яка має багато локальних мінімумів. Для подолання цієї складності автори згадуваних статей пропонують модифікації класичних алгоритмів та цільової функції. Досліджень цих локальних екстремумів не було знайдено: за якими параметрами, за яких умов і наскільки глибокі локальні мінімуми — можливо, ця інформація допомогла б знайти рішення, яке допоможе відрізнити локальний мінімум від глобального.

1.4 Проблема у розв'язку задачі

У даному підрозділі буде розглянута проблема яке з першого погляду може здатися неочевидною, проте стає досить явною при перегляді простих прикладів.

Задача, що розв'язується у згадуваних роботах, може бути описана в загальному вигляді як

$$(k^*, \theta^*) = \arg \min_{k \in K, \theta \in \Theta} E(k, \theta), \quad (1.3)$$

де k — набір параметрів, що є виходом алгоритму (форма та, у більшості випадків, текстура), θ — набір параметрів, що необхідні для генерації моделі, проте не є частиною виходу алгоритму (матриця повороту, перспективи, переносу тощо).

Шукаються найкращі k^* і θ^* , а потім відкидається θ^* і вважається, що відповіддю є k^* .

Такий підхід може здатися логічним. Модель обличчя позиціонується максимально близько до того, як розташоване реальне обличчя на пред'явленому зображенні. Далі шукається найкраща форма (і текстура). Вважається, що отримані значення є бажаним результатом, бо немає сенсу робити так, щоб на вихід впливали ті діляки зображення, на яких обличчя немає. Проте звідки відомо, що обличчя насправді позиціоновано саме так, як вказав алгоритм пошуку опорних точок? Також немає гарантії, що тривимірні моделі були розмічені абсолютно точно. Додаткову складність додає питання щодо кількості і характеру опорних точок, які дадуть максимально точний результат. Виникають такі ситуації, коли одній конфігурації опорних точок відповідає різне положення обличчя. Простіший приклад: наявність проекції лише однієї опорної точки не дає представлення щодо повороту та розміру обличчя.

Розглянемо конкретну ситуацію, коли підхід (1.3) є хибним. На рис. 1.6а зображено модель, що була згенерована за допомогою BFM. На рис. 1.6б зображено ту ж саму модель, проте до неї було застосовано матрицю перспективи. Оскільки матриця перспективи — афінне перетворення, отже оборотне, обидва зображення можуть грати роль як отриманого результату, так і результату з іншою матрицею перспективи. Нехай в кінці роботи алгоритму було отримано зображення 1.6б — обличчя кремезної людини з широким носом та вилицями і низьким лобом. Слідуючи рекомендаціям, ми зберігаємо лише параметри форми моделі та відкидаємо отриману матрицю. В результаті маємо довгасте обличчя худі людини з високим лобом.

В такому випадку можна вирішити, що перетворення, які з інтуїтивної точки зору змінюють зовнішній вигляд моделі (масштаб, перспектива, скіс і таке інше), можна зберігати разом із параметрами моделі як вихідний результат. Проте не треба забувати, що ці перетворення були внесені в задачу, щоб врахувати



(а) Візуалізована модель людського
обличчя

(б) Візуалізація з іншою матрицею
перспективи

Рисунок 1.6 — Приклад візуалізації однієї моделі з різними параметрами камери такі природні фактори, як спотворення зображення різними об'єктивами фотокамер (зокрема перспектива). Тому такий підхід теж є хибним.

У статті [13] було доведено, що стратегія, яка використовує (1.3), є негодящою, бо можна знайти іншу стратегію, яка буде давати кращі результати при будь-яких значеннях θ .

1.5 Висновки

Запропоновані методи досить складно порівняти з практичної точки зору: для цього треба реалізувати їх на одній мові програмування та перевірити на одній тестовій виборці на однакових обчислювальних машинах. Проте було вказано математичні неточності та неясності, що були знайдені при вивченні даних методів.

Щоб проаналізувати доцільність алгоритмів з точки зору статистичної теорії розпізнавання, в роботі буде запропоновано кілька постановок задачі, їх розв'язки та модифікації/спрощення, що дозволять отримати методи, які були запропоновані у розглянутих публікаціях. На основі цих даних можна буде проаналізувати якість існуючих алгоритмів з теоретичної точки зору.

2 ТЕОРЕТИЧНІ ВІДОМОСТІ

Даний розділ присвячено постановці та рішення задач з теоретичної точки зору в термінах Бассівської теорії розпізнавання образів. Спочатку буде введено основні поняття та позначення. Далі слідують конкретні постановки задач. Після цього наведено рішення наведених задач з теоретичної точки зору. Потім йдуть конкретні методи, що допоможуть знайти рішення.

2.1 Породжувальна модель обличчя

2.1.1 Побудова

За допомогою високоточного 3D сканеру було зафіксовано 100 чоловічих і 100 жіночих облич. Це люди віком від 8 до 62 років (25 в середньому), вагою від 40 до 123 кілограмів (66 в середньому).

Щоб зважена сума облич теж була обличчям, потрібно знайти відповідність між вершинами моделей різних облич. Тобто, оскільки модель складається з впорядкованого набору вершин, точка, яка відповідає, наприклад, за горбинку на носі, повинна бути на одній і тій самій позиції в масиві кожної моделі. Для цього була використана модифікація ітеративного алгоритму найближчих точок ІСР [14]. Після цього достатньо відмітити опорні точки обличчя лише на одній моделі, щоб вони були відомі на всіх інших, у тому числі на похідних обличчях.

Розраховується середня модель обличчя як середнє арифметичне по всіх моделям

$$\bar{f}_v = \frac{\sum_{f \in F} f_v}{|F|}, \quad v \in V,$$

де V — множина вершин, F — множина облич, f_v — координати вершини v в обличчі f . Середня модель теж є обличчям завдяки попередньому кроку спів-

ставлення вершин.

Останній етап, який нас цікавить, це обробка даних за допомогою методу головних компонент [15]. На вході матриця, кожному стовбцю якої відповідає координата вершини, а стрічці — обличчя. На перетині f стрічки і v_i стовбця знаходиться значення i компоненти координат вершини v обличчя f

$$M = \begin{bmatrix} M_{v_x^1}^{f^1} & M_{v_y^1}^{f^1} & M_{v_z^1}^{f^1} & M_{v_x^2}^{f^1} & \dots & M_{v_y^m}^{f^1} & M_{v_z^m}^{f^1} \\ M_{v_x^1}^{f^2} & M_{v_y^1}^{f^2} & M_{v_z^1}^{f^2} & M_{v_x^2}^{f^2} & \dots & M_{v_y^m}^{f^2} & M_{v_z^m}^{f^2} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ M_{v_x^1}^{f^n} & M_{v_y^1}^{f^n} & M_{v_z^1}^{f^n} & M_{v_x^2}^{f^n} & \dots & M_{v_y^m}^{f^n} & M_{v_z^m}^{f^n} \end{bmatrix}.$$

На виході маємо дисперсії головних компонент і лінійне перетворення U коваріаційної матриці \widetilde{M} , яке робить $U \cdot \widetilde{M} \cdot U^{-1}$ діагональною.

Оскільки маємо головні компоненти, нове обличчя отримується не як зважене середнє кількох облич, а як сума середнього обличчя та вектору параметрів x помноженого на матрицю U

$$M(x) = \overline{f} + x \cdot U.$$

Переваги головних компонент:

- 1) для зменшення обчислювальних витрат можна обирати не всі параметри, а лише кілька перших, бо вони мають найбільшу дисперсію і несуть в собі більшу частину інформації;
- 2) згідно з центральною граничною теоремою розподіл параметрів вважається нормальним, що зручно при моделюванні та обчисленнях:
 - а) фактично відсутні обмеження на параметри моделі на відміну від

- зваженого середнього, де сума параметрів повинна дорівнювати 1;
- б) за побудовою випадкові величини мають нульове середнє значення та діагональну коваріаційну матрицю;
- в) якщо нормувати параметри (перенести їх дисперсію в матрицю U), отримаємо набір незалежних випадкових величин розподілених за стандартним нормальним законом.

2.1.2 Використання

Введемо множину вершин обличчя V . Кожна вершина має певні координати в тривимірному просторі \mathbb{R}^3 . Модель обличчя — відображення, яке кожній вершині v ставить у відповідність її координати

$$M : V \rightarrow \mathbb{R}^3.$$

Породжувальна модель обличчя — відображення, яке кожному набору параметрів x ставить у відповідність модель m

$$G : X \rightarrow M.$$

Координати g вершини v моделі, згенерованої з параметрами x , позначимо

$$G_v(x) = m_v = g.$$

Координати кожної вершини v породжувальної моделі отримуються шляхом перемноження компонент параметру x на відповідний коефіцієнт λ^v , отри-

маний за допомогою методу головних компонент, та додавання результату до середнього положення поточної вершини

$$G_v(x) = g_0^v + \sum_{i \in 1}^n \lambda_i^v \cdot x_i, \quad v \in V.$$

2.2 Задача

2.2.1 Початкові умови

2.2.1.1 Зображення

Позначимо множину T зображень. Кольори — відтінки сірого, що визначаються лише інтенсивністю від 0 до 1. Введемо множину I пікселів зображення. Зображення $t \in T$ є відображення з множини пікселів на множину їх значень

$$t : I \rightarrow [0; 1].$$

Інтенсивність пікселя i в зображенні t позначимо як t_i .

Взагалі кажучи, I — множина індексів матриць однакового розміру

$$I = \{\langle i, j \rangle \mid i = 1..h, j = 1..w\}.$$

Зазвичай використовуються зображення розміром від $100 \times 100 = 10^4$ пікселів. Проте сучасні камери на фотоапаратах, смартфонах та інших пристроях можуть

зробити зображення площею кілька мільйонів пікселів і більше. При використанні $2^8 = 256$ градацій сірого маємо $2^{8 \cdot 10^4} \approx 10^{24 \cdot 10^3}$ різних зображень розміром 100×100 , тобто неймовірно багато.

Тривимірна модель обличчя визначається не тільки набором n дійсних параметрів з множини $X = \mathbb{R}^n$. Є додаткові параметри, що відіграють важливу роль при візуалізації, проте не потрібні в результаті — поворот, масштаб, перспектива тощо. Введемо відображення $\theta^M \in \Theta^M$, що застосовує необхідні перетворення до точки в тривимірному просторі

$$\theta^M : \mathbb{R}^3 \rightarrow \mathbb{R}^3.$$

Функцією, що перетворює набір параметрів на зображення, є відображення

$$f : \Theta^M \times X \rightarrow T.$$

Зображення t , згенероване з параметрами θ^M та x , позначатимемо

$$f_{\theta^M}(x) = t.$$

В подальших записах індекс θ^M записувати не будемо, коли з контексту буде зрозуміло, які перетворення застосовуються до моделі.

Для подальших обчислень введемо функцію h (hidden), яка приймає значення 0, якщо в даній точці зображення є піксель обличчя, та дорівнює 1 у протилежному випадку

$$h : \Theta^M \times X \rightarrow \{0, 1\}.$$

Введемо короткий запис для інверсії маски

$$\overline{h}_i^{\theta^M}(x) = 1 - h_i^{\theta^M}(x), \quad i \in I.$$

2.2.1.2 Опорні точки

Оскільки відповідні вершини всіх моделей мають однакове семантичне значення, за рахунок чого простір згенерованих облич є опуклим, можна скористатися інформацією деяких особливих з точки зору людини точок. Такі точки як краї губ, носа, очей і таке інше назвемо опорними точками. Множину обраних точок в контексті певної задачі позначатимемо

$$L \subset V.$$

Положення u опорної точки $l \in L$ у моделі m позначимо

$$m_l = u.$$

Відображення тривимірної точки, до якої застосовано перетворення θ^M ,

$$P_2 : \Theta^M \times \mathbb{R}^3 \rightarrow I.$$

Проекцію p точки v моделі m на площину запишемо

$$P_{\theta^M}^2(m_v) = p.$$

Також, коли з контексту зрозуміло значення θ^M , позначатимемо без нижнього індексу

$$P^2(m_v) = p.$$

Введемо відображення θ^L , що для певної опорної точки l дає координати цієї точки на даному зображенні t

$$\theta^L : L \times T \rightarrow I.$$

Координати опорної точки l на зображенні t позначимо

$$\theta_l^L(t) = i.$$

Аргумент використовувати не будемо, коли з контексту буде зрозуміло, про яке зображення йде мова.

Потрібна метрика, що буде відображати віддаленість двох точок на множині пікселів зображення. Той факт, що дві точки i і i' знаходяться одна від одної на відстані μ , запишемо як

$$\|i - i'\| = \mu.$$

Оскільки i — впорядкована пара чисел, що відповідають координатам певного пікселя, природнім чином можна впровадити звичайну евклідову метрику для розрахунку відстані між двома точками зображення.

2.2.1.3 Фон

Якщо зображення складається не лише з пікселів обличчя, обов'язково є фон. Він може бути відомим, наприклад, якщо заздалегіть було отримано зо-

браження фону, на якому буде відбуватися фотозйомка. У загальному випадку він є невідомою величиною, розподіл якої також невідомий, а може не існувати зовсім. Позначимо його як зображення θ^B

$$\theta^B \in T.$$

Фон має такий самий розмір, як і вхідне зображення.

2.2.1.4 Шум

Вважаємо, що на вхідне зображення накладено шум η , що є вектором незалежних випадкових величин, розподілених за центрованим нормальним законом з однаковою невідомою дисперсією σ_t^2 . Інтенсивність i пікселя

$$t_i = f_i(x) \cdot \bar{h}_i(x) + \theta_i^B \cdot h_i(x) + \eta_i, \quad \eta_i \sim \mathcal{N}(0, \sigma_t^2), \quad i \in I.$$

Замість реального положення опорних точок θ^L будемо розглядати її оцінку $\hat{\theta}^L$, яка отримана з похибкою, що має центрований гаусовий розподіл з дисперсією σ_L^2

$$\hat{\theta}^L = \theta_t^L(l) + \zeta_l, \quad \zeta_l \sim \mathcal{N}(0, \sigma_L^2).$$

Позначимо відхилення проекції опорних точок моделі від їх оцінки

$$\Delta(x) = P_{\theta^M}^2(G_L(x)) - \hat{\theta}^L.$$

Ймовірність того, що буде пред'явлено зображення t з параметрами x при відомих перетвореннях θ^M і оцінці положень опорних точок $\hat{\theta}^L$, є сумісною

ймовірністю

$$p_{\theta} = \mathbb{P}_{\theta} \{f_{\theta^M}(\xi) + \eta = t, \zeta = \Delta(\xi), \xi = x\}.$$

З визначення умовної ймовірності випливає

$$p_{\theta} = \mathbb{P}_{\theta} \{f_{\theta^M}(\xi) + \eta = t, \zeta = \Delta(\xi) \mid \xi = x\} \cdot \mathbb{P}(\xi = x)$$

Позбавимось умовної ймовірності, замінивши ξ на x . Також розіб'ємо перший множник на дві ймовірності, бо ці події не залежать одна від одної

$$p_{\theta} = \mathbb{P}_{\theta^M}(\eta = t - f_{\theta^M}(x)) \cdot \mathbb{P}_{\theta} \{\zeta = \Delta(x)\} \cdot \mathbb{P}(\xi = x).$$

Далі будемо позначати сумісну ймовірність без θ , коли з контексту буде зрозуміло, які параметри використані

$$\mathbb{P}(t, x) = p_{\theta}.$$

2.2.2 Узагальнення на випадок кількох зображень

Пред'явлено кілька зображень

$$\vec{t} = \langle t_1, t_2, \dots, t_m \rangle,$$

на яких знаходиться обличчя однієї й тієї ж самої людини. Для спрощення вважаємо, що форма обличчя (і колір шкіри, якщо потрібно) на всіх зображеннях однаковий, тобто $x = \text{const} \in X$. Цього важко досягти, якщо обрати фотографії зняті з різних камер в різні пори року, проте ця вимога має місце, коли на вхо-

ді дано відео, фотографії з однієї фотосесії, фотографії зроблені в один момент з різних ракурсів, або розрізане зображення, де людину видно у рівних чистих дзеркалах. Для різних зображень t^j буде відрізнятися шум η^j та параметри зйомки θ^j .

Набір зображень генерується стохастичним автоматом [16]

$$\mathfrak{U}_x = \langle \Theta \times T \times X \cup \{\varepsilon_0, \varepsilon\}, p, \varepsilon_0, \{\varepsilon\} \rangle,$$

де ε_0 — початковий стан, який не містить ніякої інформації, а ε — кінцевий стан. Функція p визначає ймовірність того, що наступним набором параметрів буде θ_j та згенерується зображення t_j . Якщо зображення не залежать одне від одного (рис. 2.1), маємо ймовірності станів, що не залежать один від одного

$$\mathbb{P}_j(\theta^j, t^j, x) = p(\theta^j, t^j).$$

У випадку відео (рис. 2.2) ймовірність генерації наступних параметрів та зображення залежить від поточного стану автомата

$$\mathbb{P}_j(\theta^j, t^j, x) = p_j(\theta^j, t^j \mid \theta^{j-1}).$$

2.2.3 Баєсова задача розпізнавання

Поставимо Баєсову задачу розпізнавання. Для цього потрібно визначитися з функцією витрат [17]

$$W : X \times X \rightarrow \mathbb{R}.$$

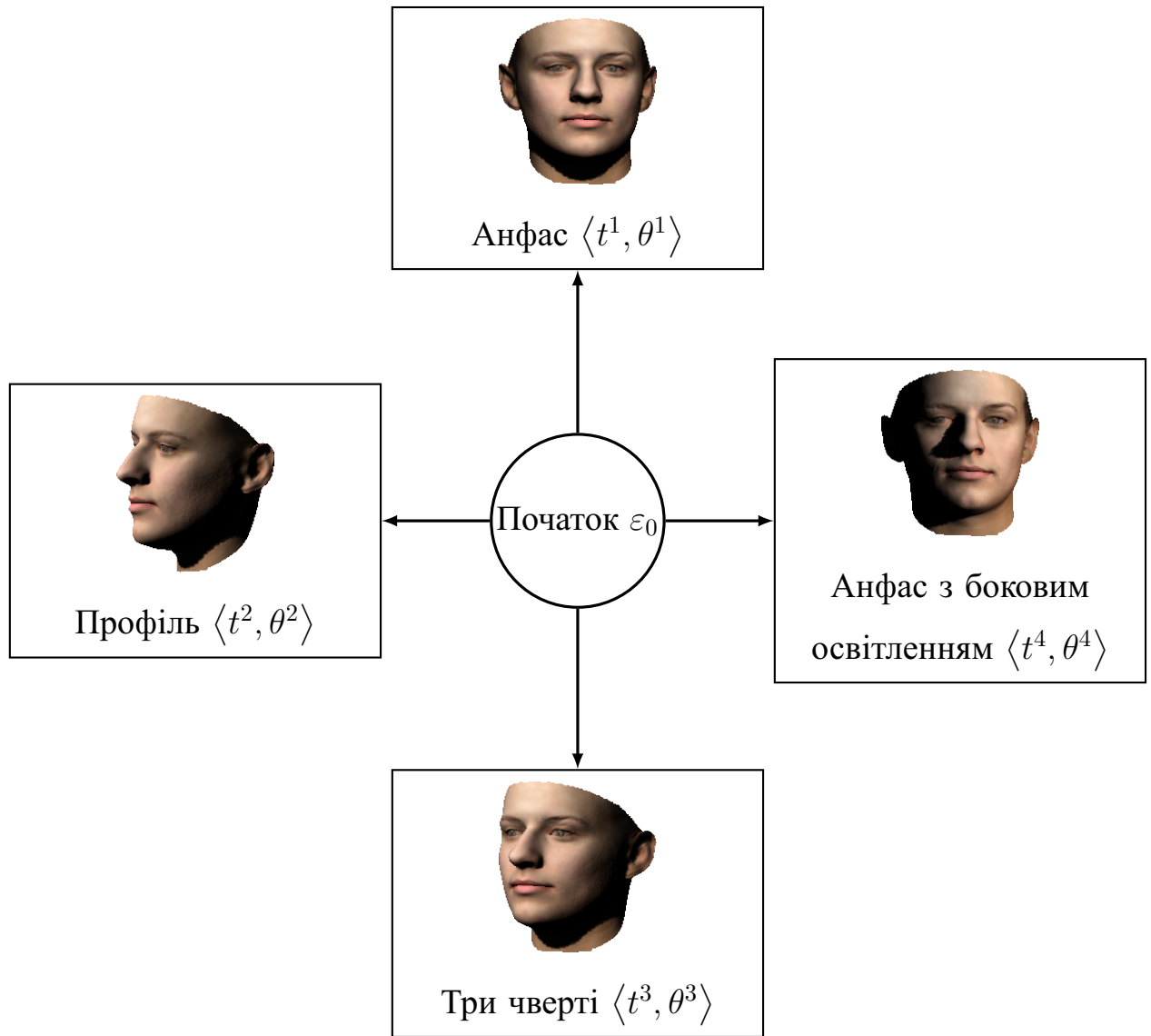


Рисунок 2.1 — Приклад роботи автомата з незалежними станами

Введемо множину стратегій розпізнавання Q як функцій, що кожному зображенню $t \in T$ та набору відображень $\theta \in \Theta$ ставлять у відповідність параметри, з якими було згенеровано обличчя на даному зображенні

$$Q = X^{\Theta \times T}.$$

Стратегію $q \in Q$

$$q : T \rightarrow X,$$

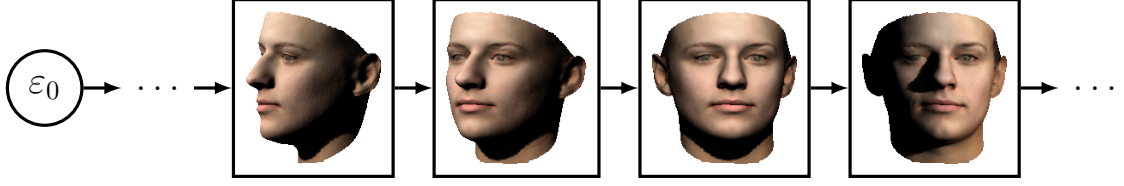


Рисунок 2.2 — Приклад автомата, що генерує відео

яка для зображення t дає результат x , позначимо

$$q(t) = x.$$

Математичне очікування функції витрат W як функції випадкової пари $\langle t, x \rangle$ для даного вирішального правила q називається Баєсовим ризиком [18]

$$R_{\theta}(q) = \sum_{t \in T} \int_X W(x, q(t)) dF_{\theta}(t, x) = \mathbb{E}_{\theta}[W(x, q(t))],$$

де F_{θ} — функція сумісного розподілу зображення та параметрів зображеної моделі при відомих θ .

Стратегія q називається баєсовою, якщо може бути представлена у вигляді [13]

$$q = \arg \min_{q \in Q} \sum_{\theta \in \Theta} \tau(\theta) \cdot R_{\theta}(q), \quad \tau(\theta) \geq 0, \quad \sum_{\theta \in \Theta} \tau(\theta) = 1. \quad (2.1)$$

Стратегії іншого вигляду є негодящими: якщо q_0 — стратегія, яку не можна представити у вигляді (2.1), обов'язково знайдеться Баєсова стратегія q така, що

$$R_{\theta}(q_0) > R_{\theta}(q), \quad \forall \theta \in \Theta.$$

Це означає, що при будь-якому значенні параметрів θ небаєсова стратегія буде розпізнавати гірше за певну баєсову.

Задача — знайти таке вирішальне правило q , за якого Бассів ризик $R_\theta(q)$ мінімальний для даної функції штрафу W при тих параметрах θ , що дають найбільший ризик [19]

$$q^* = \arg \min_{q \in Q} \max_{\theta \in \Theta} R_\theta(q).$$

Оскільки зображення t не залежать одне від одного, мінімізація і максимізація суми еквівалентна мінімізації і максимізації кожного її елементу

$$q^*(t) = \arg \min_{x' \in X} \max_{\theta \in \Theta} \int_X W(x, x') dF_\theta(t, x).$$

Можна переписати через умовне математичне очікування

$$q^*(t) = \arg \min_{x' \in X} \max_{\theta \in \Theta} \mathbb{E}_\theta [W(x, x') \mid t] \cdot \mathbb{P}(t).$$

Константний множник не впливає на розв'язок

$$q^*(t) = \arg \min_{x' \in X} \max_{\theta \in \Theta} \mathbb{E}_\theta [W(x, x') \mid t].$$

Далі зафіксуємо параметр θ . Метою цієї дії буде проведення базових математичних обчислень, що потім будуть використовуватися. Також це допоможе аналізу існуючих рішень. Після цих обчислень буде розглянуто, як саме враховувати вплив невідомих параметрів на розв'язок.

2.2.4 Бінарна функція витрат

Досить розповсюдженою, проте зазвичай неприродною є бінарна штрафна функція

$$W(x, x') = \mathbb{1}(x \neq x').$$

Для неперервного випадку така функція витрат не підходить, бо Баєсів ризик

$$R_\theta(q) = \sum_{t \in T} \int_{x \in X} \mathbb{1}(x \neq q(t)) \cdot p(t, x) dx = 1, \quad \forall q \in Q.$$

Будь-яка стратегія дає невірну відповідь у неперервному випадку з точки зору бінарної функції витрат, тому для неї буде розглядатися лише дискретний випадок.

Оберемо стратегію q^* , що мінімізує математичне очікування цієї функції витрат

$$\begin{aligned} q_\theta^*(t) &= \arg \min_{x'} \left\{ \sum_{x \in X} \mathbb{P}(t, x) \cdot \mathbb{1}(x \neq x') \right\} = \\ &= \arg \min_{x'} \left\{ \sum_{x \in X} \mathbb{P}(t, x) - \sum_{x \in X} \mathbb{P}(t, x) \cdot \mathbb{1}(x = x') \right\} = \\ &= \arg \min_{x'} \{1 - \mathbb{P}(t, x')\}. \end{aligned}$$

В результаті

$$q_\theta^*(t) = \arg \max_x \mathbb{P}(t, x).$$

Отже, якщо використовується бінарна функція витрат, потрібно обирати найбільш ймовірний набір параметрів. Аналітичний вираз для розрахунку f досить складний, тому доведеться скористатися чисельними методами, які не

завжди дають точну відповідь і можуть зупинитися у точці локального оптимуму, якщо такий є і не співпадає з глобальним.

2.2.5 Інтервальна функція витрат

Більш загальним варіантом бінарної штрафної функції є індикатор неналежності дійсного параметру x деякій δ -околиці обраного параметру x'

$$W(x, x') = \mathbb{1}(x \notin \delta(x')).$$

Оберемо стратегію, що мінімізує математичне очікування цієї функції витрат

$$\begin{aligned} q_{\theta}^*(t) &= \arg \min_{x'} \left\{ \int_X \mathbb{1}(x \notin \delta(x')) dF(t, x) \right\} = \\ &= \arg \min_{x'} \left\{ 1 - \int_X \mathbb{1}(x \in \delta(x')) dF(t, x) \right\} \end{aligned}$$

В результаті

$$q_{\theta}^*(t) = \arg \max_{x'} \int_{\delta(x')} dF_{\theta}(t, x) = \arg \max_{x'} \mathbb{P}(t, x \in \delta(x')).$$

Потрібно обрати таку точку, щоб ймовірність того, що параметри належать її δ -околиці, була більшою, ніж для δ -околиць інших точок при даному зображенні t .

2.2.6 Різниця моделей

Розглянемо більш природню функцію витрат — квадрат евклідової відстані між точками дійсної та обраної моделі.

Функція витрат має вигляд

$$\begin{aligned} W(x, x') &= \|G(x) - G(x')\|^2 = \sum_{v \in V} [G_v(x) - G_v(x')]^2 = \\ &= \sum_{v \in V} \sum_{i=1}^n [\lambda_i^v \cdot (x_i - x'_i)]^2 = \sum_{i=1}^n \left\{ (x_i - x'_i)^2 \cdot \sum_{v \in V} (\lambda_i^v)^2 \right\} = \\ &= \left| \beta_i^2 = \sum_{v \in V} (\lambda_i^v)^2 \right| = \sum_{i=1}^n \beta_i^2 \cdot (x_i - x'_i)^2. \end{aligned}$$

Оберемо стратегію q^* , що мінімізує математичне очікування цієї функції витрат

$$q_\theta^*(t) = \arg \min_{x'} \left\{ \int_X \sum_{i=1}^n \beta_i^2 \cdot (x'_i - x_i)^2 dF_\theta(t, x) \right\}.$$

Щоб мінімізувати неперервну функцію від параметрів x'_i , можна взяти по них похідну (ми можемо це зробити, бо x'_i не є змінною, по якій ведеться інтегрування — з точки зору інтегралу це лише константа)

$$\frac{\partial \int_X \sum_{i=1}^n \beta_i^2 \cdot (x'_i - x_i)^2 dF_\theta(t, x)}{\partial x'_i} = 2 \cdot \int_X \beta_i^2 \cdot (x'_i - x_i) dF_\theta(t, x), \quad i = 1..n$$

та прирівняти до нуля

$$\int_X (x'_i - x_i) dF_\theta(t, x) = 0, \quad i = 1..n.$$

Значення компоненти

$$x'_i = \frac{\int x_i dF_\theta(t, x)}{\int_X dF_\theta(t, x)} = \frac{\int x_i dF_\theta(t, x)}{\mathbb{P}(t)}$$

Результуюча стратегія

$$q_\theta^*(t) = \frac{\int x dF_\theta(t, x)}{\mathbb{P}(t)}.$$

У вигляді умовного математичного очікування

$$q_\theta^*(t) = \mathbb{E}_\theta[x \mid t].$$

Ділення на ймовірність появи зображення t можна сприймати як нормування при використанні чисельних методів розрахунку даного інтегралу.

2.2.7 Різниця параметрів

Розглянемо більш просту функцію витрат — квадрат евклідової норми різниці між дійсними та обраними параметрами моделі зображеного обличчя

$$W(x, x') = \|x - x'\|^2 = \sum_{i=1}^n (x_i - x'_i)^2.$$

Оберемо стратегію q^* , що мінімізує математичне очікування цієї функції витрат

$$q_\theta^*(t) = \arg \min_{x'} \left\{ \int_X \sum_{i=1}^n (x'_i - x_i)^2 dF_\theta(t, x) \right\}.$$

Маємо мінімізацію неперервної функції від параметрів x'_i , отже можемо взяти

по них похідну

$$\frac{\partial \int_X \sum_{i=1}^n (x'_i - x_i)^2 dF_\theta(t, x)}{\partial x'_i} = 2 \cdot \int_{x \in X} (x'_i - x_i) dF_\theta(t, x), \quad i = 1..n$$

та прирівняти до нуля

$$\int_X (x'_i - x_i) dF_\theta(t, x) = 0, \quad i = 1..n.$$

Значення компоненти

$$x'_i = \frac{\int_X x_i dF_\theta(t, x)}{\int_X dF_\theta(t, x)} = \frac{\int_X x_i dF_\theta(t, x)}{\mathbb{P}(t)}$$

Результуюча стратегія

$$q_\theta^*(t) = \frac{\int_X x dF_\theta(t, x)}{\mathbb{P}(t)}.$$

Оптимальна стратегія як умовне математичне очікування

$$q_\theta^*(t) = \mathbb{E}_\theta[x \mid t].$$

Отримана та ж стратегія, що мінімізує математичне очікування суми квадратів різниць координат вершин дійсної та обраної моделі обличчя. Тобто це вирішувальне правило розв'язує обидві задачі.

Ділення на ймовірність появи зображення t можна сприймати як нормування при використанні чисельних методів розрахунку даного інтегралу, як і в минулому прикладі.

2.3 Розв’язок

2.3.1 Дискретизація простору параметрів

Дискретизуємо простір параметрів X . Введемо невелике δx таке, щоб ймовірність того, що значення потрапляє в рамки гіперкубу $\left[x - \frac{\delta x}{2}; x + \frac{\delta x}{2}\right]$, приблизно дорівнювала добутку щільності розподілу в його центрі на його об’єм

$$\mathbb{P}\left(\xi \in \left[x - \frac{\delta x}{2}; x + \frac{\delta x}{2}\right]\right) \approx p(x) \cdot \delta x^n,$$

де

$$\left(x \pm \frac{\delta x}{2}\right)_i = x_i \pm \frac{\delta x}{2}, \quad i = 1..n.$$

Далі треба розбити простір на гіперкуби зі стороною δx і вважати, що коли величина потрапляє на територію певного гіперкубу, то вона дорівнює значенню в його центрі.

Відомо, що параметри x — набори коефіцієнтів нормованих головних компонент, тобто мають стандартний гаусовий розподіл. Ймовірність того, що було згенеровано саме такий набір

$$\mathbb{P}_X(x) = \delta x^n \cdot \prod_{i=1}^n \frac{\exp\left(-\frac{x_i^2}{2}\right)}{\sqrt{2 \cdot \pi}}.$$

Введемо зручне позначення для шуму в пікселі i

$$y_i = t_i - f_i(x).$$

Розглядається одне зображення t . З контексту буде зрозуміло, яке значення має

x . Ймовірність того, що обличчя на даному зображенні отримано з параметрами x

$$\mathbb{P}_Y(t | x) = \delta x^{w \cdot h} \cdot \prod_{i \in I} \frac{\exp \left\{ -\frac{y_i^2}{2 \cdot \sigma_t^2} \right\}}{\sqrt{2 \cdot \pi \cdot \sigma_t^2}}.$$

Ймовірність того, що параметри x приймають саме такі значення, якщо відомі координати опорних точок,

$$\mathbb{P}_L(x) = \delta x^{|L|} \cdot \prod_{l \in L} \frac{\exp \left\{ -\frac{\Delta_l^2(x)}{2 \cdot \sigma_L^2} \right\}}{\sqrt{2 \cdot \pi \cdot \sigma_L^2}}.$$

Сумісна ймовірність зображення t та параметрів x

$$\mathbb{P}(x, t) = \mathbb{P}_Y(t | x) \cdot \mathbb{P}_X(x) \cdot \mathbb{P}_L(x)$$

2.3.2 Інтервальна функція витрат

Розглянемо функцію витрат, за якої вірним є будь-який набір параметрів, що знаходиться в δ -околі дійсного набору x , а за всі інші сплачується штраф 1.

Для максимізації розглянемо логарифми ймовірностей, тому що це зручніше, ніж максимізація добутку кількох десятків тисяч або мільйонів значень

$$\begin{aligned} \mathbb{P}_X(x) &= \ln \delta x^n + \sum_{i=1}^n \left\{ -\frac{x_i^2}{2} - \frac{\ln 2 + \ln \pi}{2} \right\}, \\ \mathbb{P}_Y(t | x) &= \ln \delta x^{w \cdot h} + \sum_{i \in I} \left\{ -\frac{y_i^2}{2 \cdot \sigma_t^2} - \frac{\ln 2 + \ln \pi + \ln \sigma_t^2}{2} \right\}, \\ \mathbb{P}_L(x) &= \ln \delta x^{|L|} + \sum_{l \in L} \left\{ -\frac{\Delta_l^2(x)}{2 \cdot \sigma_L^2} - \frac{\ln 2 + \ln \pi + \ln \sigma_L^2}{2} \right\}. \end{aligned}$$

Приберемо константні доданки, що не впливають на результат максимізації ймо-

вірності появи даного зображення t з даними параметрами x

$$\sum_{i \in I} \left\{ -\frac{y_i^2}{2 \cdot \sigma_t^2} - \frac{\ln \sigma_t^2}{2} \right\} + \sum_{l \in L} \left\{ -\frac{\Delta_l^2(x)}{2 \cdot \sigma_L^2} - \frac{\ln \sigma_L^2}{2} \right\} - \sum_{i=1}^n \frac{x_i^2}{2} \rightarrow \max_{x \in X}.$$

Помножимо на -2 та винесемо логарифм дисперсії за знак суми

$$|I| \cdot \ln \sigma_t^2 + |L| \cdot \ln \sigma_L^2 + \sum_{i \in I} \frac{y_i^2}{\sigma_t^2} + \sum_{l \in L} \frac{\Delta_l^2(x)}{\sigma_L^2} + \sum_{i=1}^n x_i^2 \rightarrow \min_{x \in X}. \quad (2.2)$$

Замінімо дисперсію її оцінкою [20]

$$|I| \cdot \ln \frac{\sum_{i \in I} y_i^2}{|I| - 1} + |I| - 1 + |L| \cdot \ln \frac{\sum_{l \in L} \Delta_l^2(x)}{|L| - 1} + |L| - 1 + \sum_{i=1}^n x_i^2 \rightarrow \min_{x \in X}.$$

Позбавимось константних доданків

$$|I| \cdot \ln \frac{\sum_{i \in I} y_i^2}{|I| - 1} + |L| \cdot \ln \frac{\sum_{l \in L} \Delta_l^2(x)}{|L| - 1} + \sum_{i=1}^n x_i^2 \rightarrow \min_{x \in X}.$$

В загальному випадку зображення містить не тільки обличчя. Генерація усіх можливих фонів для кожної моделі є недоцільною. Потрібно отримати формулу, яка не прив'язана до кожного пікселя вхідного зображення t . Для цього перепишемо доданки як середньоквадратичні відхилення

$$|I| \cdot \ln \tilde{y}^2 + |L| \cdot \ln \left[\tilde{\Delta}^2(x) \right] + (n - 1) \cdot \tilde{x}^2 \rightarrow \min_{x \in X}.$$

Треба брати середньоквадратичне відхилення шуму по тим пікселям, де на згенерованому зображенні $f(x)$ знаходиться обличчя. За умовою накладений шум є

набором незалежних випадкових однаково розподілених величин, тому квадрат середньоквадратичного відхилення будь-якої їх сукупності буде оптимальною оцінкою їх дисперсії. Це означає, що квадрат середньоквадратичного відхилення шуму на обличчі буде тим точнішою оцінкою дисперсії шуму, чим більше місця займає обличчя на зображенні.

Продивимось на отримані формули і їх особливості.

Без додаткових припущень маємо

$$|I| \cdot \ln \sigma_t^2 + |L| \cdot \ln \sigma_L^2 + \sum_{i \in I} \frac{y_i^2}{\sigma_t^2} + \sum_{l \in L} \frac{\Delta_l^2(x)}{\sigma_L^2} + \sum_{i=1}^n x_i^2 \rightarrow \min_{x \in X}.$$

Оскільки дисперсія є константним невідомим параметром, можна спростити

$$\sum_{i \in I} \frac{y_i^2}{\sigma_t^2} + \sum_{l \in L} \frac{\Delta_l^2(x)}{\sigma_L^2} + \sum_{i=1}^n x_i^2 \rightarrow \min_{x \in X}. \quad (2.3)$$

Якщо замість сум квадратів розглядати вибіркві дисперсії, щоб позбутися невідомих дисперсій, задача приймає вигляд

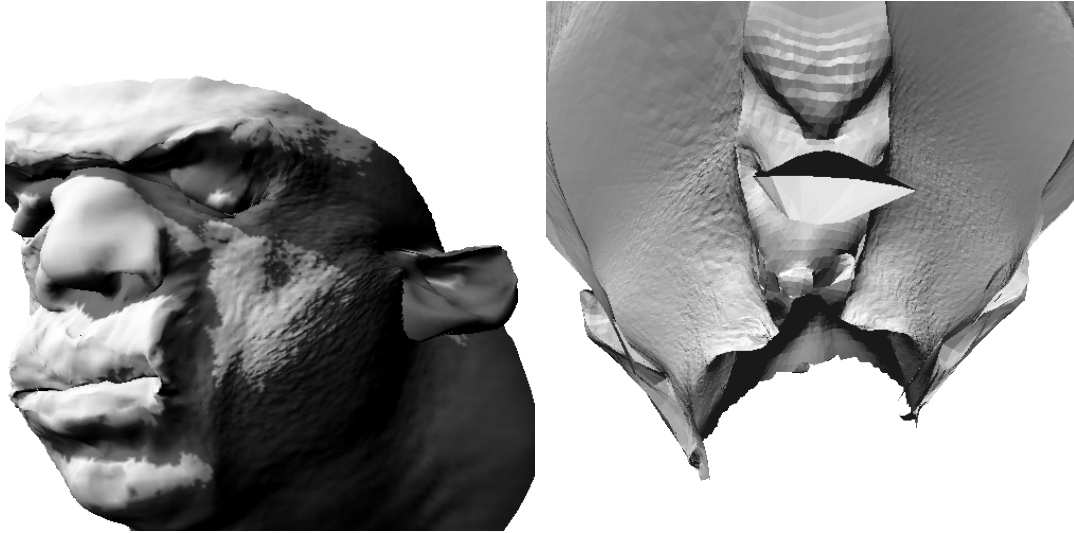
$$|I| \cdot \ln \tilde{y}^2 + |L| \cdot \ln \left[\tilde{\Delta}^2(x) \right] + (n-1) \cdot \tilde{x}^2 \rightarrow \min_{x \in X}.$$

Найпростіша цільова функція — сума квадратів відхилень між дійсним та синтезованим зображенням. Доречна у випадку, коли відомо лише те, що на зображення накладено гаусовий шум, а функція витрат бінарна і розглядається дискретний випадок

$$\sum_{i \in I} y_i^2 \rightarrow \min_{x \in X}.$$

Проте в даній задачі її застосовувати не можна. Справа в тому, що коли параметри моделі x приймають завеликі абсолютні значення, отримані зображення приймають гротескну форму (рис. 2.3а) або зовсім не схожі на обличчя люди-

ни (рис. 2.3б). Тому доданок, що містить параметри x , необхідний у цільовій функції. Суму квадратів цих параметрів також називають умовою регуляризації, яка допомагає запобігти того стану моделі, коли вона вже не описує об'єкт реального світу.



(а) Модель голови схожа на Квазімодо (б) Модель голови не схожа на голову

Рисунок 2.3 — Приклади моделей з завеликими значеннями параметрів

2.3.3 Різниця параметрів

У даному підрозділі не будемо враховувати опорні точки, щоб запобігти завеликих формул. Проте отриманий результат не втрачає гнучкості і до нього без проблем можна додати P_L .

Ймовірність того, що дане зображення t було отримано саме з параметрами x , є добутком розглянутих ймовірностей

$$\mathbb{P}(x, t) = \mathbb{P}_I(t | x) \cdot \mathbb{P}_X(x) = \delta x^{n+w \cdot h} \cdot \prod_{i \in I} \frac{\exp \left\{ -\frac{(t_i - f_i(x))^2}{2 \cdot \sigma_t^2} \right\}}{\sqrt{2 \cdot \pi \cdot \sigma_t^2}} \cdot \prod_{i=1}^n \frac{\exp \left\{ -\frac{x_i^2}{2} \right\}}{\sqrt{2 \cdot \pi}}.$$

Винесемо константи з добутків

$$\mathbb{P}(x, t) = \left(\frac{\delta x}{\sqrt{2 \cdot \pi}} \right)^{n+w \cdot h} \cdot \prod_{i \in I} \frac{\exp \left\{ -\frac{(t_i - f_i(x))^2}{2 \cdot \sigma_t^2} \right\}}{\sqrt{\sigma_t^2}} \cdot \prod_{i=1}^n \exp \left\{ -\frac{x_i^2}{2} \right\}.$$

В даному випадку не можна замінити дисперсію на її оцінку, бо для кожного x вона буде різною, через що інтеграл ймовірностей не буде дорівнювати одиниці, що порушує умову, якій повинна задовольняти ймовірнісна міра [21]. Введемо константу

$$c_t = \frac{(2 \cdot \pi)^{-\frac{n+w \cdot h}{2}} \cdot (\sigma_t^2)^{-\frac{w \cdot h}{2}}}{\mathbb{P}(t)}$$

Ймовірність певного набору параметрів x на даному зображенні t

$$\mathbb{P}(x | t) = c_t \cdot \delta x^{n+w \cdot h} \cdot \exp \left\{ -\frac{\|t - f(x)\|^2}{2 \cdot \sigma_t^2} \right\} \cdot \exp \left\{ -\frac{\|x\|^2}{2} \right\}.$$

Стратегія розпізнавання

$$q_{\theta}^*(t) = c_t \cdot \delta x^{n+w \cdot h} \cdot \sum_{x \in X} x \cdot \exp \left\{ -\frac{\|t - f(x)\|^2}{2 \cdot \sigma_t^2} \right\} \cdot \exp \left\{ -\frac{\|x\|^2}{2} \right\}.$$

Щільність розподілу набору параметрів x на даному зображенні t

$$p(x | t) = c_t \cdot \exp \left\{ -\frac{\|t - f(x)\|^2}{2 \cdot \sigma_t^2} \right\} \cdot \exp \left\{ -\frac{\|x\|^2}{2} \right\}.$$

Стратегія розпізнавання у неперервному випадку

$$q_{\theta}^*(t) = c_t \cdot \int_{x \in X} x \cdot \exp \left\{ -\frac{\|t - f(x)\|^2}{2 \cdot \sigma_t^2} \right\} \cdot \exp \left\{ -\frac{\|x\|^2}{2} \right\} dx.$$

2.3.4 Попередні результати в контексті Бассової теорії розпізнавання

Тепер очевидно, що у попередніх роботах використовувалася бінарна функція витрат. Знову розглянемо цільову функцію (1.2)

$$E(x, \theta) = \omega_c \cdot \sum_{i \in I' \subset I} \frac{y_i^2}{|I'|} + \omega_r \cdot \sum_{j=1}^n x_j^2 + \omega_l \cdot \sum_{l \in L} \frac{\Delta_l^2(x)}{|L|} \rightarrow \min_{x \in X},$$

для якої коефіцієнти ω_c , ω_r і ω_l обираються рівними 1, $2.5 \cdot 10^{-5}$ і 10 відповідно. Завдяки (2.3) стає зрозуміло, звідки беруться ваги ω . Треба лише замінити суму квадратів елементів шуму y_i добутом розміру картини $|I|$ зі зміщеною оцінкою його дисперсії на пікселях з обличчям $I' \subset I$

$$|I| \cdot \sum_{i \in I' \subset I} \frac{y_i^2}{|I'| \cdot \sigma_t^2} + |L| \cdot \sum_{l \in L} \frac{\Delta_l^2(x)}{|L| \cdot \sigma_L^2} + \sum_{i=1}^n x_i^2 \rightarrow \min_{x \in X}.$$

Оскільки $\omega_c = 1$, а ділення енергії на константу не змінює результату мінімізації, можна помножити вираз (2.3) на дисперсію шуму картини та поділити на її розмір

$$\sum_{i \in I' \subset I} \frac{y_i^2}{|I'|} + \frac{\sigma_t^2}{\sigma_L^2} \cdot \frac{|L|}{|I|} \cdot \sum_{l \in L} \frac{\Delta_l^2(x)}{|L|} + \frac{\sigma_t^2}{|I|} \cdot \sum_{i=1}^n x_i^2 \rightarrow \min_{x \in X}.$$

Маємо систему рівнянь

$$\begin{cases} \omega_r &= \frac{\sigma_t^2}{|I|}, \\ \omega_l &= \frac{\sigma_t^2}{\sigma_L^2} \cdot \frac{|L|}{|I|}, \end{cases} \implies \begin{cases} \sigma_t^2 &= \omega_r \cdot |I|, \\ \sigma_L^2 &= \frac{\omega_r}{\omega_l} \cdot |L|. \end{cases}$$

Підставимо значення $\omega_r = 2.5 \cdot 10^{-5}$ і $\omega_l = 10$

$$\begin{cases} \sigma_t^2 &= 2.5 \cdot 10^{-5} \cdot |I|, \\ \sigma_L^2 &= 2.5 \cdot 10^{-6} \cdot |L|. \end{cases}$$

Бачимо, що дисперсія шуму лінійно зростає з площею зображення. Для картинки 500×500 вона дорівнює одиниці, а для одного мегапікселя це значення вже сягає 25. Для більших зображень сама модель голови має замалу точність, тому немає сенсу розглядати поведінку дисперсії на них. Дисперсія відхилення опорних точок від дійсного положення лінійно залежить від їх кількості. Точних даних немає, проте використаний для знаходження особливих точок алгоритм зазвичай демонструється з набором із 194 опорних точок, які було запропоновано у базі PUT Face [22]. В такому випадку дисперсія дорівнює $4.85 \cdot 10^{-4}$.

2.4 Чисельні методи

2.4.1 Метод Монте-Карло

Оскільки функція f , що дає зображення t для даного набору параметрів x , є складно влаштованою, вона не представлена в явному вигляді. Інтегрування виразу, що містить цю функцію, не може бути проведено аналітично. Метод Монте-Карло — чисельний метод, що підходить до розв’язку даної задачі.

Треба підрахувати інтеграл

$$q^*(t) = c_t \cdot \int_{x \in X} x \cdot \exp \left\{ -\frac{\|t - f(x)\|^2}{2 \cdot \sigma_t^2} \right\} \cdot \exp \left\{ -\frac{\|x\|^2}{2} \right\} dx,$$

який є математичним очікуванням функції стандартного гаусового n -вимірного вектора

$$q^*(t) = c_t \cdot \mathbb{E}_\xi \left[\xi \cdot \exp \left\{ -\frac{\|t - f(\xi)\|^2}{2 \cdot \sigma_t^2} \right\} \right], \quad \xi_i \sim \mathcal{N}(0, 1).$$

Внесемо все в експоненту для полегшення машинних обчислень

$$q^*(t) = \mathbb{E}_\xi \left[\frac{\xi}{|\xi|} \cdot \exp \left\{ \ln |\xi| - \frac{\|t - f(\xi)\|^2}{2 \cdot \sigma_t^2} + \ln c_t \right\} \right].$$

Для отримання оцінки інтегралу треба згенерувати N векторів з n незалежних випадкових величин, що мають стандартний нормальний розподіл

$$\xi^{(i)} = \langle \xi_1^{(i)}, \xi_2^{(i)}, \dots, \xi_n^{(i)} \rangle, \quad \xi_j^{(i)} \sim \mathcal{N}(0, 1).$$

Для кожного вектору розрахувати значення підінтегрального виразу

$$Q^{(i)} = \frac{\xi^{(i)}}{|\xi^{(i)}|} \cdot \exp \left\{ \ln |\xi^{(i)}| - \frac{\|t - f(\xi^{(i)})\|^2}{2 \cdot \sigma_t^2} + \ln c_t \right\}.$$

Значення константи c_t визначимо з умови нормування

$$c_t \cdot \sum_{i=1}^N \exp \left\{ \frac{\|t - f(\xi^{(i)})\|^2}{2 \cdot \sigma_t^2} + \frac{\|\xi^{(i)}\|^2}{2} \right\} = 1.$$

Явний вигляд

$$c_t = \frac{1}{\sum_{i=1}^N \exp \left\{ \frac{\|t - f(\xi^{(i)})\|^2}{2 \cdot \sigma_t^2} + \frac{\|\xi^{(i)}\|^2}{2} \right\}}.$$

Оцінкою інтегралу буде середнє значення

$$\overline{Q}_N = \frac{\sum_{i=1}^N \hat{Q}^{(i)}}{N}.$$

Визначимо оптимальне значення N . Згадаємо центральну граничну теорему

$$\sqrt{N} \cdot \frac{\overline{Q}_N - M(Q)}{\sqrt{D(Q)}} \Rightarrow \mathcal{N}(0, 1).$$

Для певного $z_\gamma > 0$

$$\mathbb{P} \left(\sqrt{N} \cdot \frac{\overline{Q}_N - M(Q)}{\sqrt{D(Q)}} \leq z_\gamma \right) = \Phi(z_\gamma),$$

де Φ — інтегральна функція стандартного гаусового розподілу. Нас цікавить ймовірність

$$\mathbb{P} \left(-z_\gamma \leq \sqrt{N} \cdot \frac{\overline{Q}_N - M(Q)}{\sqrt{D(Q)}} \leq z_\gamma \right) = \Phi(z_\gamma) - \Phi(-z_\gamma).$$

Використаємо модуль для перевірки входження виразу в довірчий інтервал та скористаємося симетричністю гаусового розподілу

$$\mathbb{P} \left(\left| \sqrt{N} \cdot \frac{\overline{Q}_N - M(Q)}{\sqrt{D(Q)}} \right| \leq z_\gamma \right) = 2 \cdot \Phi(z_\gamma) - 1.$$

Поділимо обидві частини нерівності на модуль середнього значення виборки

$$\mathbb{P} \left(\left| \sqrt{N} \cdot \frac{\overline{Q}_N - M(Q)}{\overline{Q}_N \cdot \sqrt{D(Q)}} \right| \leq \frac{z_\gamma}{|\overline{Q}_N|} \right) = 2 \cdot \Phi(z_\gamma) - 1.$$

Позначимо відносну похибку

$$\varepsilon = \left| \frac{\overline{Q}_N - M(Q)}{\overline{Q}_N} \right|$$

та перепишемо вираз

$$\mathbb{P} \left(\left| \sqrt{N} \cdot \frac{\varepsilon}{\sqrt{D(Q)}} \right| \leq \frac{z_\gamma}{|\overline{Q}_N|} \right) = 2 \cdot \Phi(z_\gamma) - 1.$$

Перенесемо все окрім кореню від кількості спроб в праву частину нерівності

$$\mathbb{P} \left(\left| \sqrt{N} \right| \leq \left| \frac{z_\gamma \cdot \sqrt{D(Q)}}{\varepsilon \cdot \overline{Q}_N} \right| \right) = 2 \cdot \Phi(z_\gamma) - 1$$

та піднесемо обидві частини нерівності до квадрату

$$\mathbb{P} \left(N \leq \frac{z_\gamma^2 \cdot D(Q)}{\varepsilon^2 \cdot \overline{Q}_N^2} \right) = 2 \cdot \Phi(z_\gamma) - 1.$$

Перепозначимо отриману ймовірність

$$\mathbb{P} \left(N \leq \frac{z_\gamma^2 \cdot D(Q)}{\varepsilon^2 \cdot \overline{Q}_N^2} \right) = \gamma.$$

Щоб виразити z_γ , потрібно скористатися квантилем нормального розподілу

$$\Phi^{-1} \left(\frac{\gamma + 1}{2} \right) = z_\gamma.$$

Щоб відносна похибка обчислення математичного очікування методом Монте-

Карло не перевищувала ε з ймовірністю γ і вище, необхідна та достатня кількість ітерацій рахується за формулою

$$N = \left\lceil \frac{z_\gamma^2 \cdot D(Q)}{\varepsilon^2 \cdot \overline{Q}_N^2} \right\rceil, \quad z_\gamma = \Phi^{-1} \left(\frac{\gamma + 1}{2} \right). \quad (2.4)$$

Коли дисперсія отриманої випадкової величини Q невідома, використовується її незміщена оцінка

$$\tilde{Q}_N = \frac{\sum_{i=1}^N \left(\hat{Q}^{(i)}(x) \right)^2 - \overline{Q}_N^2}{N - 1}.$$

Формула (2.4) використовується лише для перевірки того, чи дає наявна вибірка з N елементів відповідь з вказаною похибкою. Спочатку необхідно дати алгоритму “розігрітись” і не порівнювати значення N з необхідною кількістю кроків. Також немає необхідності у перевірці відповідності значення N на кожній ітерації, особливо якщо це займає багато часу — достатньо робити перерахунок кожні 100 або 1000 кроки в залежності від інтегралу.

2.4.2 Методи мінімізації цільової функції

При мінімізації цільової функції (2.2) виникає кілька складностей:

- 1) наявність локальних мінімумів;
- 2) складний аналітичний вираз функції візуалізації;
- 3) недиференційованість.

Останній пункт може викликати сумніви, тому переглянемо причини його справедливості. Множина пікселів дискретна за умовою, а опорні точки прив’язані до неї. Множина можливих кольорів дискретна за технічними особливостями комп’ютерної техніки, і будь-яка зміна кольору відбувається дискретно. Алгоритми екранного згладжування межі моделі зроблено для того, щоб межа

полігонів була трохи розмитою. При недостатньому згладжуванні цю межу явно видно (рис. 2.4), що дає різкий стрибок цільової функції, коли модель починає (перестає) обіймати нові (старі) пікселі. При збільшенні розмірів візуалізованого зображення ефект не зникне, тому що це специфіка відображення діагональних ліній на множині пікселів.



Рисунок 2.4 — Візуалізація обличчя зі збільшеними діагональними лініями

Функція візуалізації є складною, бо

- 1) треба визначити невидимі вершини;
- 2) проводиться фільтрація, згладжування та інші процедури покращення зображення;

- 3) розрахунок самозатінення містить функцію візуалізації, проте без тіней та текстури.

Для запобігання локальних мінімумів використовують координати знайдених опорних точок обличчя.

Диференціювати цільову функцію можна чисельними методами з таким dx , при якому стрибки не будуть суттєвими.

Щоб диференціювати цільову функцію аналітично, автори попередніх робіт один раз на певну кількість ітерацій (наприклад, 1000) шукають видимі вершини/трикутники та розраховують самозатінення на них. Далі з'являється можливість розрахувати похідні аналітично і використовувати у методах, що потребують Якобіан (градієнтний спуск, метод Ньютона-Гаусса тощо).

2.5 Висновки

Проведено аналіз задачі реконструкції та пред'явлено дві основні постановки, розв'язки яких суттєво відрізняються. У випадку бінарної функції витрат на дискретизованій множині параметрів моделі достатньо мінімізувати апостеріорну ймовірність шуканих величин. Саме цей підхід і використовується у всіх відомих на даний момент методах. Більш доречно для даного випадку функція витрат, що дорівнює сумі квадратів відхилень отриманого рішення від реального, потребує пошуку умовного математичного очікування шуканих величин за відомим фото. Якщо вважати, що глобальний мінімум цільової функції достатньо глибокий і математичне очікування буде приблизно дорівнювати екстремуму, два дані підходи можна вважати близькими.

Той факт, що для коректного розв'язку поставленої задачі треба шукати математичне очікування, також вказує на те, що знайдене рішення можна покращити за допомогою параметрів, що знаходяться близько до розв'язку.

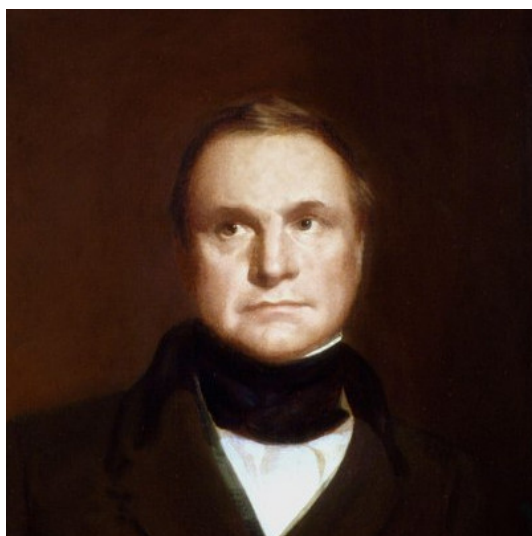
Одним з важливіших результатів є виявлення теоретичної некоректності підходу, який полягає в тому, що рішення шукається для найбільш ймовірних невідомих параметрів (таких як освітлення, поворот тощо). Це означає, що можна знайти стратегію, яка при будь-яких значеннях невідомих параметрів буде давати менший ризик, тобто буде менше помилятися.

3 ПРАКТИЧНІ РЕЗУЛЬТАТИ

У даному розділі наведено практичні результати просторової реконструкції із зазначенням використаних методів, постановки задачі, вхідних даних та візуалізації результуючої моделі.

3.1 Портрет Чарльза Беббіджа

Є портрет відомого англійського математика Чарльза Беббіджа (рис. 3.1а). За допомогою бінарної функції витрат та дискретизації множини параметрів було отримано навпрочуд реалістичну модель (рис. 3.1б).



(а) Фрагмент портрету

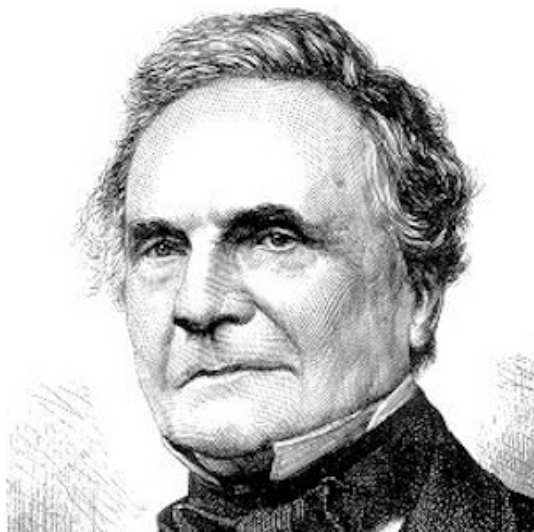


(б) Реконструйована модель

Рисунок 3.1 — Чарльз Беббідж молодий

Проте реалістична текстура та тіні — це не привід вважати, що алгоритм відпрацював добре. Тому було взято інший портрет вченого, де він старше та дивиться в іншу сторону (рис. 3.2а). До наявної моделі застосовано лише інший поворот та освітлення (рис. 3.2б), ніякої додаткової інформації з портрету система не дістає. Тепер можна оцінити якість реконструкції на око, бо дійсної

тривимірної моделі голови Беббіджа немає.



(а) Портрет



(б) Модель з поворотом

Рисунок 3.2 — Чарльз Беббідж

З рисунку видно, що деякі ділянки реконструйовані некоректно. Проте взагалом результат задовільний. До отриманої моделі можна застосувати будь-який поворот (рис. 3.3а) та тінь (рис. 3.3б).



(а) Вигляд збоку



(б) Інша тінь

Рисунок 3.3 — Реконструйована модель Чарльза Беббіджа

3.2 Висновки

Наведені результати демонструють базові можливості просторової реконструкції людського обличчя за портретом. Використані методи працюють та дають реалістичні результати. Проте необхідні подальші дослідження, щоб виявити слабкі та сильні сторони різних підходів.

ВИСНОВКИ

В результаті виконання роботи вдалося закласти теоретичне підґрунтя для подальшої роботи над задачею створення оптимізованого методу реконструкції просторової конфігурації людського обличчя в системах біометричної ідентифікації.

Попередні роботи приголомшують своїми результатами. Їх аналіз допоміг задати правильні питання та шукати на них відповідь. Також вони містять хитрощі, до яких треба вдаватися, щоб обходити певні складні ситуації.

За допомогою статистичної теорії розпізнавання образів було зроблену коректну постановку та теоретичний розв'язок поставленої задачі. Перевірені чисельні методи вказали, як обходити обчислювальні складнощі, що виникають при практичному вирішенні задачі.

Було реалізовано демонстративне програмне забезпечення, що дає гарну реконструкцію просторової конфігурації людського обличчя за певних умов.

ПЕРЕЛІК ПОСИЛАНЬ

1. Parke, Frederick I. Computer Generated Animation of Faces / Frederick I. Parke // *Proceedings of the ACM Annual Conference - Volume 1.* — 1972. — Pp. 451–457.
2. Parke, F.I. A Parametric Model for Human Faces / F.I. Parke. — Univerisity Microfilms, 1974.
3. A 3D Face Model for Pose and Illumination Invariant Face Recognition / P. Paysan, R. Knothe, B. Amberg et al. // *Proceedings of the 6th IEEE Internati-onal Conference on Advanced Video and Signal based Surveillance (AVSS) for Security, Safety and Monitoring in Smart Environments.* — 2009.
4. Blanz, V. A morphable model for the synthesis of 3D faces / V. Blanz, T. Vetter // *Proceedings of the 26th annual conference on Computer graphics and interactive techniques.* — 1999. — Pp. 187–194.
5. Jones, M. J. Multidimensional morphable models / M. J. Jones, T. Poggio. — 1998. — Jan. — Pp. 683–688.
6. Blanz, V. Face Identification across Different Poses and Illuminations with a 3D Morphable Model / V. Blanz, S. Romdhani, T. Vetter // *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on.* — 2002. — Pp. 192–197.
7. Blanz, V. Face recognition based on fitting a 3D morphable model / V. Blanz, T. Vetter // *IEEE Transactions on Pattern Analysis and Machine Intelligence.* — 2003. — Sept. — Vol. 25, no. 9. — Pp. 1063–1074.
8. Tyshchenko, M. 3D Reconstruction of Human Face Based on Single or Several Images / M. Tyshchenko // *Управляющие системы и машины.* — 2011. — № 2. — С. 79–85.

9. Rossi, Francesca. Handbook of Constraint Programming (Foundations of Artificial Intelligence) / Francesca Rossi, Peter van Beek, Toby Walsh. — New York, NY, USA: Elsevier Science Inc., 2006. — P. 978.
10. Saito, Shunsuke. Real-Time Facial Segmentation and Performance Capture from RGB Input / Shunsuke Saito, Tianye Li, Hao Li // *Computer Vision – ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VIII*. — 2016. — Pp. 244–261.
11. Face2Face: Real-time Face Capture and Reenactment of RGB Videos / J. Thies, M. Zollhöfer, M. Stamminger et al. // *Proc. Computer Vision and Pattern Recognition (CVPR), IEEE*. — 2016.
12. Kazemi, Vahid. One Millisecond Face Alignment with an Ensemble of Regression Trees / Vahid Kazemi, Josephine Sullivan // *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*. — 2014. — Pp. 1867–1874.
13. Schlesinger, M. I. Nearly Optimal Statistical Recognition and Learning / M. I. Schlesinger, E. V. Volodazkiy // *Proceedings of 4th International conference on Inductive Modelling*. — 2013.
14. Amberg, Brian. Optimal Step Nonrigid ICP Algorithms for Surface Registration. / Brian Amberg, Sami Romdhani, Thomas Vetter // *2007 IEEE Conference on Computer Vision and Pattern Recognition*. — 2007.
15. Айвазян, С.А. Прикладная статистика: Классификация и снижение размерности : Справочное издание / С.А. Айвазян. — М.: Финансы и статистика, 1989. — 606 с.
16. Rabin, Michael O. Probabilistic automata / Michael O. Rabin // *Information and Control*. — 1963. — Vol. 6, no. 3. — Pp. 230–245.

17. Berger, J.O. Statistical Decision Theory: Foundations, Concepts, and Methods / J.O. Berger // Springer Series in Statistics. — Springer New York, 1980. — P. 428.
18. Wald, A. Selected Papers in Statistics and Probability / A. Wald, I.M. Statistics // No. 10. — McGraw-Hill, 1955. — P. 702.
19. Schlesinger, M.I. Ten Lectures on Statistical and Structural Pattern Recognition / M.I. Schlesinger, V. Hlavác // Computational Imaging and Vision. — Springer Netherlands, 2002. — P. 522.
20. Боровков, А. А. Математическая статистика / А. А. Боровков // Учебники для ВУЗов. Специальная литература. — 4 изд. — Санкт-Петербург: Лань, 2010. — С. 705.
21. Дороговцев, А. Я. Элементы общей теории меры и интеграла / А. Я. Дороговцев. — К.: Выща школа, 1989. — С. 152.
22. Kasiński, A. The PUT Face Database / A. Kasiński, A. Florek, A. Schmidt // *Image Processing & Communications*. — 2008. — Vol. 13, no. 3–4. — Pp. 59–64.