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 ________ *(підпис)*
Керівник к.т.н., Барановський Олексій Миколайович ________ *(підпис)*

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

Київ — 2017 року

РЕФЕРАТ

Диссертация содержит 72 страницы, 18 иллюстраций и список использованной литературы из 26 наименований.

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 — разработка метода реконструкции пространственной конфигурации человеческого лица в системах биометрической идентификации. Это даст возможность использовать корректно распознанные модели в системах биометрической идентификации.

Для достижения цели были использованы

- Базельская порождающая модель поверхности лица;
- байесова теория распознавания образов для правильной постановки и решения задачи и анализа существующих решений;
- численные методы для расчёта решения.

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

ABSTRACT

The thesis contains 72 pages, 18 figures and 26 references.

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 design an optimized method of human face inverse rendering in biometric identification systems. This will allow to use correctly reconstructed models in biometric identification systems.

To perform the study

- Basel morphable face model was used;
- bayesian patterns recognition theory was used to formulate the problem correctly and analyse existent solutions;
- numerical methods were used to compute the solution.

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

РЕФЕРАТ

Дисертація містить 72 сторінки, 18 ілюстрацій і бібліографію з 26 найменувань.

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

Метою роботи є розробка оптимізованого методу реконструкції просторової конфігурації людського обличчя в системах біометричної ідентифікації. Це допоможе використовувати коректно розпізнані моделі в системах біометричної ідентифікації.

Для досягнення мети було використано

- Базельську породжувальну модель поверхні обличчя;
- баєсову теорію розпізнавання образів для коректної постановки задачі і аналізу існуючих рішень;
- чисельні методи для розрахунку рішення.

БАЄСОВА СТРАТЕГІЯ, РОЗПІЗНАВАННЯ ОБРАЗІВ, ПРОСТОРОВА РЕКОНСТРУКЦІЯ, ПОРОДЖУВАЛЬНА МОДЕЛЬ, БІОМЕТРИЧНА ІДЕНТИФІКАЦІЯ

ЗМІСТ

ВСТУП

Актуальність роботи. Примітивні системи біометричної ідентифікації 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, розпізнавання опорних точок обличчя та просторова реконструкція обличчя. Методи для реалізації останнього підходу наразі розвиваються, різні дослі-джувачі пропонують свої алгоритми, які швидше або точніше, ніж у інших.

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. Реконструкцію тривимірної моделі поверхні обличчя можна використовувати в системах біометричної ідентифікації, для реконструкції облич історичних постатей, у відеоіграх і так далі. Неточності, що були знайдені в існуючих алгоритмах, допоможуть побудувати програмне забезпечення, що робить реконстукцію обличчя точніше за аналоги.

Публікації. XV Всеукраїнська науково-практична конференція студентів, аспірантів та молодих вчених «Теоретичні і прикладні проблеми фізики, математики та інформатики».

1 ПОПЕРЕДНІ РОБОТИ ПРИСВЯЧЕНІ РЕКОНСТРУКЦІЇ МОДЕЛІ ОБЛИЧЧЯ

В першому розділі розглянуто коротку історію досліджень, що пов'язані з реконструкцією просторової конфігурації людського обличчя за одним зображенням. Проводиться аналіз та порівняння існуючих методів реконструкції. Розбір попередніх робіт дає змогу чітко поставити задачу розпізнавання, що розв'язується в другому розділі дисертації.

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 заздалегідь було нанесено необхідні полігони (рис. 1.1).

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).

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.

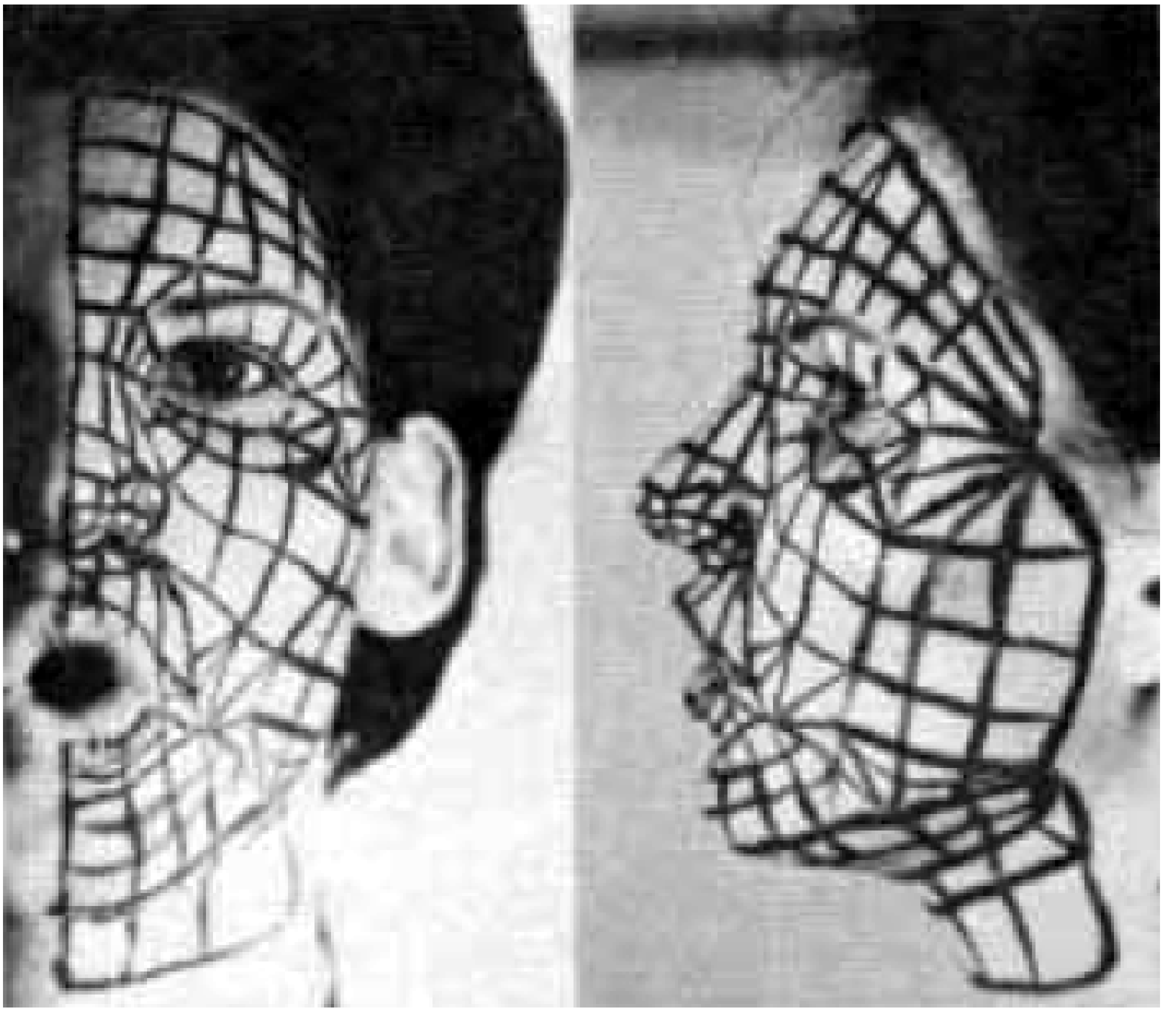


Рисунок 1.1 — Пара ортогональних фотографій асистента з нанесеними полігонами

1.1.2 Модель Базельського університету

У даній роботі використовується породжувальна модель обличчя 3D Basel Face Model (BFM) розроблена командою Базельського університету [3]. Перший підхід, яких 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 обличчя Тома Хенкса (рис. 1.3б) за допомогою одного кадру з кінофільму “Форрест Гамп”

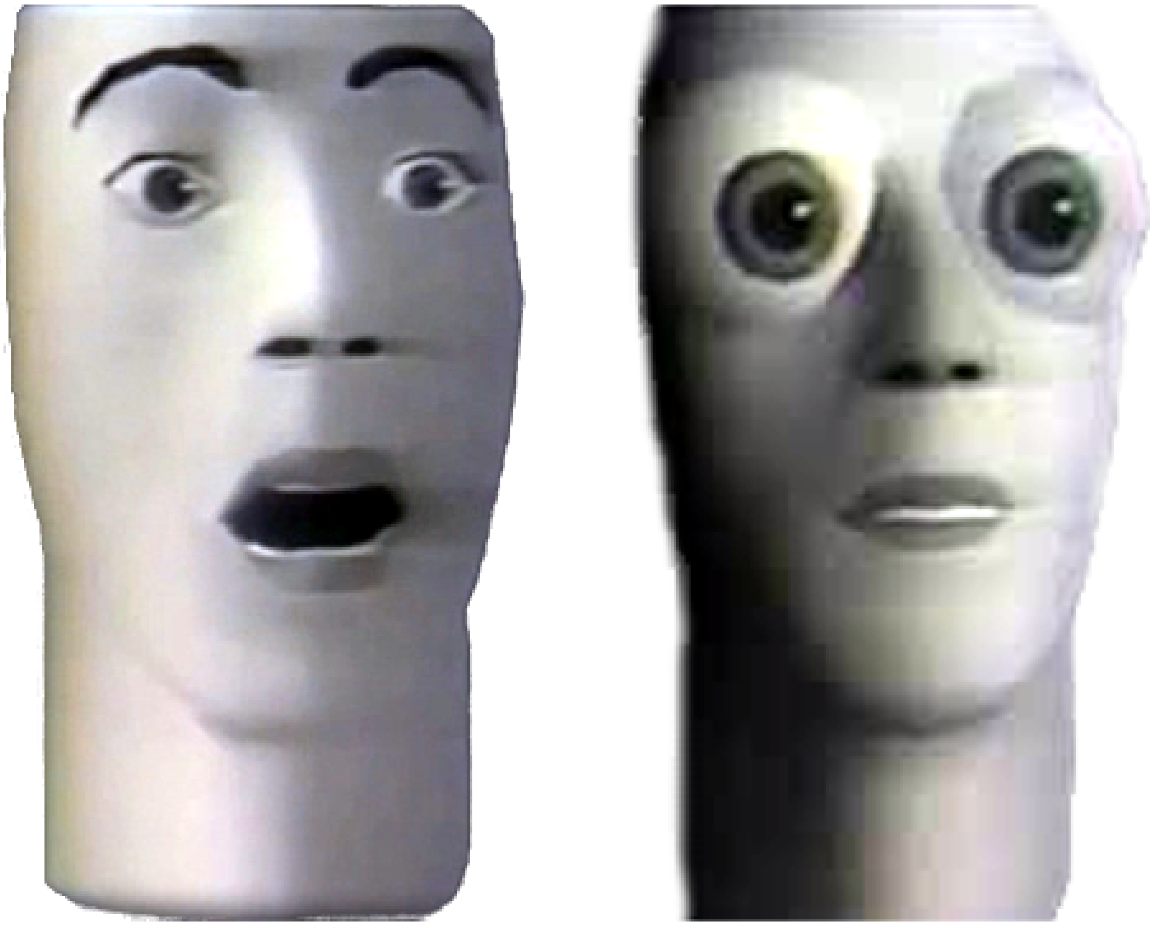
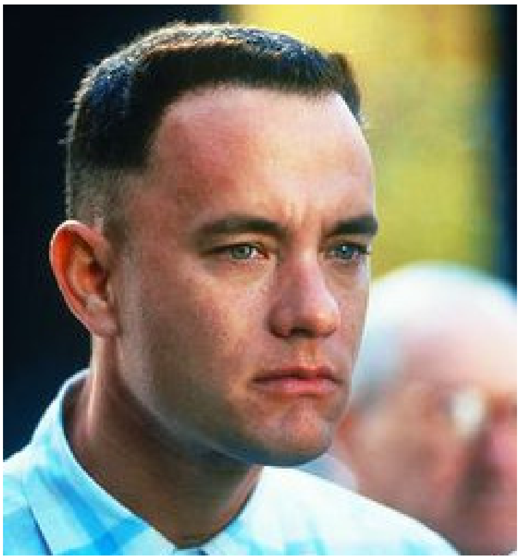


Рисунок 1.2 — Моделі з різною формою очей та виразами облич

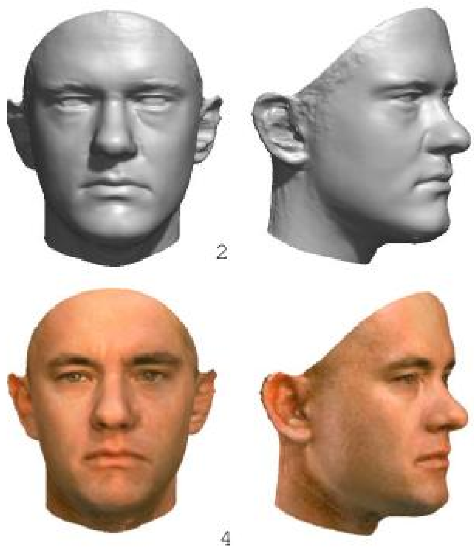
(рис. 1.3а).

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 породжувальної 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.



(а) Кадр з кінофільму “Форрест Гамп”



(б) Реконструкція голови Тома Хенкса

Рисунок 1.3 — Демонстрація роботи алгоритму дослідників з Базельського університету

1.1.3 Вітчизняна робота

У 2011 році Максим Тищенко, випускник Національного технічного університету України “Київський політехнічний інститут”, працюючи у Міжнародному науково-навчальному центрі інформаційних технологій та систем над кандидатською дисертацією, 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 облич генеруються як зважене середнє. Проте було наявне спрощення: самозатінення не бралося до уваги, що дозволяло дуже швидко знаходити оптимальне освітлення за пласкою моделлю затінення методом найменших квадратів.

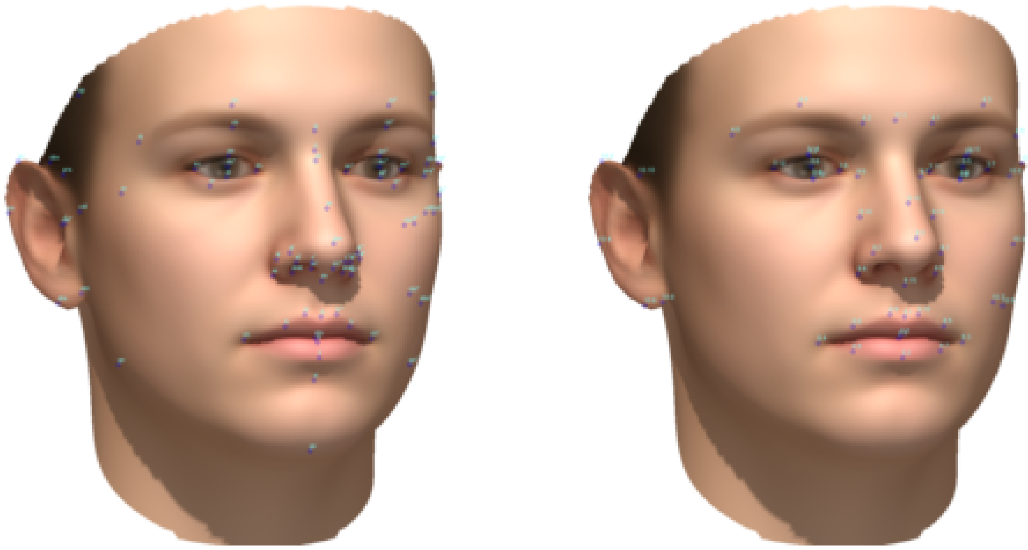


Рисунок 1.4 — Усереднені моделі облич з відміченими опорними точками

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. Останньою відомою роботою є Large Scale 3D Morphable Models [12] — це генеративна модель обличчя, яка отримана з тривимірних знімків 10 тисяч людей різної статі, віку та раси, що робить її дуже різноманітною та корисною (рис. 1.5).

1.2 Постановка задачі

Спільною рисою методів, що дають добрі результати, є використання породжувальної моделі обличчя для виконання аналізу через синтез. 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 форми

Рисунок 1.5 — Приклади облич, що побудовані за допомогою LSFM

моделі обличчя, що мають стандартний нормальний розподіл, та суми квадратів відхилень параметрів камери від початкових значень, що визначалися людиною-оператором:

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

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 чином 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), потім поступово додаються нові компоненти;
3) спочатку дисперсія σ_N^2 обирається достатньо великою, аби більше звертати увагу на грубіші відхилення, а з кожною ітерацією її зменшують, щоб отримувати все більш схожу модель.

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 [13]. Результати, що дає алгоритм, можна побачити на рис. 1.6.

Також популярним є евристичний підхід, який полягає у використанні зміщеної оцінки дисперсії замість суми квадратів відхилень. Це дозволяє брати різну кількість точок на кожній ітерації. Нова цільова функція має вигляд

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

де Δ_l^2 — функція, яка перетворює параметри моделі на квадрат відхилення її опорних точок від реальних, а коефіцієнти ω_c, ω_r і ω_l обираються рівними 1, $2.5 \cdot 10^{-5}$ і 10 відповідно.

Рисунок 1.6 — Приклади знаходження опорних точок обличчя

1.3 Питання щодо постановки задачі

Перше фундаментальне питання полягає в тому, навіщо потрібно мінімізувати саме ті функції, що було вказано вище. В роботі [6] вибір функції витрат (1.1) мотивується оцінкою апостеріорного максимуму. Проте функція (1.2) не має підґрунтя окрім того, що її мінімізація дає блискучі результати, проте не ідеальні (рис. 1.7).

У формулі (1.2) є вагові коефіцієнти ω_c, ω_r і ω_l, походження яких не пояснюється; невідомо, чому значення мають бути сами такими, і чи вони

(а) Фрагмент оригінального фото Дональда Трампа

(б) Реконструкція поверхні обличчя Дональда Трампа

Рисунок 1.7 — Некоректно реконструйований ніс

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 дасть змогу відрізнити локальний мінімум від глобального.

1.4 Проблема у розв'язку задачі

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

$$(k^*, \theta^*) = \underset{k \in K, \theta \in \Theta}{\arg\min} E(k, \theta), \tag{1.3}$$

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 (освітлення, матриця повороту, перспективи, переносу тощо). Шукаються накращі 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 ділянки зображення, на яких обличчя немає. Проте звідки відомо, що обличчя насправді позиціоновано саме так, як вказав алгоритм оцінки положення обличчя? Також немає гарантії, що тривимірні моделі були розмічені абсолютно точно. Додаткову складність додає питання щодо кількості і характеру опорних точок, що використовуються для отримання приблизного положення, які дадуть максимально точний результат. Виникають такі ситуації, коли одній конфігурації опорніх точок відповідає різне положення обличчя. Найпростіший 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. 1.8а зображено модель, що була згенерована за допомогою BFM. На рис. 1.8б зображено ту ж саму модель, проте до неї було застосовано матрицю перспективи. Оскільки матриця перспективи — аффінне 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 1.8б — 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 худої людини з високим лобом.

(а) Візуалізована модель людського обличчя

(б) Візуалізація з іншою матрицею перспективи

Рисунок 1.8 — Приклад візуалізації однієї моделі з різними параметрами камери

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 такі природні фактори, як спотворення зображення різними об’єктивами фотокамер (зокрема перспектива). Тому такий підхід теж є хибним.

У статті [14] 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.

Висновки до розділу 1

Проведено огляд існуючих алгоритмів реконструкції з теоретичної точки зору, більш детальне дослідження яких наведено у наступних розділах дисертації. Запропоновані методи досить складно порівняти з практичної точки зору: для цього треба реалізувати їх на одній мові програмування та перевірити на одній тестовій вибірці на однакових обчислювальних машинах. Проте було вказано математичні неточності, що були знайдені при вивченні даних методів. В результаті аналізу стало зрозуміло, що на даний момент для просторової реконструкції людського обличчя за одним зображенням доцільно використовувати генеративну модель обличчя. Цей факт використано в наступному розділі для постановки задачі розпізнавання.

2 ПОСТАНОВКА ЗАДАЧІ РЕКОНСТРУКЦІЇ З ВИКОРИСТАННЯМ ГЕНЕРАТИВНОЇ МОДЕЛІ

Другий розділ присвячено постановці та розв'язку задачі з теоретичної точки зору в термінах баєсової теорії розпізнавання образів. Досліджуються теоретичні неточності існуючих підходів. Пропонується альтернативна постановка проблеми просторової реконструкції обличчя за однією фотографією. Наводиться метод біометричної ідентифікації за реконструйованими моделями.

2.1 Початкові умови

2.1.1 Породжувальна модель обличчя

2.1.1.1 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 ICP [15]. Після цього достатньо відмітити опорні точки обличчя лише на одній моделі, щоб вони були відомі на всіх інших, у тому числі на похідних обличчях.

Середня модель обличчя — середнє арифметичне всіх наявних моделей

$$\overline{f}_v = \frac{\sum\limits_{f \in F} f_v}{|F|}, \qquad v \in V,$$

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

Останній етап, який нас цікавить, це обробка даних за допомогою методу головних компонент [16]. 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) пом'якшені обмеження на параметри моделі в порівнянні зі зваженим середнім, 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.

2.1.1.2 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

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

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

$$G : X \to M.$$

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

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

2.1.2 Зображення

Позначимо множину зображень T і множину кольорів C. Введемо множину пікселів зображення I. Зображення $t \in T$ є відображення з множини пікселів у множину їх значень

$$t : I \to C.$$

Колір пікселя i в зображенні t позначимо як t_i.

Найчастіше 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 робити 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

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

Функцією, що перетворює набір параметрів на зображення, є відображення

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

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.

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

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

Введемо короткий запис для інверсії маски

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

2.1.3 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,

$$P^2 : \Theta^M \times \mathbb{R}^3 \to I.$$

Проекцію p точки v моделі m на площину запишемо

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

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

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

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 та i' знаходяться одна від одної на відстані μ, запишемо як

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

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

2.1.4 Фон

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

2.1.5 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. Колір пікселя i

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

2.1.6 Оцінка опорних точок

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

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

Позначимо відхилення проекції опорних точок моделі від їх оцінки

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

2.1.7 Оцінка сегменту обличчя

Введемо ймовірність того, що даний піксель i належить обличчю. За відомих x і θ^M ці ймовірності визначені однозначно і приймають значення 0 або 1

$$\mathbb{P}_{\theta^M}\left(i \in I' \mid x\right) = \overline{h}_i^{\theta^M}(x).$$

Дана ймовірність заздалегідь невідома, бо невідомі x і θ, тому єдина можливість отримати її оцінку — проаналізувати зображення. Зображення може містити декілька облич, тому перевірка належності одного пікселя i множині пікселів обличчя I' не має сенсу. Треба перевіряти усю сукупність пікселів I' на те, чи знаходиться там обличчя

$$\mathbb{P}\left(h_i^{\theta^M} = 0, \forall i \in I'\right) = \mathbb{P}\left(I' \mid t\right).$$

2.1.8 Результуюча ймовірність

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

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

З визначення умовної ймовірності випливає

$$p_\theta = \mathbb{P}_\theta\left\{f_{\theta^M}(\xi) \cdot \overline{h}(x) + \theta^B \cdot h(x) + \eta = t, \zeta = \Delta(\xi) \,\middle|\, \xi = x\right\} \cdot \mathbb{P}\left\{\xi = x\right\}$$

Позбавимося умовної ймовірності, замінивши ξ на x і розіб'ємо першу ймовірність на добуток двох ймовірностей, вважаючи, що похибка оцінки опорних точок не залежить від шуму на зображенні

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

Подібним чином в обчислення можна включити сегментацію замість опорних точок

$$p_\theta = \mathbb{P}_{\theta^M}\left\{\eta = t - f_{\theta^M}(x) \cdot \overline{h}(x) - \theta^B \cdot h(x)\right\} \cdot \mathbb{P}_\theta\left\{I'\right\} \cdot \mathbb{P}\left\{\xi = x\right\}.$$

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

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

2.1.9 Узагальнення на випадок кількох зображень

Пред'явлено кілька зображень

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

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 $x = const \in X$. Цього важко досягти, якщо обрати фотографії зняті з різних камер в різні пори року, проте ця вимога має місце, коли на вході 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 [17]

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

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

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

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

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

2.2 Задача

2.2.1 Баєсова задача розпізнавання

Поставимо баєсову задачу розпізнавання [18]. Для цього потрібно визначитися з функцією витрат [19]

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

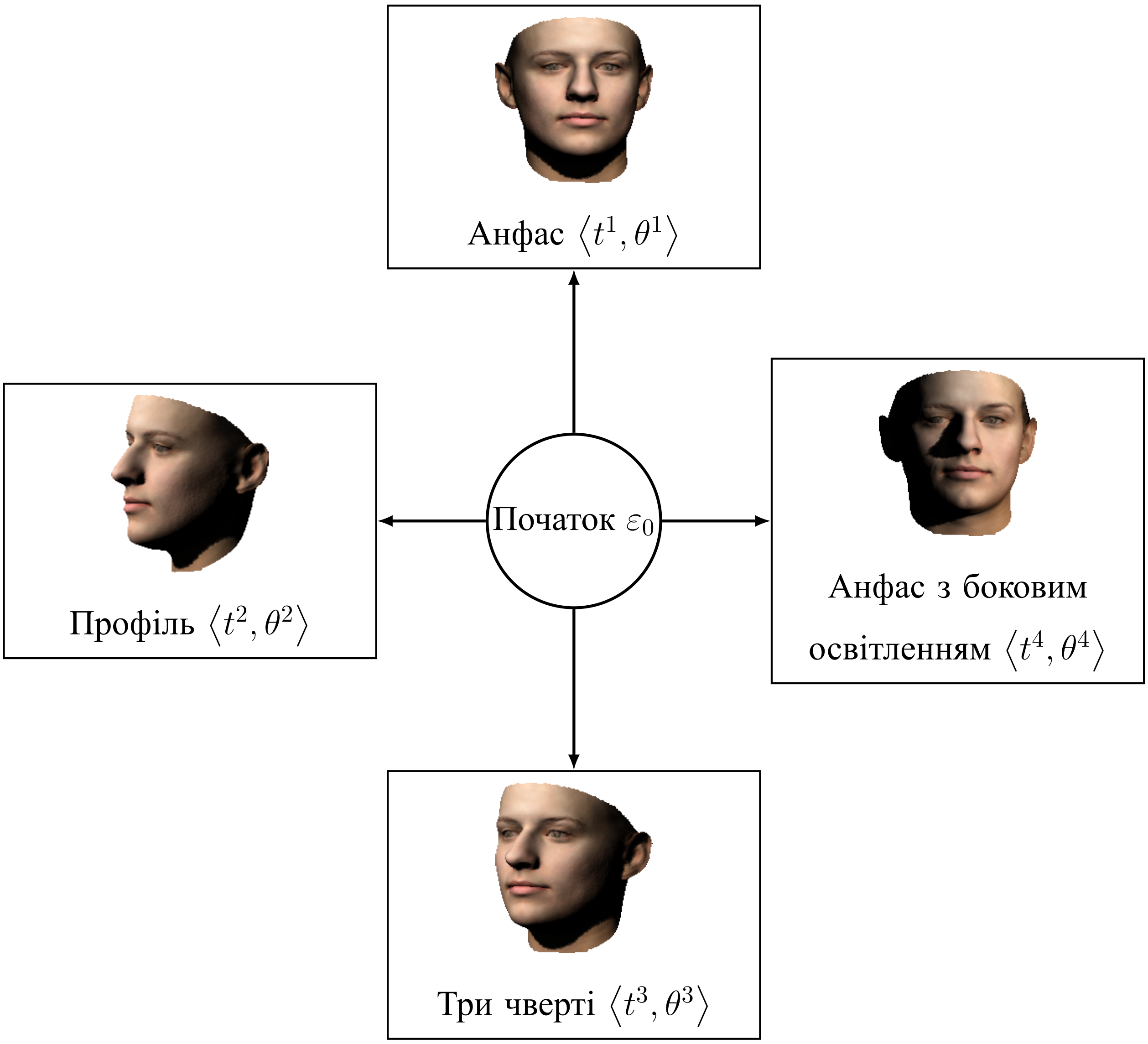


Рисунок 2.1 — Приклад роботи автомата з незалежними станами

Стратегію

$$q : T \to X,$$

яка для зображення t дає результат x, позначимо

$$q\left(t\right) = x.$$

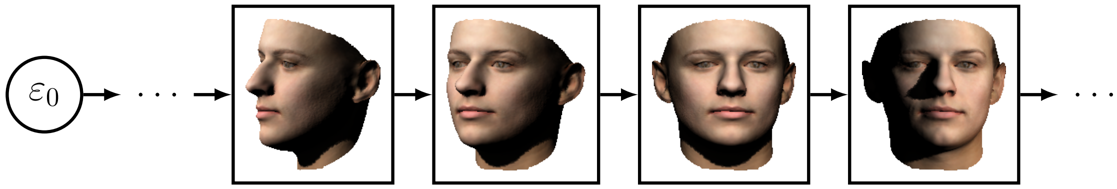


Рисунок 2.2 — Приклад автомата, що генерує відео

Математичне сподівання функції витрат W як функції випадкової пари $\langle t, x \rangle$ для даного вирішального правила q називається баєсовим ризиком [20]

$$R_\theta(q) = \int\limits_{T,X} W(x, q(t))\, dF_\theta(t, x) = \mathbb{E}_\theta \left[W(x, q(t)) \right],$$

де F_θ — функція сумісного розподілу зображення та параметрів зображеної моделі при відомих θ.

2.2.2 Баєсова стратегія

Стратегія q^* називається баєсовою, якщо може бути представлена у вигляді [14]

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

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 при будь-якому значенні параметрів θ небаєсова стратегія буде розпізнавати гірше за певну баєсову

$$R_\theta(q_0) > R_\theta(q), \qquad \forall \theta \in \Theta.$$

Оскільки зовнішній вигляд найкращих стратегій декларовано, для кожної задачі треба шукати лише коефіцієнти $\tau(\theta)$. Прикладом доцільної задачі у даному випадку може бути пошук такої стратегії, що помиляється якнайменше при кожному θ. За визначенням мінімальні витрати гарантує та стратегія, якій заздалегідь відомо параметр θ, з яким вона працюватиме. Така стратегія називається оптимальною

$$q^{\theta}(t) = \arg\min_{q \in Q} R_{\theta}(q). \tag{2.2}$$

Можемо сформулювати оптимізаційну задачу з обмеженнями

$$\begin{cases} c \to \min, \\ R_{\theta}(q) - R_{\theta}\left(q^{\theta}\right) < c, \qquad \forall \theta \in \Theta. \end{cases} \tag{2.3}$$

В результаті розв'язку отримаємо ваги $\tau(\theta)$ і стратегію q. Доведемо від протилежного, що знайдена стратегія не буде негодящою: якщо знайдеться інша стратегія q', яка краща за дану q, то знайдеться таке невід'ємне δ, що

$$R_{\theta}(q') - R_{\theta}\left(q^{\theta}\right) < c - \delta, \qquad \forall \theta \in \Theta.$$

Якщо таке $\delta > 0$ існує, то c можна зменшити на δ і оптимізаційну задачу (2.3) не було розв'язано до кінця. Проте δ може бути наперед визначеною точністю обчислень, якою можна поступитися на користь швидкості розв'язку. Стратегія, що є розв'язком (2.3), називається субоптимальною.

Далі зафіксуємо параметр θ, ніби він відомий, і розглянемо побудову оптимальних стратегій для різних штрафних функцій.

2.2.3 Бінарна функція витрат

Досить розповсюдженою, проте зазвичай неприродною є бінарна штрафна функція

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

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

$$R_\theta\left(q\right) = \sum_{t \in T} \int\limits_{x \in X} \mathbb{1}\left(x \neq q\left(t\right)\right) \cdot p\left(t, x\right) d\, x = 1, \qquad \forall q \in Q.$$

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 сподівання цієї функції витрат

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

В результаті маємо максимізацію апостеріорної ймовірності, адже t дано і фіксовано, його ймовірність константа, і на неї можна поділити

$$q^*_\theta\left(t\right) = \arg\max_{x} \mathbb{P}\left(t, x\right) = \arg\max_{x} \mathbb{P}\left(x \mid t\right).$$

Отже, якщо використовується бінарна функція витрат, потрібно обирати найбільш ймовірний набір параметрів. Аналітичний вираз для розрахування f досить складний, тому доведеться скористатися чисельними методами, які не завжди дають точну відповідь і можуть зупинитися у точці локального оптимуму, який часто не співпадає з глобальним у даній задачі.

2.2.4 Інтервальна функція витрат

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

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

Оберемо стратегію, що мінімізує математичне сподівання цієї функції витрат

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

В результаті

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

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

2.2.5 Різниця моделей

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

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

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

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

Щоб мінімізувати неперервну функцію від параметрів x_i', можна взяти по них похідну, бо x_i' не є змінною, по якій ведеться інтегрування — з точки зору інтегралу це лише константа

$$
\frac{\partial \int\limits_X \sum\limits_{i=1}^{n}\beta_i^2 \cdot \left(x_i' - x_i\right)^2 dF_\theta\left(t, x\right)}{\partial x_i'} = 2 \cdot \int\limits_X \beta_i^2 \cdot \left(x_i' - x_i\right) dF_\theta\left(t, x\right), \qquad i = 1..n
$$

та прирівняти до нуля

$$
\int\limits_X \left(x_i' - x_i\right) dF_\theta\left(t, x\right) = 0, \qquad i = 1..n.
$$

Значення компоненти

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

Результуюча стратегія

$$q^*_\theta\left(t\right) = \frac{\int\limits_X x dF_\theta\left(t, x\right)}{\mathbb{P}\left(t\right)}.$$

У вигляді умовного математичного сподівання

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

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

2.2.6 Різниця параметрів

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

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

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

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

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

$$\frac{\partial \int\limits_X \sum\limits_{i=1}^{n} \left(x_i' - x_i\right)^2 dF_\theta\left(t, x\right)}{\partial x_i'} = 2 \cdot \int\limits_{x \in X} \left(x_i' - x_i\right) dF_\theta\left(t, x\right), \qquad i = 1..n$$

та прирівняти до нуля

$$\int\limits_X \left(x_i' - x_i\right) dF_\theta\left(t, x\right) = 0, \qquad i = 1..n.$$

Значення компоненти

$$x_i' = \frac{\int\limits_X x_i dF_\theta\left(t, x\right)}{\int\limits_X dF_\theta\left(t, x\right)} = \frac{\int\limits_X x_i dF_\theta\left(t, x\right)}{\mathbb{P}\left(t\right)}$$

Результуюча стратегія

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

Оптимальна стратегія як умовне математичне сподівання

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

Отримана та ж стратегія, що мінімізує математичне сподівання 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

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

де

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

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 ґаусовий розподіл. Ймовірність того, що було згенеровано саме такий набір

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

Введемо зручне позначення для шуму в пікселі i

$$y_i = t_i - f_i\left(x\right) \cdot \overline{h}_i\left(x\right) - \theta_i^B \cdot h_i\left(x\right), \qquad i \in I.$$

Розглядається одне зображення t. З контексту буде зрозуміло, яке значення має x. Ймовірність того, що обличчя на даному зображенні отримано з параметрами x

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

Ймовірність того, що параметри x приймають саме такі значення, якщо відома оцінка координат опорних точок,

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

Сумісна ймовірність зображення t та параметрів x

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

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

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

Приберемо константні доданки, що не впливають на результат максимізації ймовірності появи даного зображення t з даними параметрами x

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

Помножимо на -2 та винесемо логарифм дисперсії за знак суми

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

Оскільки дисперсія є константним невідомим параметром, можна спростити

$$\sum_{i\in I}\frac{y_i^2}{\sigma_t^2}+\sum_{l\in L}\frac{\Delta_l^2\left(x\right)}{\sigma_L^2}+\sum_{i=1}^{n}x_i^2\to\min_{x\in X}. \tag{2.4}$$

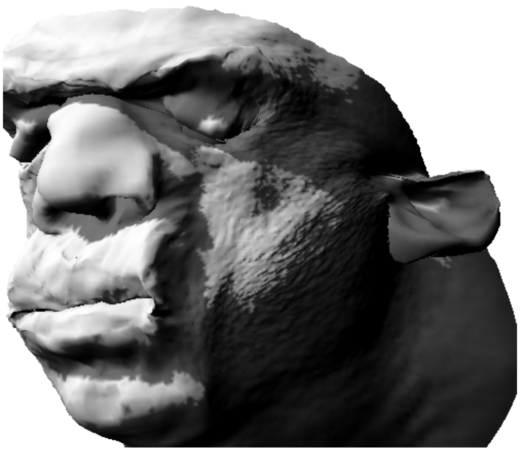
В загальному випадку зображення містить не тільки обличчя. Генерація усіх можливих фонів для кожної моделі є недоцільною. Потрібно отримати формулу, яка не прив'язана до кожного пікселя вхідного зображення t. Для цього треба брати середньоквадратичне відхилення шуму по тим пікселям, де на згенерованому зображенні $f\left(x\right)$ 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 [21]. Це означає, що квадрат середньоквадратичного відхилення шуму на обличчі буде тим точнішою оцінкою дисперсії шуму, чим більше місця займає обличчя на зображенні.

$$\sum_{i\in I}\frac{y_i^2}{\sigma_t^2}+\sum_{l\in L}\frac{\Delta_l^2\left(x\right)}{\sigma_L^2}+\sum_{i=1}^{n}x_i^2\to\min_{x\in X}. \tag{2.5}$$

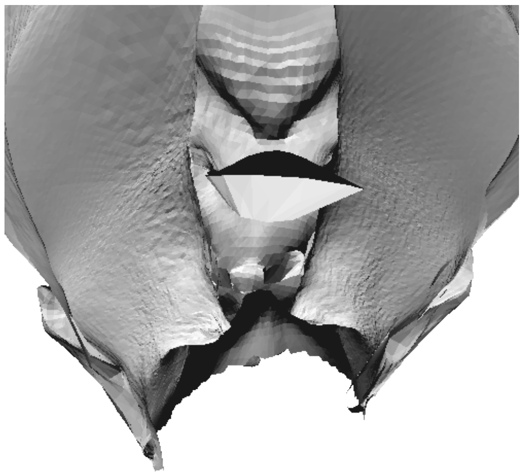
Найпростіша цільова функція — сума квадратів відхилень між дійсним та синтезованим зображенням. Доречна у випадку, коли відомо лише те, що на зображення накладено ґаусовий шум, а функція витрат бінарна і розглядається дискретний випадок

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

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 людини (рис. 2.3б). Тому доданок, що містить параметри x, необхідний у цільовій функції. Суму квадратів цих параметрів також називають умовою регуляризації, яка допомогає 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 ймовірність P_L.

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

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

Винесемо константи з добутків

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

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 [22]. Введемо константу

$$c_t = \frac{(2 \cdot \pi)^{-\frac{n+|I|}{2}} \cdot \left(\sigma_t^2\right)^{-\frac{|I|}{2}}}{\mathbb{P}(t)}$$

Ймовірність певного набору параметрів x на даному зображенні t

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

Стратегія розпізнавання

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

Щільність розподілу набору параметрів x на даному зображенні t

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

Стратегія розпізнавання у неперервному випадку

$$q_\theta^*\left(t\right) = c_t \cdot \int\limits_{x \in X} x \cdot \exp\left\{-\frac{y_i^2}{2 \cdot \sigma_t^2}\right\} \cdot \exp\left\{-\frac{\|x\|^2}{2}\right\} d\, x.$$

2.3.4 Попередні результати в контексті баєсової теорії розпізнавання

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

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

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

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

Оскільки $\omega_c = 1$, а ділення енергії на константу не змінює результату мінімізації, можна помножити вираз (2.5) на дисперсію шуму картинки та поділити на її розмір

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

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 [23]. В такому випадку дисперсія дорівнює $4.85\cdot 10^{-4}$, тобто вважається, що оцінка положення точок ідеальна.

2.3.5 Врахування фону

Методи оптимізації використовують порівняння значень цільової функції в різних точках для визначення кращого набору параметрів. Оскільки в досліджуваній задачі йдеться мова саме про мінімізацію певної функції E, значення на наступному кроці $j+1$ повинно бути меншим за значення на поточному кроці j

$$E\left(x^j\right) > E\left(x^{j+1}\right).$$

Введемо позначення для тієї множини пікселів, що на даній ітерації вважається множиною пікселів обличчя

$$I^j = \left\{i \mid i \in I, h_i\left(x^j\right) = 0\right\}.$$

Відому частину енергії, що не містить фону, позначимо

$$E'\left(x^j\right) = \sum_{i \in I_j} \frac{\left\|f\left(x^j\right) - t_i\right\|^2}{\sigma_t^2} + \sum_{l \in L} \frac{\Delta_l^2\left(x\right)}{\sigma_L^2} + \sum_{i=1}^{n} x_i^2.$$

Частину, що містить фон, позначимо

$$E^B\left(x^j\right) = \sum_{i \in I \setminus I_j} \frac{\left\|\theta_i^B - t_i\right\|^2}{\sigma_t^2}.$$

Цільова функція

$$E\left(x^j\right) = E'\left(x^j\right) + E^B\left(x^j\right).$$

Порівняння цільової функції на сусідніх кроках оптимізації приймає вигляд

$$E'\left(x^j\right) + E^B\left(x^j\right) > E'\left(x^{j+1}\right) + E^B\left(x^{j+1}\right).$$

Перенесемо функції, що містять фон, в праву частину нерівності, а відомі в ліву

$$E'\left(x^j\right) - E'\left(x^{j+1}\right) > E^B\left(x^{j+1}\right) - E^B\left(x^j\right). \tag{2.6}$$

Суми з правої сторони нерівності містять невідомі величини, проте за умовою фон на одному зображенні не змінюється, тому θ_i^B має одне й те саме значення на будь-якій ітерації. Оскільки

$$(I \setminus I_{j+1}) \cap (I \setminus I_j) = \overline{I_{j+1}} \cap \overline{I_j} = \overline{I_{j+1} \cup I_j}$$

і

$$\begin{cases} \overline{I_{j+1}} \setminus \overline{I_{j+1} \cup I_j} = \overline{I_{j+1}} \cap (I_{j+1} \cup I_j) = I_j \setminus I_{j+1}, \\ \overline{I_j} \setminus \overline{I_{j+1} \cup I_j} = \overline{I_j} \cap (I_{j+1} \cup I_j) = I_{j+1} \setminus I_j, \end{cases}$$

маємо

$$E^B\left(x^{j+1}\right) - E^B\left(x^j\right) = \sum_{i \in I_j \setminus I_{j+1}} \frac{\left\|\theta_i^B - t_i\right\|^2}{\sigma_t^2} - \sum_{i \in I_{j+1} \setminus I_j} \frac{\left\|\theta_i^B - t_i\right\|^2}{\sigma_t^2}.$$

Нерівність (2.6) приймає вигляд

$$E'\left(x^j\right) - E'\left(x^{j+1}\right) > \sum_{i \in I_j \setminus I_{j+1}} \frac{\left\|\theta_i^B - t_i\right\|^2}{\sigma_t^2} - \sum_{i \in I_{j+1} \setminus I_j} \frac{\left\|\theta_i^B - t_i\right\|^2}{\sigma_t^2}.$$

Згадаємо, що за умовою на зображення накладено центрований адитивний ґаусовий шум з дисперсією σ_t^2

$$E'\left(x^j\right) - E'\left(x^{j+1}\right) > \sum_{i \in I_j \setminus I_{j+1}} \frac{\eta_i^2}{\sigma_t^2} - \sum_{i \in I_{j+1} \setminus I_j} \frac{\eta_i^2}{\sigma_t^2}.$$

Суму квадратів незалежних випадкових величин, що мають стандартний нормальний розподіл, розподілено за законом χ^2

$$\begin{cases} E'\left(x^j\right) - E'\left(x^{j+1}\right) > \chi \sim \zeta_1 - \zeta_2, \\ \zeta_1 \sim \chi^2_{|I_j \setminus I_{j+1}|}, \\ \zeta_2 \sim \chi^2_{|I_{j+1} \setminus I_j|}. \end{cases} \tag{2.7}$$

Для того, щоб нерівність (2.7) виконувалася з ймовірністю 1, коли $E'\left(x^j\right) > E'\left(x^{j+1}\right)$, необхідно, щоб кожне наступне обличчя було не менше поточного, доки не буде заповнено все зображення. Альтернативний варіант — обмежити коливання силуету обличчя, щоб різниця його площі на різних ітераціях була якнайменшою, завдяки чому ймовірність збереження нерівності буде якнайбільшою.

Оскільки не було декларовано, що зображення — прямокутна область, вона може бути довільної форми. Дане зауваження дозволяє комбінувати обидва міркування щодо коректної стратегії пошуку вірної моделі. Якщо заздалегідь за допомогою деякого алгоритму отримати оцінку силуету обличчя $\hat{I}'$, його можна використовувати в якості множини пікселів I. Отримали задачу

$$\begin{cases} E\left(x\right) \to \min\limits_{x}, \\ \left|I' \Delta \hat{I}'\right| < c, \qquad c > 0, \end{cases}$$

де c — наперед визначена точність оцінки силуету обличчя.

2.3.6 Біометрична ідентифікація

У статті дослідників, що працювали над створенням базельскої моделі обличчя, вказано метод ідентифікації людей з використанням реконструкцій просторових конфігурацій їх облич, який дає добрі експериментальні результати [6].

Вводиться скалярний добуток параметрів двох моделей обличчя

$$(x, x') = \sum_{i=1}^{n} x_i \cdot x'_i.$$

Оскільки головні компоненти нормовані, вони мають одиничну дисперсію, ділення на яку не впливає на результат. Такий скалярний добуток породжує відстань Махаланобіса [24], що в даному випадку еквівалентна звичайній евклідовій нормі

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

Було доведено [25], що для ідентифікації краще використовувати аналог косинуса кута між двома обличчями в якості міри несхожості

$$d\,(x, x') = \frac{(x, x')}{\|x\| \cdot \|x'\|}.$$

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

$$q^*(t) = c_t \cdot \int\limits_{x \in X} x \cdot \exp\left\{-\frac{\|t - f(x)\|^2}{2 \cdot \sigma_t^2}\right\} \cdot \exp\left\{-\frac{\|x\|^2}{2}\right\} d\, x, \tag{2.8}$$

який є математичним сподіванням функції стандартного ґаусового n-вимірного вектора

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

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

$$\xi^{(i)} = \left\langle \xi_1^{(i)}, \xi_2^{(i)}, \ldots, \xi_n^{(i)} \right\rangle, \qquad \xi_j^{(i)} \sim \mathcal{N}(0, 1).$$

Для кожного вектору розрахувати значення підінтегрального виразу

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

Значення константи c_t визначимо з умови нормування

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

Явний вигляд

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

Оцінкою інтегралу буде середнє значення

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

Визначимо оптимальне значення N. Згадаємо центральну граничну теорему

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

Для певного $z_\gamma > 0$

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

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

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

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

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

Поділимо обидві частини нерівності на модуль середнього значення виборки

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

Позначимо відносну похибку

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

та перепишемо вираз

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

Перенесемо все окрім кореню від кількості спроб в праву частину нерівності

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

та піднесемо обидві частини нерівності до квадрату

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

Перепозначимо отриману ймовірність

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

Щоб виразити z_γ, потрібно скористатися квантилем нормального розподілу

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

Щоб відносна похибка обчислення математичного сподівання методом Монте-Карло не перевищувала ε з ймовірністю γ і вище, необхідна та достатня кількість ітерацій рахується за формулою

$$N = \left\lceil \frac{z_\gamma^2 \cdot D\left(Q\right)}{\varepsilon^2 \cdot \overline{Q}_N^2} \right\rceil, \qquad z_\gamma = \Phi^{-1}\left(\frac{\gamma + 1}{2}\right). \tag{2.9}$$

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

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

Формула (2.9) 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 Мінімізація цільової функції

При мінімізації цільової функції (2.4) виникає кілька складностей:

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

Останній пункт може викликати сумніви, тому переглянемо причини його справедливості. Множина пікселів дискретна за умовою, а опорні точки прив’зані до неї. Множина можливих кольорів дискретна за технічними особливостями комп’ютерної техніки, тому 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, через специфіку відображення діагональних ліній.

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

Рисунок 2.4 — Візуалізація обличчя зі збільшеними діагональними лініями

можливість розрахувати похідні аналітично і використовувати у методах, що потребують якобіан (градієнтний спуск, метод Ньютона-Ґауса тощо).

Висновки до розділу 2

Пред'явлено дві основні постановки задачі реконструкції просторової конфігурації людського обличчя, розв'язки яких суттєво відрізняються. У випадку бінарної функції витрат на дискретизованій множині параметрів моделі треба максимізувати апостеріорну ймовірність шуканих величин. Саме цей підхід і використовується у всіх відомих на даний момент методах. Більш доречна для даного випадку функція витрат, що дорівнює сумі квадратів відхилень отриманого рішення від реального, потребує пошуку умовного математичного сподівання шуканих величин за відомим фото. Якщо глобальний мінімум цільової функції настільки глибокий, що математичне сподівання приблизно дорівнює екстремуму, два дані підходи повинні дати однакові результати. Той факт, що для коректного розв'язку поставленої задачі треба шукати математичне сподівання, також вказує на те, що знайдене рішення можна покращити за допомогою параметрів, що знаходяться близько до розв'язку.

Одним з найважливіших 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.

Врахування фону призвело до альтернативної постановки задачі, яка не була розглянута іншими дослідниками.

З минулих робіт виявлено, як використовувати реконструкцію у системах біометричної ідентифікації.

3 ПРАКТИЧНІ РЕЗУЛЬТАТИ РЕКОНСТРУКЦІЇ МОДЕЛІ ОБЛИЧЧЯ ЗА ОДНИМ ФОТО

У третьому розділі наведено проблеми, які виникають при реалізації теоретично сильних методів, їх детальний аналіз та засоби обійти їх, якщо це можливо. Показано практичні результати просторової реконструкції із зазначенням використаних методів, постановки задачі, вхідних даних та візуалізації результуючої моделі.

3.1 Обчислювальна складність

3.1.1 Повний перебір

Для якісної реконструкції потрібно мати хоча б 20 параметрів. Якщо множина значень кожного параметру складається з двох елементів $\{-1, +1\}$, потрібно перебрати $2^{20} \approx 10^6$ різних наборів. Якщо генерація одного обличчя потребує одну соту долю секунди, перебрати 10^6 значень вдасться за три години. При використанні трьох значень $\{-1, 0, +1\}$ отримаємо рік роботи на $3^{20} \approx 3 \cdot 10^9$ векторів x.

Зважаючи на те, що насправді потрібно знати значення параметрів хоча б до цілої частини, а їх значення знаходяться в проміжку $[-3; +3]$ за правилом $3 \cdot \sigma$, стає зрозуміло, що на практиці цей метод використовувати не вийде, бо 20 параметрів, кожен з яких приймає 7 різних значень, потрібно підбирати десять мільярдів років ($7^{20} \approx 8 \cdot 10^{16}$ різних моделей).

3.1.2 Метод Монте-Карло

Розрахуємо відношення ймовірностей (2.8) для облич, що відрізняються в одному пікселі на 1 (мінімальна різниця кольорів)

$$\frac{\mathbb{P}_{+1}}{\mathbb{P}_0} = \frac{\exp\left\{-\Sigma - \frac{(t_i - f_i \pm 1)^2}{2\cdot\sigma_t^2}\right\}}{\exp\left\{-\Sigma - \frac{(t_i \mp f_i)^2}{2\cdot\sigma_t^2}\right\}} = \exp\left\{\frac{2\cdot(t_i \mp f_i) + 1}{2\cdot\sigma_t^2}\right\}.$$

Якщо дисперсія дорівнює 1, маємо

$$\frac{\mathbb{P}_{+1}}{\mathbb{P}_0} = \exp\left\{t_i \mp f_i + \frac{1}{2}\right\}.$$

Оскільки в даному випадку розглядаються цілочисельні зміни кольорів, модуль показника не може бути меншим за 0.5. Мінімальне відношення між ймовірностями двох різних облич становить $\sqrt{e} \approx 1.7$. Зазвичай різниця між двома моделями помітна на око, бо немає сенсу обирати таку точність, яку людина не буде помічати. З кожним навмання обраним набором параметрів різниця квадратів змінюється на досить помітне число. Якщо 10 пікселів змінять інтенсивність на 2 одиниці, ймовірність може не змінитися, якщо кольори будуть знаходитися за одиницю до та після реального кольору. Несиметрична зміна кольору на 2 одиниці в 10 пікселях призведе до відношення ймовірностей $e^{-10} \approx 2 \cdot 10^{-9}$. Навіть такі непомітні зміни призведуть до того, що одна ймовірність буде більшою за іншу в два мільярди разів, а їх середнє буде повністю складатися з більшого доданку.

На практиці виявилося, що інтегрування методом Монте-Карло для розв'язку задачі з квадратичною функцією витрат — задача неможлива. Різниця ймовірностей між навмання обраними моделями обличчя неймовірно велика, тому метод перетворювався на пошук глобального максимуму функції шляхом генерації випадкових параметрів. Такий підхід може бути доцільним для генерації початкових умов алгоритму локальної оптимізації, проте без цієї модифікації випадкова генерація моделей не має ніяких переваг над іншими методами.

3.1.3 Баєсова стратегія

Субоптимальна баєсова стратегія, що є розв'язком системи (2.3)

$$\begin{cases} c \to \min, \\ R_\theta\left(q\right) - R_\theta\left(q^\theta\right) < c, \qquad \forall \theta \in \Theta, \end{cases}$$

де q за теоремою про дихотомію має вигляд [14]

$$\begin{cases} \sum\limits_{\theta \in \Theta} \tau\left(\theta\right) = 1, \\ \tau\left(\theta\right) \geq 0, \qquad \forall \theta \in \Theta, \\ q = \underset{q \in Q}{\arg\min} \sum\limits_{\theta \in \Theta} \tau\left(\theta\right) \cdot R_\theta\left(q\right), \end{cases}$$

потребує розрахунку ризиків оптимальних стратегій (2.2)

$$q^\theta = \underset{q \in Q}{\arg\min}\, R_\theta\left(q\right)$$

для кожного θ.

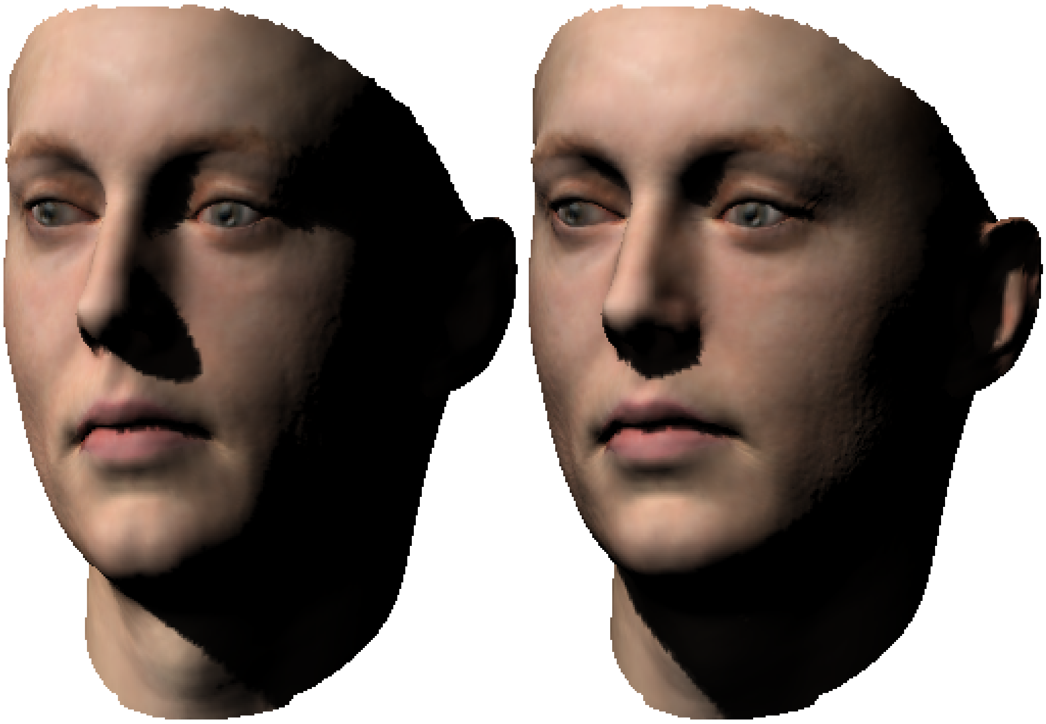
Дискретизуємо простір параметрів моделі так, щоб кожен параметр приймав лише по 7 цілих значень з проміжку $[-3; +3]$, візьмемо бінарну функцію витрат та введемо припущення, що локальний екстремум співпадає з глобальним. Розрахунок ризику — це інтегрування по множині можливих параметрів моделі. У попередньому підпункті було вказано, що навіть візуалізувати необхідну кількість різних облич неможливо, але дане завдання ускладнюється необхідністю багатократного запуску алгоритму локальної оптимізації. Припустимо, що було взято дуже мало параметрів – наприклад, два. Отримаємо лише 49 різних облич. При фіксованому фоні, рівні шуму, освітленні, повороті та розмірі обличчя можна легко підібрати обидва параметри повним перебором.

Наступною проблемою є множина можливих параметрів. Візьмемо кути нахилу від $-\frac{\pi}{2}$ до $+\frac{\pi}{2}$. Якщо використовувати досить помітну зміну кута $\frac{\pi}{6}$ (рис. 3.1), отримаємо 6 різних позицій для кожного кута Ейлера. Оскільки їх 3, доведеться вести розрахунки для 216 різних поворотів.

(а) Початкове положення (б) Поворіт вбік (в) Поворіт вбік и вгору

Рисунок 3.1 — Повороти обличчя на 30° по різним осям

Освітлення є одним з важливіших параметрів, бо саме воно формує зовнішній вигляд обличчя, і може як допомогти, так і ввести в оману. Положення джерела світла описується лише двома кутами, проте діапазон можливих значень повинен бути ширше за можливі значення повороту голови, бо джерело може світити ззаду, через що модель буде майже повністю темна. З кроком $\frac{\pi}{6}$ (рис. 3.2 та рис. 3.3) та значеннями від $-\pi$ до $+\pi$ маємо 12 значень на кожну компоненту та 144 загалом. Також треба мати хоча б 5 різних значень інтенсивності світла.



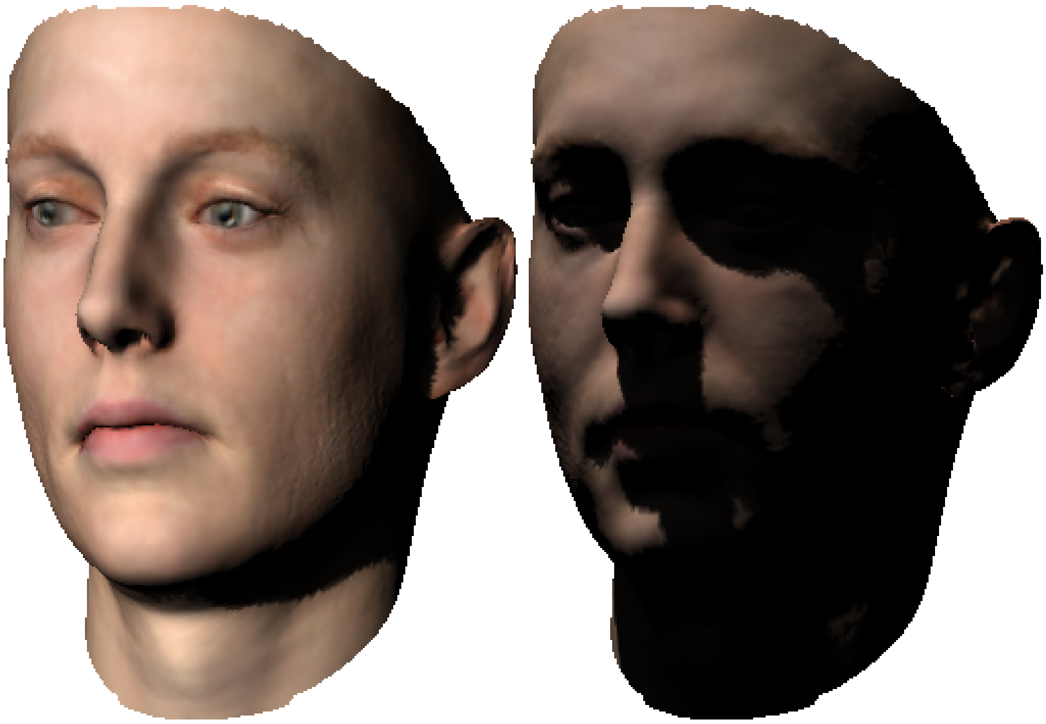
(а) Початкове положення (б) Зміна азимутального кута на 30°

Рисунок 3.2 — Переміщення джерела світла по азимуту

Бачимо, що навіть без врахування фону, переміщення і шуму потрібно перебрати $144 \cdot 5 \cdot 216 = 155520 \approx 1.5 \cdot 10^5$ різних параметрів для кожного з 49 облич. Їх генерація забере 21 годину. Для кожного згенерованого обличчя потрібно запустити процедуру оптимізації, щоб визначити точність роботи алгоритму за даних параметрів. Якщо використати одну ітерацію покомпонентного спуску, знадобиться перебрати лише по 14 облич для кожного зображення, що потребує 12 діб роботи на одному комп'ютері, що візуалізує одну модель за одну соту долю секунди.

На рисунку 3.1 видно, що обрана зміна кута повороту дуже помітна, проте з рисунку 3.3 помітно, що для зенітного кута освітлення обраний крок дискретизації є досіть значним, а чим зеніт ближчий до прямого кута, тим значніший азімутний поворіт. Обрана роздільна сітка не підходить і потрібно ще зменшити крок зміни положення джерела освітлення.

Якщо грамотно поставити експеримент та паралельно обчислювати значення ризиків, то для моделей з дуже малою кількістю параметрів за реальний



(а) Зменшення зенітного кута на 30°   (б) Збільшення зенітного кута на 30°

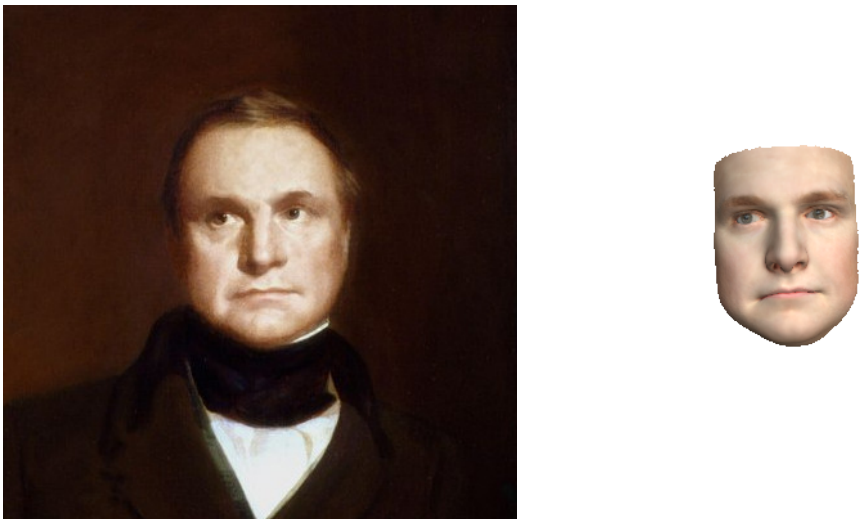
Рисунок 3.3 — Переміщення джерела світла по зеніту

час можна зібрати дані для однієї ітерації розв’язку задачі (2.3). Хоча вказаний час не є прийнятним для кількох десятків кроків алгоритму пошуку субоптимальної стратегії, зібрані дані вкажуть максимальну якість розпізнавання, на яку можна сподіватися при певних значеннях невідомих параметрів, що в свою чергу можна використовувати як відомості про обмеження на умови експлуатації системи реконструкції.

На даний момент цей спосіб являє собою виключно теоретичний інтерес, проте з розвитком комп’ютерної техніки у найближчі роки можна буде провести даний експеримент з більшою точністю за прийнятний час та ціну, що дозволить використовувати субоптимальну стратегію на практиці.

3.2 Портрет Чарльза Беббіджа

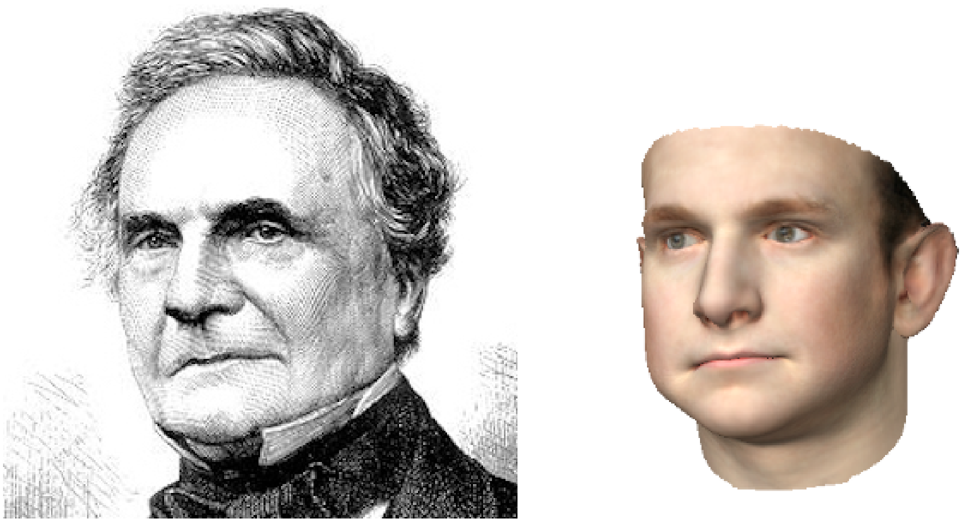
Є портрет відомого англійського математика Чарльза Беббіджа (рис. 3.4а). За допомогою бінарної функції витрат та дискретизації множини параметрів було отримано напрочуд реалістичну модель (рис. 3.4б).



(а) Фрагмент портрету (б) Реконструйована модель

Рисунок 3.4 — 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. 3.5а). До наявної моделі застосовано лише інший поворот та освітлення (рис. 3.5б), 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 3.5 — Чарльз Беббідж

З рисунку видно, що деякі ділянки реконструйовані некоректно. Проте загалом результат задовільний. До отриманої моделі можна застосувати будь-який поворот (рис. 3.6а) та тінь (рис. 3.6б).

Такого результату вдалося досягти за 40 хвилин з використанням 80 параметрів для текстури і 320 параметрів для форми моделі на комп'ютері з процесором Intel Core i3-2328M 2.2GHz та відеокартою nVidia GeForce 620M. Програмне забезпечення було написано на Python 3 спеціально для цієї роботи, для оптимізації використано метод COBYLA (Constrained Optimization by Linear Approximation) [26] з бібліотеки SciPy. Реконструкція з меншою кількістю параметрів, проте з задовільною якістю, потребує лише 3–6 хвилин.

(а) Вигляд збоку (б) Інша тінь

Рисунок 3.6 — Реконструйована модель Чарльза Беббіджа

Висновки до розділу 3

Проведено теоритичний аналіз проблем у використанні запропонованих в минулому розділі методів. Виявилося, що найбільш коректний з теоретичної точки зору метод реконструкції, що використовує метод Монте-Карло, практично неможливо використовувати на даний момент. Баєсова стратегія з інтервальною функцією витрат може бути реалізована досить скоро. Сучасні методи базуються на оцінці найбільшої правдоподібності не тільки тому, що цей підхід дає достатню точність, але тому, що він є одним з небагатьох, який доцільно використовувати.

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 розробити оптимізований метод 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 зроблено коректну постановку та теоретичний розв'язок поставленої задачі. Добре відомі чисельні методи вказали, як обходити обчислювальні складнощі, що виникають при практичному розв'язку задачі.

Було реалізовано демонстративне програмне забезпечення, що дає гарну реконструкцію просторової конфігурації людського обличчя за певних умов. Реалізації деяких методів вказали на необхідність різного роду евристичних підходів та неможливість практичного застосування деяких теоретично вірних міркувань.

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. — P. 108.

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 International Conference on Advanced Video and Signal based Surveillance (AVSS) for Security, Safety and Monitoring in Smart Environments*. — 2009. — Sep.

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 Lovász, L. Submodular functions and convexity / L. Lovász // Mathematical Programming The State of the Art: Bonn 1982 / Ed. by Achim Bachem, Bernhard Korte, Martin Grötschel. — Berlin, Heidelberg: Springer Berlin Heidelberg, 1983. — Pp. 235–257.

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. — Pp. 2387–2395.

12 Large Scale 3D Morphable Models / J Booth, A Roussos, A Ponniah et al. // *International Journal of Computer Vision*. — 2017. — Pp. 1–22.

13 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.

14 Schlesinger, M. I. Nearly Optimal Statistical Recognition and Learning / M. I. Schlesinger, E. V. Volodazkiy // *Proceedings of 4th International conference on Inductive Modelling*. — 2013.

15 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.

16 Айвазян, С.А. Прикладная статистика: Классификация и снижение размерности : Справочное издание / С.А. Айвазян. — М.: Финансы и статистика, 1989. — 606 с.

17 Rabin, Michael O. Probabilistic automata / Michael O. Rabin // *Information and Control*. — 1963. — Vol. 6, no. 3. — Pp. 230–245.

18 Duda, R.O. Pattern Classification and Scene Analysis / R.O. Duda, P.E. Hart. — Wiley, 1973. — P. 512.

19 Berger, J.O. Statistical Decision Theory: Foundations, Concepts, and Methods / J.O. Berger // Springer Series in Statistics. — Springer New York, 1980. — P. 428.

20 Wald, A. Selected Papers in Statistics and Probability / A. Wald, I.M. Statistics // No. 10. — McGraw-Hill, 1955. — P. 702.

21 Боровков, А. А. Математическая статистика / А. А. Боровков // Учебники для ВУЗов. Специальная литература. — 4 изд. — Санкт-Петербург: Лань, 2010. — С. 705.

22 Дороговцев, А. Я. Элементы общей теории меры и интеграла / А. Я. Дороговцев. — К.: Выща школа, 1989. — С. 152.

23 Kasiński, A. The PUT Face Database / A. Kasiński, A. Florek, A. Schmidt // *Image Processing & Communications*. — 2008. — Vol. 13, no. 3–4. — Pp. 59–64.

24 Haykin, Simon. Neural Networks: A Comprehensive Foundation / Simon Haykin. — 2nd edition. — Upper Saddle River, NJ, USA: Prentice Hall PTR, 1998. — P. 842.

25 Moon, H. Computational and performance aspects of pca-based face-recognition algorithms / H. Moon, P. J. Phillips // *Perception*. — 2001. — Vol. 30. — Pp. 303–321.

26 Powell, M. J. D. A Direct Search Optimization Method That Models the Objective and Constraint Functions by Linear Interpolation / M. J. D. Powell //

Advances in Optimization and Numerical Analysis. — Dordrecht: Springer Netherlands, 1994. — Pp. 51–67.