

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Інфокомунікацій
(повна назва)

Кафедра Інформаційно-мережної інженерії
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА Пояснювальна записка

рівень вищої освіти перший (бакалаврський)
Дослідження методів сегментації зображень документів

(тема)

Виконав:
здобувач 4 року навчання,
групи ТРИМІ-21-2
Ростислав Григор`єв
(власне ім'я, прізвище)

Спеціальність 172 Телекомунікації та
радіотехніка
(код і повна назва спеціальності)

Тип програми освітньо-професійна
Освітня програма Інформаційно-мережна
інженерія
(повна назва освітньої програми)

Керівник ст. викл. Олексій Федоров
(посада, власне ім'я, прізвище)

Допускається до захисту

Завідувач кафедри ІМІ
(підпис)

Валерій Безрук
(власне ім'я, прізвище)

2025 р.

Не містить відомостей, заборонених до відкритого публікування

Студент _____ / Григор`єв Р.О. /
(підпис) (прізвище та ініціали)

Керівник _____ / Федоров О.В /
(підпис) (прізвище та ініціали)

Харківський національний університет радіоелектроніки

Факультет Інфокомунікацій
Кафедра Інформаційно мережної інженерії
Рівень вищої освіти перший (бакалаврський)
Спеціальність 172 Телекомунікації та Радіотехніка
(код і повна назва)
Тип програми освітньо-професійна
Освітня програма Інформаційно-мережна інженерія
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)
«___» _____ 20__ р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

здобувачеві Григор'єву Ростиславу Олександровичу
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження методів сегментації зображень документів

затверджена наказом університету від 23 травня 2025 р. № 410Ст

2. Термін подання здобувачем роботи до екзаменаційної комісії 19 червня 2025 р.

3. Вихідні дані до роботи Дослідити методи сегментації зображень документів. Ознайомитись з гістограмним методом, що використовується для бінарзації зображень. Особливу увагу приділити використанню X-Y дерев та модифікованим X-Y деревам, що дозволяють врахувати інформацію про структуру даних документа

4. Перелік питань, що потрібно опрацювати в роботі _____

Вступ

1 Загальні відомості про машинний зір

2 Формування й уведення зображень

3 Області й сегментація зображення

4 Використання операторів згортки для виділення границь на зображенні

5 Сегментація структурованих документів

Висновки

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (п.5 включається до завдання за рішенням випускової кафедри) _____
 слайди презентації в форматі MS Power Point

6. Консультанти розділів роботи (п.6 включається до завдання за наявності консультантів згідно з наказом, зазначеним у п.1)

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Строк / терміни виконання етапів роботи	Примітка
1.	Ознайомлення із завданням. Уточнення ТЗ.	26.05.2025	вик.
2.	Підбір літератури за темою роботи	26.05 – 28.05.2025	вик.
3.	Виконання розділу 1	28.05 – 02.06.2025	вик.
4.	Виконання розділу 2	30.05 – 6.06.2025	вик.
5.	Виконання розділу 3	05.06 – 10.06.2025	вик.
6.	Виконання розділу 4	09.06 – 13.06.2025	вик.
7.	Виконання розділу 5	04.06 – 12.06.2025	вик.
8.	Оформлення презентаційного матеріалу, підготовка до захисту в ЕК	12.06 – 17.06.2025	вик.

Дата видачі завдання 26 травня 2025р.

Здобувач _____
 (підпис)

Керівник роботи _____ ст. викл. Олексій Федоров
 (підпис) (посада, власне ім'я, прізвище)

РЕФЕРАТ

Пояснювальна записка: 65 с., 20 рис., 18 джерел, 1 додаток.

Об'єктом дослідження є методи сегментації зображень документів.

Мета кваліфікаційної роботи – дослідження методів сегментації цифрових зображень документів з використанням X-Y дерев та модифікованих X-Y дерев.

Робота присвячена методам сегментції зображень документів. Розглядається низхідний підхід до сегментації та представлення документів, що містять табличні структури. Прикладами таких документів є рахунки-фактури та технічні документи з таблицями. Сегментація базується на залученні розширення апарату X-Y дерев. В якості роздільників (розрізів) між областями виступають або лінії, або зони утворені пробілами. Листя X-Y дерев описують області, що містять однорідну інформацію, та роздільники розрізів. Зв'язки суміжності між листям дерева описують локальні зв'язки між відповідними областями.

СЕГМЕНТАЦІЯ, ЗОБРАЖЕННЯ, СТРУКТУРА, ДОКУМЕНТИ,
ДЕРЕВА, МОДИФІКАЦІЯ

ABSTRACT

Explanatory note: 65 p., 20 fig., 18 sources, 1 app.

Object of study is methods for document image segmentation.

Purpose of the qualification work is to study methods for segmenting digital images of documents using X-Y trees and modified X-Y trees.

The work is devoted to methods of document image segmentation. A top-down approach to segmentation and representation of documents containing tabular structures is considered. Examples of such documents are invoices and technical documents with tables. Segmentation is based on the use of an extension of the X-Y tree apparatus. The dividers (cuts) between regions are either lines or blanks. The leaves of X-Y trees describe the regions containing homogeneous information and the dividers of the sections. Adjacency links between tree leaves describe local connections between the corresponding regions.

SEGMENTATION, IMAGE, STRUCTURE, DOCUMENTS, TREES,
MODIFICATION

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ.....	9
ВСТУП	10
1 ЗАГАЛЬНІ ВІДОМОСТІ ПРО МАШИННИЙ ЗІР	11
1.1 Машинний зір	11
1.2 Завдання, що розв'язуються системою машинного зору.....	12
1.3 Зв'язок з іншими областями	13
2 ФОРМУВАННЯ Й УВЕДЕННЯ ЗОБРАЖЕНЬ	16
2.1 Два аспекти формування зображення	17
2.1.1 Центральна проекція	17
2.1.2 Ортогональна проекція	19
2.2 Яскравість.....	21
2.3 Лінзи.....	24
2.5 Уведення зображення.....	29
2.5.1 Сприйняття кольору	30
2.5.2 Випадковість і шум.....	31
3 ОБЛАСТІ Й СЕГМЕНТАЦІЯ ЗОБРАЖЕННЯ	33
3.1 Методи граничного розділення.....	33
3.2 Побудова гістограми	35
3.3 Просторовий взаємозв'язок	38
3.4 Сегментація зображень	41
3.5 Використання кольору	42
3.6 Нарощування й розбивка областей.....	43
4 ВИКОРИСТАННЯ ОПЕРАТОРІВ ЗГОРТКИ ДЛЯ ВИДІЛЕННЯ ГРАНИЦЬ НА ЗОБРАЖЕННІ	46
4.1 Методи виділення границь на зображенні.....	47
4.2 Приклад застосування оператора Собеля.....	50
5 СЕГМЕНТАЦІЯ СТРУКТУРОВАНИХ ДОКУМЕНТІВ.....	52
5.1 Загальні відомості.....	52

5.2 Сегментація на базі X-Y дерев.....	53
5.3 Модифікація X-Y дерев	57
5.4 Результати експерименту	59
ВИСНОВКИ.....	63
ПЕРЕЛІК ПОСИЛАНЬ.....	64
ДОДАТОК А СЛАЙДИ ПРЕЗЕНТАЦІЇ.....	66

ПЕРЕЛІК СКОРОЧЕНЬ

КЗ - Комп'ютерний зір

М-Х-У дерево – модифіковане Х-У дерево

OCR - Optical Character Recognition

RLSA - Run Length Smoothing Algorithm

IEEE - the Institute of Electrical and Electronics Engineers

ВСТУП

Машинний зір - це відносно молода галузь знань, що швидко розвивається і спирається на використання нетривіального математичного апарату. Крім того, глибоке розуміння питань, що відносяться до царини машинного зору, вимагає досконального розуміння процесів, що відбуваються при формуванні зображення.

Якщо говорити серйозніше, машинний зір часто розглядають просто як сукупність засобів аналізу візуальної інформації для системи зі штучним інтелектом. Справді, у системах штучного інтелекту він посідає відносно скромне місце через значну його математичну складність. З іншого боку, неправильно розглядати машинний зір і систему керування роботом просто як «вхід-вихід» системи штучного інтелекту. Проблеми, пов'язані із зором, маніпулюванням предметами, автоматизацією опрацювання даних, поданих в текстовому вигляді, цікаві самі по собі, але досить складні й необхідні для їх розв'язання засоби нетривіальні.

Сегментація зображень є складовою частиною машинного зору, оскільки для вдалого аналізу сцени й інтерпретації того, що відбувається навколо необхідно вміти відокремлювати частини зображення одна від одної. В залежності від поставленої задачі, необхідно підбирати відповідні алгоритми сегментації. Дана кваліфікаційна робота бакалавра присвячена вивченню питань сегментації зображень документів. Зокрема розглядається низхідний підхід до сегментації та представлення документів, що містять табличні структури. Прикладами таких документів є рахунки-фактури та технічні документи з таблицями.

1 ЗАГАЛЬНІ ВІДОМОСТІ ПРО МАШИННИЙ ЗІР

1.1 Машинний зір

Зір є самим потужним з наших органів почуттів. Він постачає нам разюче більший обсяг інформації про те, що нас оточує, і дає можливість вільно взаємодіяти із зовнішнім світом, причому все це без безпосереднього фізичного контакту. Завдяки йому ми дізнаємося про розташування об'єктів, ідентифікуємо їх, співвідносимо їх один з одним і тому відчуваємо значну незручність, якщо позбавлені цього почуття. Тому не дивно, що, як тільки цифрові обчислювальні машини виявилися досить доступними, почали робити спроби наділити їх здатністю «бачити» [1, 2].

У той же час зір є самим складним з органів почуттів. Накопичені нами знання про те, як функціонують біологічні зорові системи, дотепер фрагментарні й обмежені в основному тими етапами переробки інформації, які впливають безпосередньо за надходженням сигналів від сенсорів. Що ми дійсно знаємо, так це те, що біологічні зорові системи складні. Тому не є несподіваним той факт, що багато спроб наділити машини здатністю «бачити» закінчилися невдачею. Проте, вдалося досягти значного прогресу, і зараз можна знайти зорові системи, які включені до складу машинних комплексів і успішно функціонують в умовах навколишнього середовища, що змінюється [1, 2]. Такі зорові системи можуть бути побудовані або із застосуванням класичних методів, або спиратись на використання нейронних мереж, які знаходять все більше застосування в завданнях інтелектуального аналізу даних, яким, по суті, є система машинного (комп'ютерного) зору.

1.2 Завдання, що розв'язуються системою машинного зору

Система машинного зору аналізує зображення й видає опис того, що зображено (рис. 1.1). Ці описи повинні містити ті характеристики зображених об'єктів, які потрібні для виконання деякого певного завдання. Тим самим ми розглядаємо систему машинного зору як частину більш загальної системи, що взаємодіє з навколишнім середовищем. Її можна вважати елементом контуру зворотного зв'язку, що відноситься до сприйняття, у той час як інші елементи призначені для прийняття й виконання рішень [1].

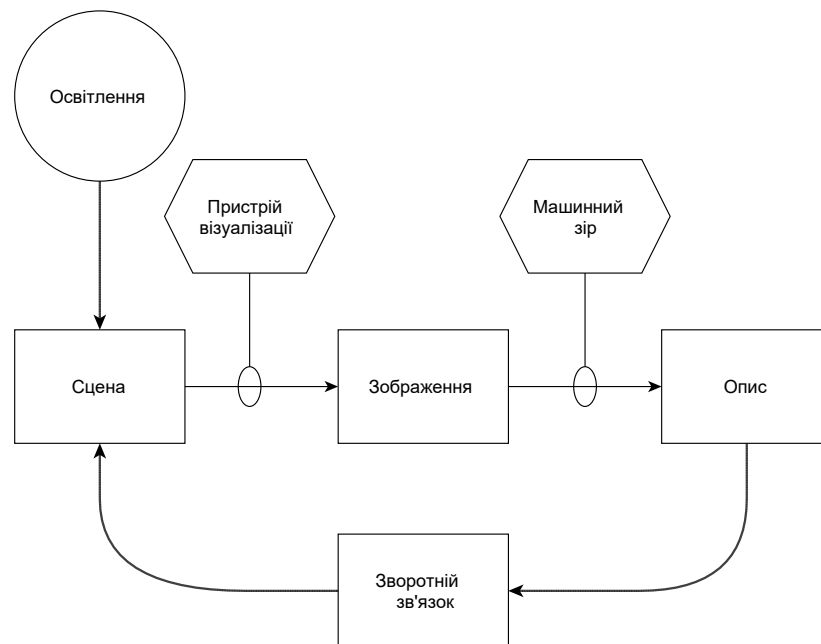


Рисунок 1.1 - Мета роботи системи машинного зору полягає в отриманні символічного опису того, що зображено

Входом системи машинного зору служить одне або кілька зображень, а виходом - опис, що повинен задовольняти двом умовам [1]:

- він повинен бути якимось пов'язаним з тим, що зображено;

- він повинен містити вичерпну інформацію, необхідну для виконання деякого конкретного завдання.

Перша умова забезпечує певну залежність опису від візуального входу, а друга - що отримана інформація є корисною.

Опис об'єкта не є однозначним; ми можемо уявити собі безліч рівнів деталізації й точок зору. Неможливо вичерпним образом описати об'єкт. На щастя, концентруючись на тім завданні, для якого призначений опис, ми можемо уникнути потенційно виникаючого логічного протиріччя, тобто ми не потребуємо просто якогось опису, а тільки такий, котрий дасть можливість зробити відповідну дію [1].

1.3 Зв'язок з іншими областями

Машинний зір тісно стикається із трьома областями (рис. 1.2): [1] обробкою зображень, класифікацією образів і аналізом сцен.

Обробка зображень (Image Processing) в основному стосується одержання нових зображень за вже наявними. Більшість методів, які тут використовуються, запозичено з теорії лінійних систем. На новому зображенні можна послабити шум, усунути розмиття або підкреслити границі. Остаточний результат являє собою зображення, що, як правило, призначено для інтерпретації людиною або машиною [1].

Основним завданням класифікації образів (Pattern Classification) є віднесення «образа» до того або іншого класу, причому образ звичайно представляється у вигляді набору чисел, що відповідають вимірним характеристикам об'єкта (висоті, ширині). Хоча на вхід класифікатора подається не зображення, самі методи класифікації образів виявляються іноді корисними й для аналізу зображень, одержуваних за допомогою зорової системи. Розпізнати об'єкт - це значить віднести його до одного з відомих класів. Однак потрібно підкреслити, що розпізнавання - це лише одне з багатьох завдань, що стоять перед зоровою системою [1].

Дослідники, що займалися розпізнаванням, розробили багато різноманітних методів отримання характеристик за зображеннями.

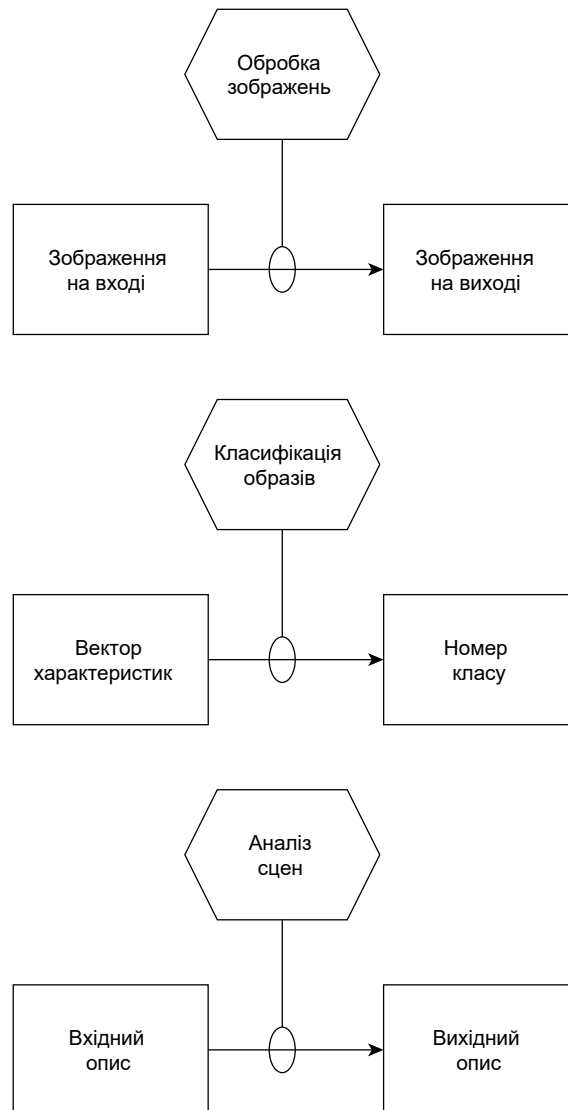


Рисунок 1.2 - Три початкові парадигми машинного зору: обробка зображень, класифікація образів і аналіз сцен

В аналізі сцен (Scene Analysis) розглядається завдання переходу від простих описів, отриманих безпосередньо по зображеннях, до більш складних, представлених у формі, що більше пристосована для виконання певного завдання [1]. Класичною ілюстрацією цього служить інтерпретація контурних рисунків (рис. 1.3). Тут опис зображення сукупності

багатогранників дається у вигляді множини прямолінійних відрізків. Перш ніж його використати, необхідно визначити, які області, обмежені цими відрізками, відносяться до тих самих об'єктів. Також бажано знати, як об'єкти розташовані друг відносно друга. Подібним чином із простого опису можна отримати складний символічний опис зображення [1].

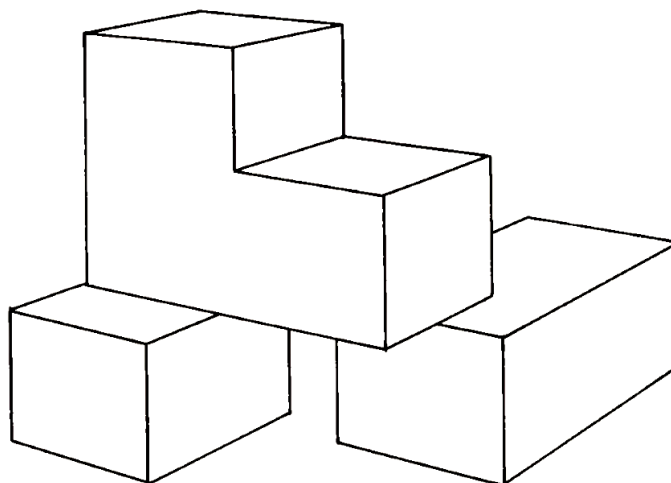


Рисунок 1.3 - В аналізі сцен символічний опис низького рівня (контурні рисунки) використовується для складання опису високого рівня

Значна частина методів аналізу зображень призначена для обробки не стільки всього зображення, скільки областей, що відносяться до окремих об'єктів. Оскільки на зображення, як правило, потрапляє велика кількість поверхонь навколишніх предметів, то перш, ніж використати згадані методи, зображення необхідно розбити на області, що відповідають різним об'єктам, тобто виконати сегментацію зображення [1]. Дана кваліфікаційна робота буде присвячена вивченню особливого виду сегментації зображень: сегментації зображень документів.

2 ФОРМУВАННЯ Й УВЕДЕННЯ ЗОБРАЖЕНЬ

У цьому розділі розглядається питання формування зображення й спосіб його уведення в комп'ютер. Знання процесу формування зображення є необхідною умовою глибокого розуміння методів відновлення інформації із зображень. Аналізуючи процес проектування тривимірного простору на двовимірну площину зображення, ми виділимо два ключових питання, що стосується формування зображення [1]:

- чим визначається положення образу будь якої точки на зображенні;
- чим визначається яскравість будь якої поверхні на зображенні.

Відповіді на ці два запитання потребують знання механізмів відображення спостережуваної сцени на площину зображення і процесів випромінювання, що відбуваються при цьому. Ці теми будуть розглянуті по відношенню до простих оптичних систем, складених з лінз [1].

Вирішальним фактором при вивченні формування зображень служить своєрідність світу, що ми його спостерігаємо. Він має особливі властивості, наявність яких дозволяє відновлювати інформацію про тривимірні сцени на основі одного або декількох двовимірних зображень [1]. Ми обговоримо це питання й укажемо ситуації, в яких така специфіка відсутня отже, витягти інформацію з цих зображень набагато складніше.

Будуть також розглянуті принципи роботи типових зорових пристроїв і способи отримання й обробки інформації для різних частин спектра. Після швидкого ознайомлення з питаннями колірною зору розділ закінчується обговоренням проблеми шуму й нагадуванням деяких понять із теорії ймовірностей і статистики. Тут самий підходящий момент ввести одномірну згортку [1, 2] - математичний прийом, що буде вживатися у вигляді, узагальненому на двовимірний випадок.

2.1 Два аспекти формування зображення

Перш ніж аналізувати зображення, необхідно зрозуміти, як воно формується. Зображення являє собою двовимірну картину розподілу яскравості. Питання про те, як в оптичній зоровій системі виникає ця картина, найкраще вивчати, розбивши його на дві частини: спочатку необхідно знайти відповідність між точками сцени й точками зображення, а потім з'ясувати, від чого залежить яскравість у кожній конкретній точці зображення [1, 2].

2.1.1 Центральна проекція

Уявимо собі камеру із точковим отвором, розташовану на фіксованій відстані від площини зображення (рис. 2.1). Припустимо, що камера влаштована так, що тільки світло, що проходить через отвір, може досягти площини зображення. Оскільки світло поширюється прямолінійно, кожній точці зображення відповідає напрямком, обумовлений променем, що йде із цієї точки через отвір. У результаті виникає добре знайома центральна (перспективна) проекція. На практиці зручно користуватися системою координат, в якій площина xu паралельна площини зображення, а початок координат збігається із точковим отвором O . Вісь z спрямована уздовж оптичної осі [1, 2].

У цьому спрощеному випадку визначимо оптичну вісь як перпендикуляр, опущений з отвору на площину зображення. Тепер ми можемо ввести зручну декартову систему координат з початком в точковому отворі й віссю z , що збігається з оптичною віссю й спрямована убік зображення. При такому виборі напрямку осі точки, що перебувають перед камерою, мають негативну z -координату. Незважаючи на цей недолік, будемо дотримуватися прийнятої домовленості, оскільки завдяки цьому ми маємо у своєму розпорядженні зручну праву систему координат (з віссю x , спрямованою вправо, а віссю y - вгору) [1].

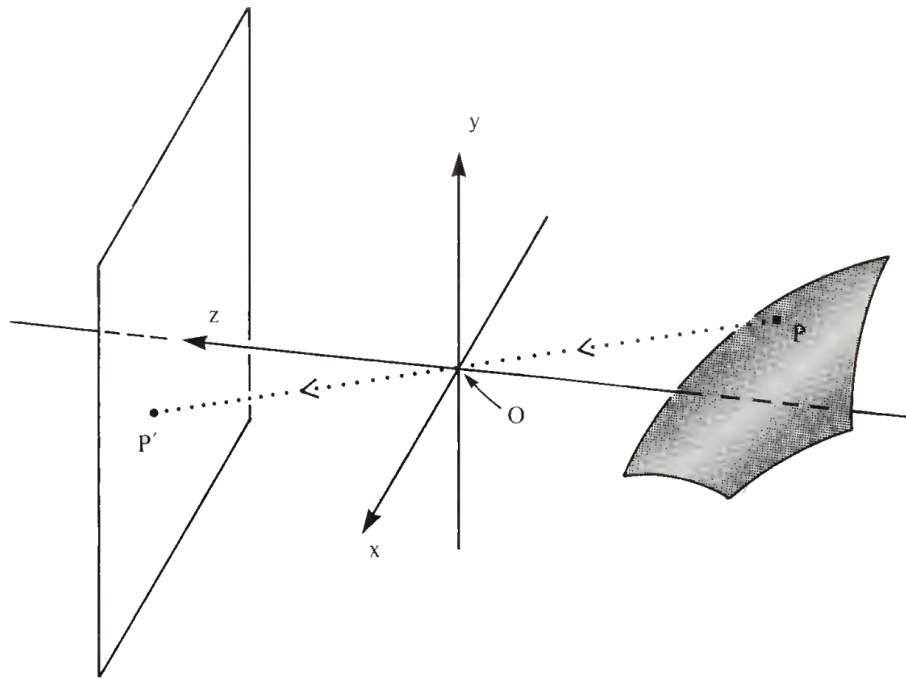


Рисунок 2.1 - Камера із точковим отвором, що дає зображення, що є центральною проекцією

Нам потрібно визначити положення образу P' деякої точки P об'єкта, що перебуває перед камерою (рис. 2.1). Будемо вважати, що ніякий інший об'єкт не перетинає промінь, що йде із точки P в точку O . Нехай $\vec{r} = (x, y, z)^T$ - вектор, що з'єднує O з P , а $\vec{r}' = (x', y', f')^T$ - вектор, що з'єднує O з P' [1]. (Операція транспонування, позначається верхнім індексом T .)

Тут f' - відстань від площини зображення до отвору, а x' й y' - координати точки P' на цій площині. Два вектори \vec{r} й \vec{r}' колінеарні й відрізняються тільки (від'ємним) скалярним множником. Якщо промінь, що з'єднує точки P й P' , утворює кут α з оптичною віссю, то довжина вектора \vec{r} визначається виразом [1]

$$\vec{r} = -z \sec \alpha = -(\vec{r} \cdot \hat{z}) \sec \alpha, \quad (2.1)$$

де \hat{z} - одиничний вектор уздовж оптичної осі. (Варто пам'ятати, що величина z від'ємна для точки, що розташована перед камерою.)

Довжина \vec{r}' становить $r' = f' \sec \alpha$, і тому $(1/f')\vec{r}' = 1/(\vec{r} \hat{z})\vec{r}$. Цей вираз можна представити у вигляді компонентів $x'/f' = x/z$, $y'/f' = y/z$. Іноді для спрощення рівнянь проєкції координати зображення нормалізують, розділивши x' й y' на f' [1].

2.1.2 Ортогональна проєкція

Розглянемо процес формування зображення площини, що паралельна площини зображення й задається рівнянням $z = z_0$. У цьому випадку можна визначити коефіцієнт (поперечного) збільшення m як відношення відстані між двома точками на площині зображення до відстані між відповідними точками на зображуваній площині [1]. Розглянемо малий відрізок $(\delta x, \delta y, 0)^T$ на площині й відповідний йому малий інтервал $(\delta x', \delta y', 0)^T$ на зображенні. Тоді [1]

$$m = \frac{\sqrt{(\delta x')^2 + (\delta y')^2}}{\sqrt{(\delta x)^2 + (\delta y)^2}} = \frac{f'}{-z_0}, \quad (2.2)$$

де $-z_0$ - відстань від площини до отвору. Коефіцієнт збільшення той самий для всіх точок площини (при цьому $m < 1$ для всіх зображень, крім тих, що отримуються за допомогою мікроскопа).

Зображення невеликого об'єкта, розташованого на середній відстані $-z_0$, збільшується в m разів за умови, що зміна z -координати точок його видимої поверхні мало в порівнянні з величиною $-z_0$. Площа зображення об'єкта пропорційна m^2 . Коефіцієнт збільшення об'єктів, розташованих на різних відстанях від зорової системи, буде, природно, різним. Нехай

глибина сцени - це розкид відстаней від видимих поверхонь до камери. Значення коефіцієнта збільшення приблизно постійно, коли глибина спостережуваної сцени мала в порівнянні із середньою відстанню від видимих поверхонь до камери. У цьому випадку ми можемо спростити рівняння проекції в такий спосіб [1]:

$$x' = -m x, \quad y' = -m y, \quad (2.3)$$

де $m = f'/(-z_0)$, а $-z_0$ - середнє значення величини $-z$. Часто для зручності скалярний множник m вважають рівним 1 або -1 . Тоді рівняння ще більше спрощуються й приймають вигляд $x' = x$, $y' = y$. Цю ортогональну (паралельну) проекцію (рис. 2.2) можна моделювати променями, що поширюються паралельно оптичній осі (а не проходять через початок координат). Різниця між центральною й ортогональною проекціями незначна, якщо відстань до сцени в багато разів більша в порівнянні з розкидом відстаней між об'єктами на сцені [1]. Зауважимо, що при ортогональній проекції промені від точок сцени поширюються паралельно напрямку проекції до перетинання із площиною зображення.

Кутом поля зору зорової системи є кут розкриття конуса всіляких променів, спрямованих до об'єктів спостережуваної сцени. Природно, цей конус має ті ж самі форму й розмір, що й конус, утворений прямими, що з'єднують центр проектування із границею площини зображення (екрана). «Нормальні» лінзи характеризуються кутом поля зору величиною приблизно $25 - 40^\circ$. До телескопічних (довгофокусних) лінз відносяться такі лінзи, які мають фокусну відстань, більшу у порівнянні з розмірами зображення, і тому малим кутом поля зору. Навпаки, ширококутні (короткофокусні) лінзи мають фокусну відстань, меншу у порівнянні з розмірами зображення, і тому широким полем зору. Приблизно можна вважати, що ефекти перспективи істотні при використанні ширококутових лінз, у той час як зображення, що отримуються за допомогою

телескопічних лінз, приблизно апроксимуються ортогональною проекцією [1, 2, 10].

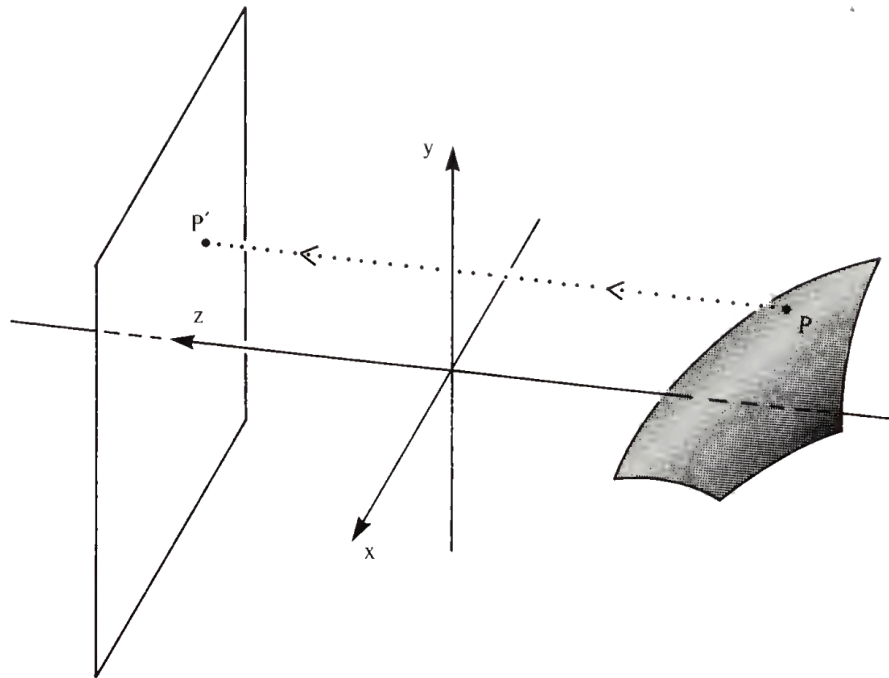


Рисунок 2.2 - Можливість апроксимації центральної проекції ортогональною проекцією, коли глибина сцени мала в порівнянні із середньою відстанню від камери

2.2 Яскравість

До більш складного й цікавого питання процесу формування зображення відносяться фактори, що визначають яскравість кожної конкретної його точки. Яскравість - це неформальний термін, що вживається для позначення принаймні двох різних понять: яскравості зображення і яскравості сцени [1]. При розгляді зображення яскравість пов'язана з потоком енергії, що падає на площину зображення; його величина може вимірятися декількома способами. Тут замість неформального поняття яскравість зображення буде використатися термін освітленість зображення. Освітленість - це потужність (світлової) енергії,

що падає на поверхню, віднесена до одиниці площі ($\text{Вт} \cdot \text{м}^2$ — ват на квадратний метр), див. рис. 2.3, а). На рисунку 2.3 E позначає освітленість, а δP — потужність світлового потоку, що падає на нескінченно малу ділянку поверхні площею δA . У кожній конкретній точці зображення освітленість буде залежати від кількості світла, що попадає в неї з відповідної точки об'єкта (ця точка знаходиться шляхом простежування променя, що виходить із точки зображення і йде через зіницю оптичної системи до перетинання з поверхнею об'єкта) [1].

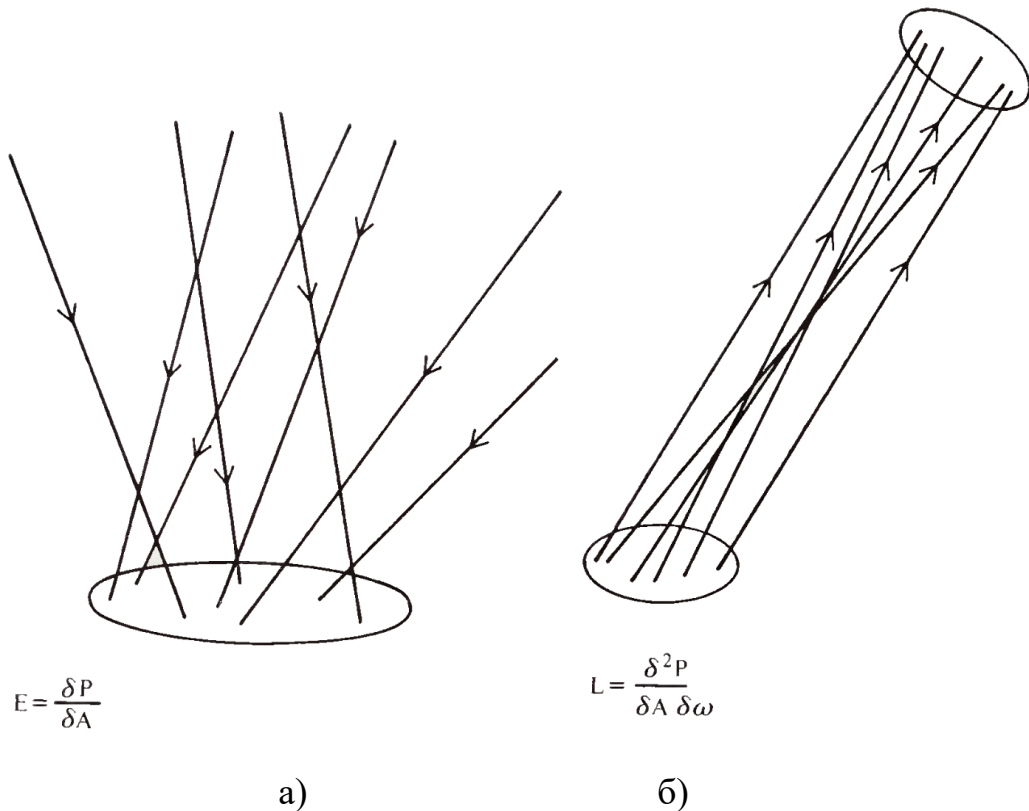


Рисунок 2.3 – Яскравість: а) - освітленість, що є потужністю світлового потоку, що падає на одиницю площі поверхні; б) - енергетична яскравість, що є потужністю світлового потоку, випромінюваного одиницею поверхні й міститься всередині конуса напрямків. Термін яскравість неформально відносять до обох понять

При розгляді сцени яскравість пов'язана з потоком енергії, що випромінюється поверхнею. Різні точки на об'єктах, розташованих перед зоровою системою, будуть мати різну яскравість, що залежить від їхньої освітленості й здатності відбивати світло. Для заміни неформального поняття яскравість сцени вводиться термін енергетична яскравість поверхні [1]. Вона дорівнює потужності енергії, що випромінюється одиницею площі поверхні усередині одиничного тілесного кута ($\text{Вт} \cdot \text{м}^2 \cdot \text{ср}^{-1}$ - ват на квадратний метр на стерадіан), див. рис. 2.3, 6). На рисунку L позначає енергетичну яскравість, а $\delta^2 P$ — потужність потоку, що випромінюється нескінченно малою ділянкою поверхні площі δA всередині нескінченно малого тілесного кута $\delta \omega$ [1]. Уявна складність визначення яскравості поверхні виникає з того, що множина можливих напрямків випромінювання світла поверхнею утворює півсферу, і тому величина яскравості виявляється ненульовою лише при розгляді кінцевого тілесного кута цих напрямків. У загальному випадку яскравість поверхні буде змінюватися залежно від напрямку, під яким спостерігається об'єкт [1].

Яскравість ділянок поверхні об'єкта нас цікавить остільки, оскільки ми вимірюємо освітленість зображення, що пропорційна яскравості сцени. Коефіцієнт пропорційності залежить від властивостей оптичної системи. Щоб зібрати на екрані кінцеву кількість світла, необхідно мати апертуру кінцевих розмірів. Розглянутий у попередньому підрозділі точковий отвір повинен мати ненульовий діаметр. Тому вивчена вище спрощена схема проєктування відтепер стає неприйнятною, оскільки точка сцени тепер зображується у вигляді невеликого кола. Останнє впливає з розгляду променів, що проходять крізь круглу апертуру, які утворюють конус із вершиною в точці об'єкта [1].

Ми не можемо зробити апертуру занадто малою й з іншої причини. Через хвильову природу світла на її краях виникає явище дифракції, і

світло «розтікається» по всьому зображенню. В міру зменшення діаметру апертури все більш значна частина світла, що проходить, відхиляється від первісного напрямку.

2.3 Лінзи

Щоб уникнути проблем, пов'язаних з малістю діаметра отвору камери, розглянемо можливість використання лінз у системі формування зображень [1]. Ідеальна лінза забезпечує той же самий тип проекції, що й точковий отвір, але в той же час збирає кінцеву кількість світла (рис. 2.4). Чим більше лінза, тим більший тілесний кут вона охоплює (якщо дивитися з боку об'єкта), і відповідно тим більше відбитого (або випромененого) об'єктом світла вона збирає. Промінь, що проходить через центр лінзи, не відхиляється. В добре сфокусованій системі інші промені відхиляються так, щоб досягти тієї ж самої точки зображення, що й центральний промінь [1].

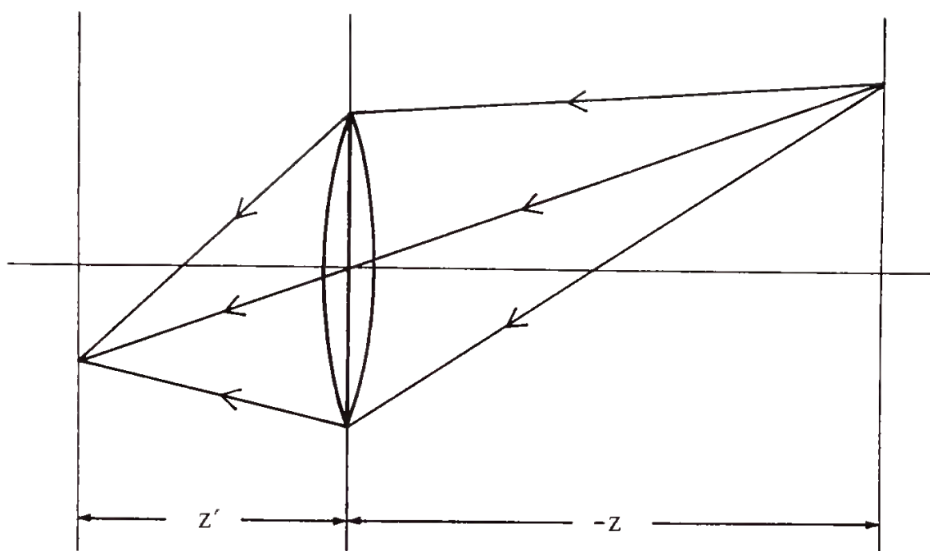


Рисунок 2.4 - Використання лінзи замість ідеального точкового отвору для одержання ненульової освітленості в площині зображення

Ідеальна лінза володіє тим недоліком, що вона фокусує тільки світло, що випромінюється точками, що перебувають на відстані $-z$, що визначається відомим рівнянням лінзи $1/z' + 1/(-z) = 1/f$, де z' - відстань від площини екрана до лінзи, а f - фокусна відстань (рис. 2.4) [1]. Точки, що перебувають на інших відстанях, зображуються у вигляді невеликих кіл. У цьому можна переконатися, якщо розглянути конус світлових променів, що проходять крізь лінзу, з вершиною в точці, в якій вони точно сфокусовані. Розмір кола нерізкості можна визначити в такий спосіб: точка на відстані $-\bar{z}$ зображується точкою на відстані \bar{z}' від лінзи, причому $1/\bar{z}' - 1/\bar{z} = 1/f$, і тому [1]

$$(\bar{z}' - z') = \frac{f}{(\bar{z} + f)} \frac{f}{(z + f)} (\bar{z} - z). \quad (2.4)$$

Якщо площину екрана розташувати так, аби виходили точно сфокусовані зображення об'єктів, що перебувають на відстані $-z$, то точки, що перебувають на відстані $-\bar{z}$, призведуть до появи кіл нерізкості діаметром [1]

$$(d/z') \cdot |\bar{z}' - z|, \quad (2.5)$$

де d - діаметр лінзи. Глибиною різкості називається діапазон відстаней, на яких об'єкти фокусуються «досить добре» у тому розумінні, що діаметр кола нерізкості менше роздільної здатності зорового пристрою. Звичайно, глибина різкості залежить від сприймаючої апаратури, однак у будь-якому випадку ясно, що чим ширше апертура лінзи, тим менше глибина різкості. Ясно також, що при використанні широкої апертури помилки фокусування стають більше значними [1, 2, 10].

Розумінню властивостей оптичних систем, утворених з декількох лінз, можуть допомогти прості правила трасування променів. Як вже згадувалося, промінь, що проходить крізь центр лінзи, не відхиляється. Промені, що надходять у лінзу паралельно оптичній осі, сходяться в точці, що перебуває на оптичній осі на фокусній відстані. Це впливає з визначення фокусної відстані як відстані, на якій фокусується зображення нескінченно віддаленого об'єкта. І навпаки, промені, що виходять із точки оптичної осі, що перебуває на відстані, що дорівнює фокусній відстані, відхиляються й стають паралельними оптичній осі по іншу сторону лінзи. Це впливає із властивості зворотності поширення променів [1]. На межі двох середовищ із різними показниками заломлення утворюються одні й ті самі кути відбиття та заломлення незалежно від напрямку поширення світла [1].

Проста лінза виточується зі скляної заготовки і полірується так, щоб обидві її поверхні набули сферичної форми. Оптична вісь є прямою, що проходить крізь центри двох сфер. Будь-яка проста лінза матиме ряд дефектів або аберацій. Тому часто комбінують кілька простих лінз, ретельно вирівнюючи їхні оптичні осі, щоб одержати композитну лінзу із кращими характеристиками [1].

Зручною моделлю такої системи лінз є товста лінза (рис. 2.5). Для такої лінзи можна визначити дві головні площини, перпендикулярні оптичній осі, і дві головні точки, в яких ці площини перетинають оптичну вісь. Промінь, що проходить крізь передню головну точку, виходить через задню головну точку, не змінюючи свого напрямку. Ця властивість визначає той тип проєкції, що реалізується лінзою. Відстань між двома головними точками називається товщиною лінзи [1]. Тонка лінза - це така лінза, у якій дві головні точки можна розглядати як такі, що співпадають.

Навіть теоретично неможливо виготовити ідеальну лінзу. Проєкція ніколи не буде точно такою, як проєкція ідеального точкового отвору. Ще важливіше те, що неможливо досягти точного фокусування всіх променів.

Виникають різного роду аберації. У добре розроблених лінзах ці дефекти зводяться до мінімуму, однак при збільшенні апертури лінзи домогтися цього стає усе важче [1]. Таким чином, існує компроміс між здатністю лінзи збирати світло (світлосилою) і якістю зображення.

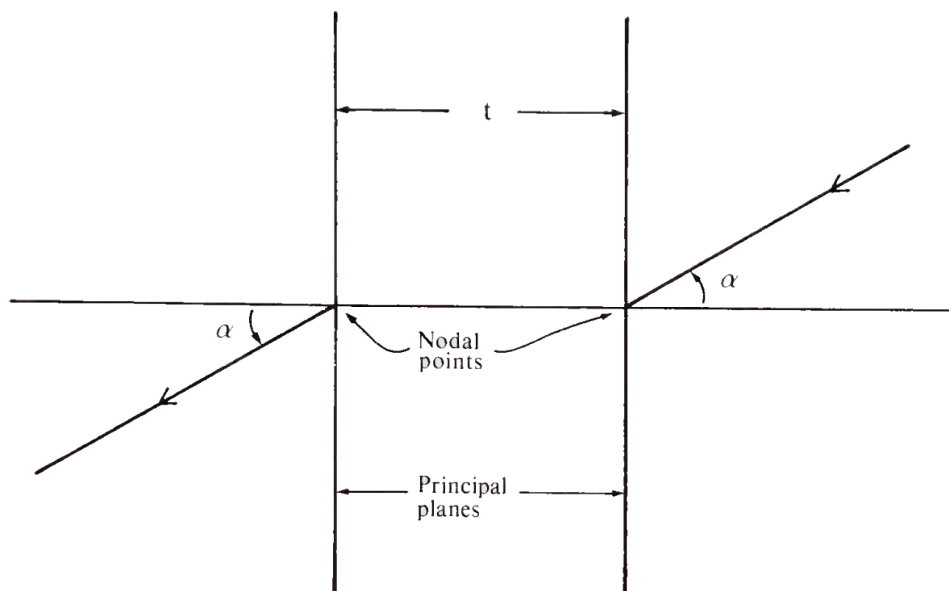


Рисунок 2.5 - Ідеальна товста лінза, що є прийнятною моделлю для більшості реальних лінз

Дефект, що становить для нас у цьому випадку особливий інтерес, називається віньєтуванням [1]. Уявіть собі кілька кругових діафрагм різного діаметра, встановлених одна за іншою так, що їхні центри лежать на одній прямій (рис. 2.6). Якщо дивитися уздовж цієї прямої, то погляд буде обмежувати діафрагма з мінімальним діаметром. Якщо переміщатися вбік від лінії, то сильніше обмежувати погляд почнуть інші діафрагми, поки, нарешті, взагалі нічого не буде видно. В простих лінзах всі промені, що потрапили на передню поверхню, так чи інакше фокусуються на зображенні. В складених лінзах деякі промені, що пройшли першу лінзу, можуть частково затримуватися другою лінзою й т.д. Результат залежить від нахилу падаючого променя до оптичної осі і його відстані від передньої

головної точки [1]. Таким чином, здатність лінзи збирати світло виявляється слабшою для точок зображення, що не належить оптичній осі, ніж для точок, що належать оптичній осі. Чутливість зменшується з відстанню від центра зображення.

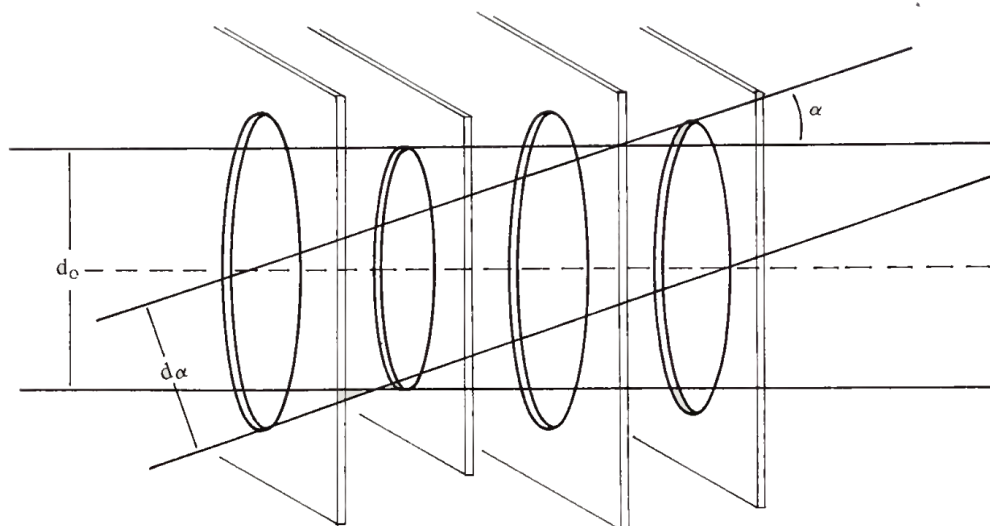


Рисунок 2.6 - Виньетування, що представляє собою зменшення збірної сили лінзи зі збільшенням нахилу променів щодо оптичної осі

Інша важлива властивість полягає в тім, що величина аберації лінзи зростає як ступінь кута між падаючим променем і оптичною віссю. Аберації класифікуються за їхнім порядком, тобто показником ступеня, що фігурує в цій залежності. Точки на оптичній осі можна добре сфокусувати, в той час як кутові точки зображення будуть змазані [1]. Через це додатною виявляється тільки обмежена частина площини зображення. Крім того, величина аберації збільшується як ступінь відстані від точки падіння променя на передню поверхню лінзи до оптичної осі. Тому якість лінзи можна поліпшити, якщо використати тільки її центральну частину [1].

Одна з причин встановлення діафрагм в оптичній системі саме й полягає в тому, щоб поліпшити якість зображення тоді, коли немає необхідності повністю використати збірну силу лінзи. Як ми вже відзначали, установка діафрагми забезпечує те, що промені, що падають під більшим кутом до оптичної осі, не пройдуть крізь периферійну зону ні однієї з лінз. Це підвищує якість зображення по краях, однак одночасно значно підсилює віньєтування [1]. Як правило, при використанні лінз це не має великого значення, оскільки люди дивовижним чином несприйнятливі до поступової зміни яскравості зображення, однак це важливо в системах машинного зору, оскільки в них по вимірах яскравості зображення (освітленості) визначається яскравість сцени (енергетична яскравість поверхні) [1].

Ефект віньєтування можна пояснити інакше: це явище викликане апертурами системи лінз, які загороджують частину пучка світла в міру його проходження крізь систему. Віньєтування призведе до гладкого, але іноді досить значного спаду чутливості при наближенні до країв області зображення [1].

2.5 Уведення зображення

Робота майже всіх пристроїв отримання зображень заснована на ефекті утворення електронно-діркових пар при бомбардуванні фотонами підходящого матеріалу. Це основний процес як у зорі тварин, так і у фотографії. Пристрої відрізняються методом виміру потоку заряджених часток. В деяких з них для виривання електронів з поверхні, всередині якої вони звільняються, використовуються електричне поле й вакуум. В інших електрони протікають крізь збіднений шар (d) $n-p$ -напівпровідника (рис. 2.7) [1].

Не всі падаючі фотони породжують електронно-діркові пари. Деякі проходять чутливий шар наскрізь, деякі відбиваються, інші по різних

причинах втрачають свою енергію. Більше того, не всі звільнені електрони попадають у вимірювальне коло. Відношення числа звільнених електронів до числа падаючих фотонів називається квантовим виходом і позначається через $q(\lambda)$ [1]. Квантовий вихід залежить від енергії падаючих фотонів і, отже, від їхньої довжини хвилі λ . Він також залежить від типу матеріалу й методу, що використовується для збору звільнених електронів. Старі вакуумні прилади, як правило, мають поверхневий шар з відносно низьким квантовим виходом, в той час як напівпровідникові пристрої для деяких довжин хвиль близькі до ідеального. Фото- або кіноплівки звичайно мають низький квантовий вихід [1, 10].

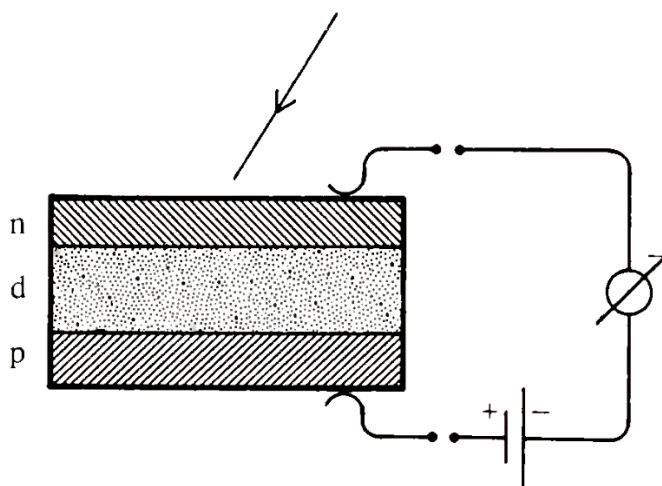


Рисунок 2.7 - Фотони, що бомбардують підходящу поверхню, генерують носії заряду, які збираються й вимірюються для визначення освітленості

2.5.1 Сприйняття кольору

Чутливість пристрою змінюється залежно від довжини хвилі падаючого світла. Фотони з малою енергією мають тенденцію просто проходити крізь матеріал, у той час як фотони з дуже великою енергією можуть затримуватися ще до досягнення чутливого шару. Кожний

матеріал має свою характеристику залежності квантового виходу від довжини хвилі [1].

Для малого інтервалу довжин хвиль $\delta\lambda$ позначимо через $b(\lambda)\delta\lambda$ потік фотонів з довжиною хвилі, яка дорівнює або є більшою від λ , але меншою від $\lambda + \delta\lambda$. Тоді число звільнених електронів визначається інтегралом [1]

$$\int_{-\infty}^{\infty} b(\lambda)q(\lambda)d\lambda. \quad (2.6)$$

Якщо ми використаємо пристрої з різними фоточутливими матеріалами, то через їхню різну чутливість отримаємо різні зображення. Це може допомогти під час розпізнавання поверхонь, що мають схожі напівтонові зображення, отримані за допомогою одного пристрою, але дають різні напівтонові зображення під час використання іншого. Інший шлях досягнення того ж ефекту полягає у використанні того самого чутливого матеріалу, але розміщенням перед камерою фільтрів, які мають селективне поглинання різних частин спектра [1]. Якщо спектральний коефіцієнт пропускання i -го фільтра становить $f_i(\lambda)$, то результуючий квантовий вихід комбінації даного приладу з фільтром - $f_i(\lambda) \cdot q(\lambda)$ [1].

Скільки різних фільтрів необхідно використати? Чим більше зображень отримано за допомогою більшої кількості фільтрів, тим вище здатність розпізнавання матеріалів. Однак виміри корелюють, оскільки більшість поверхонь має гладку залежність відбивної здатності від довжини хвилі. Вигода від використання великої кількості фільтрів звичайно невелика [1, 2, 10].

2.5.2 Випадковість і шум

Важко точно виміряти яскравість зображення. Розглянемо коротко спотворення зображення, що викликані шумом. На результати вимірів

впливають флуктуації у вимірюваному сигналі. При повторному вимірі результати можуть виявитися трохи іншими. Звичайно вони розташовуються біля «справжнього» значення. Ми можемо говорити про ймовірність влучення результату виміру в певний інтервал. Грубо кажучи, вона відповідає границі відношення числа вимірів, що попадають у цей інтервал, до загального числа випробувань при наближенні останнього до нескінченності [1].

3 ОБЛАСТІ Й СЕГМЕНТАЦІЯ ЗОБРАЖЕННЯ

Багато алгоритмів машинного зору розраховані на використання відносно зображення якої-небудь однієї поверхні. Якщо зображення поверхні не заповнює всього поля зору, то увагу необхідно зосередити на тій області, що відповідає цій поверхні. У даному розділі розглядаються методи сегментації зображень на області, які повинні являти собою зображення різних поверхонь. Для цього ми насамперед звернемося до гістограм напівтонових рівнів і спробуємо використати просторовий взаємозв'язок, що проявляється на зображеннях, об'єктів. Зазначимо, що завдання виділення різних областей зображення спрощується, якщо є можливість одержати зображення, що нас цікавлять, в різних областях спектра. Іншими словами, інформація про колір допомагає справі [1, 2].

Будемо розрізняти методи, засновані на підрозділі наявних областей, від методів, що комбінують ці області [1]. У багатьох випадках тільки після виділення істотної інформації про зображені об'єкти можна досить надійно зробити сегментацію. На жаль, більшість методів добування подібної інформації, як правило, вимагає, щоб зображення вже було сегментовано.

3.1 Методи граничного розділення

Щоб одержати бінарне зображення з напівтонового, необхідно встановити граничне значення (рис. 3.1). Елементи зображення, в яких рівень яскравості вище граничного, у відповідних позиціях бінарного зображення дають одиниці, а елементи, у яких рівень нижче, - нулі (або навпаки). Як вибрати це граничне значення? Очевидно, нам потрібно знайти метод автоматичного вибору порога, що враховує як зміни освітленості, так і особливості зображеної поверхні [1, 2]. Один зі шляхів

полягає в аналізі розподілу напівтонових рівнів незалежно від їхнього положення на зображенні.

У випадку неперервного зображення ми можемо визначити щільність розподілу яскравості $p(x)$. При малих δx величина $p(x)\delta x$ - частка тої частини зображення, у якій яскравість більше або дорівнює x , але менше $x + \delta x$. Інтегрування дає інтегральну функцію розподілу яскравості $P(x)$. Для заданого значення x величина $P(x)$ - частка тої частини зображення, у якій яскравість менше або дорівнює x [1]:

$$P(x) = \int_0^x p(t)dt. \quad (3.1)$$



Рисунок 3.1 - Бінарне зображення, що легко формується шляхом граничного розділення значень яскравості тла та об'єкта

У дискретному випадку ми можемо побудувати (диференціальну) гістограму розподілу півтонів, що дає число елементів зображення, що мають заданий рівень яскравості. Інтегральна гістограма розподілу півтонів отримується з неї послідовним підсумовуванням [1].

3.2 Побудова гістограми

Уявіть собі ідеальний випадок, коли об'єкт і тло мають рівномірну яскравість. У цьому випадку поріг вибрати дуже легко, якщо тільки значення яскравості об'єкта й тла різні. Можна взяти будь-яку величину, розташовану між цими значеннями, наприклад їх середнє арифметичне [1].

В дійсності елементи зображення (пікселі), що відповідають об'єкту, не будуть мати точно той самий рівень яскравості (рис. 3.2). Це обумовлено низкою чинників, зокрема наявністю шуму. Визначення рівня яскравості включає вимір відносно невеликого числа електронів, що звільняються падаючими фотонами, що приводить до деяких флуктуацій у вимірюваних струмах. Значний шум вносить підсилювач струму. Цей шум проявляється в розмиванні функції розподілу ймовірності; точніше: щільність розподілу ймовірності, яку б ми одержали при відсутності шуму, згортається з щільністю розподілу адитивного шуму. Крім того, датчики в різних точках звичайно реагують по-різному, а об'єкти не володіють у точності постійною відбивною здатністю або рівномірним освітленням. В силу цих і інших причин ми повинні очікувати наявності розкиду в рівнях яскравості як для об'єкта, так і для тла [1].

Якщо нам пощастить, то цей розкид виявиться настільки малим, що все одно буде існувати граничне значення, що розділяє дві групи рівнів. Його можна знайти, якщо побудувати гістограму рівнів [1, 2].

На гістограмі можна побачити два сплески, які грубо відповідають об'єкту й тлу, і западину між ними (рис. 3.2). В ідеалі між сплесками буде повний провал, але якщо навіть дві групи рівнів трохи перекриваються, за граничне значення можна прийняти мінімальний рівень, що відповідає мінімуму гістограми. Це дозволить зробити бінарізацію зображення лише з незначним числом помилок [1, 2].

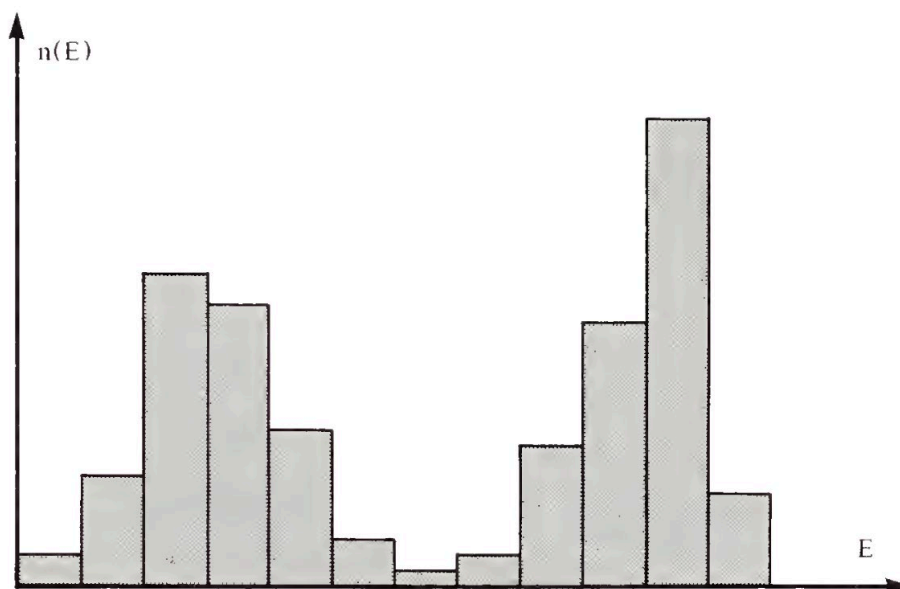


Рисунок 3.2 - Гістограма напівтонів, яка іноді виявляється корисною для визначення порога, що дозволяє розбити зображення на області, тут $n(E)$ - число елементів зображення, що мають рівень яскравості E

Якщо «сигнал» (різниця між рівнями яскравості об'єкта й тла) ненабагато перевершує «шум» (розкид у розподілі рівнів яскравості), то подібним чином успіху, звичайно, домогтися не можна. У цьому випадку може допомогти усереднення значень сусідніх елементів. Це знизить шум пропорційно квадратному кореню із числа усереднених елементів. На рівень сигналу така процедура не вплине, але просторова роздільність погіршиться [1].

Дотепер ми припускали, що кожне значення гістограми відповідає одному рівню яскравості. Можна побудувати більш грубу гістограму, поєднуючи сусідні рівні. При побудові гістограми завжди виникає важка дилема: яка повинна бути ширина інтервалів? Кожному інтервалу відповідає число елементів зображення, рівень яскравості яких лежить між його нижньою й верхньою границями. Якщо інтервали занадто широкі, то роздільна здатність у визначенні рівнів є низькою (у граничному випадку є один інтервал і нульова роздільна здатність) [1]. Якщо ж інтервали занадто

вузькі, то в кожний з них потрапить дуже мало значень і гістограма виявиться сильно порізаною (у граничному випадку майже всі інтервали, крім деяких, не будуть містити жодного виміру). У цьому останньому випадку важко локалізувати положення мінімуму. Природно, що подібний ефект виражений слабкіше, коли об'єкт займає значну площу. Гістограму часто доводиться згладжувати, поєднуючи сусідні інтервали, але це погіршує роздільну здатність [1].

Мають місце й інші труднощі, пов'язані з елементами зображення, розташованими на границі об'єкта (рис. 3.3). Такі елементи будуть мати проміжні значення рівнів, що призведе до змазування й злиття країв двох сплесків. Ступінь прояву ефекту залежить від частки елементів зображення, що попадають на границі. У свою чергу ця частка обернено пропорційна квадратному кореню з відношення площі об'єкта до площі одного елемента зображення. Таким чином, проблема не настільки гостра при високій роздільній здатності зображення об'єкта. І навпаки, вона збільшується при застосуванні методу усереднення, запропонованого вище для збільшення відношення сигнал/шум [1, 2].

Інша проблема виникає, коли площа частини зображення, що займає об'єкт, значно більше або значно менше площі тла. При цьому менший сплеск на гістограмі може виявитися подавленим краєм основного сплеску. Мінімального значення, що відповідає порогу, може не знайтися. Необхідно також пам'ятати, що гістограма буде «зашумленою» через те, що число наявних відліків скінчене. Тому, якщо навіть розподіл яскравості, що лежить в основі зображення, являє собою гладку функцію з яскраво вираженими максимумами, не можна очікувати того ж самого стосовно гістограми рівнів [1,2].

Якщо відома частина площі, займаної об'єктом, то можлива проста модифікацію нашого методу. Насамперед побудуємо інтегральну гістограму підсумовуванням значень диференціальної гістограми. Як згадувалося вище, інтегральна гістограма для кожного значення яскравості

дає число елементів зображення з тим же або меншим значенням яскравості. Оберемо граничне значення таким, щоб число елементів зображення, що мають граничну або меншу яскравість, становило задану частку від усього числа елементів [1]. Зауважте, що тепер вже ми не можемо скористатися обчисленням значення площі бінарного зображення, оскільки ми її задали при визначенні порога!

	10	10	10	10	10	10	10	10	10	17
	10	10	10	10	10	10	10	13	23	29
	10	10	10	10	10	10	18	27	30	30
	10	10	10	10	14	24	30	30	30	30
	10	10	10	19	28	30	30	30	30	30
	10	16	25	30	30	30	30	30	30	30
	22	29	30	30	30	30	30	30	30	30

Рисунок 3.3 - Елементи зображення, що перетинають границю між двома областями (Вони будуть мати проміжні рівні яскравості щодо рівнів яскравості цих областей)

3.3 Просторовий взаємозв'язок

На типових зображеннях рівні яскравості сусідніх елементів не є незалежними. Наприклад, можна заздалегідь передбачити, що зображення багатогранного об'єкта буде містити області із практично постійною яскравістю. Аналогічно яскравість зображень об'єктів із гладкою

поверхнею змінюється поступово [1]. Ми скористаємося таким просторовим взаємозв'язком декількома способами.

Під час сегментації зображення обов'язково знайдеться трохи неправильно класифікованих елементів (в основному завдяки шуму). Якщо відношення сигнал/шум мало, майже всі ці помилки будуть ізольовані, за винятком елементів, розташованих поблизу границі об'єкта. Такий ефект називається крапчастим шумом (рис. 3.4). Після порогового розділення подібні помилки неважко усунути, якщо врахувати, що сусідні елементи будуть мати протилежне значення [1].

Однак можна спробувати на першому ж етапі уникнути виникнення такого сорту помилок шляхом дослідження рівнів яскравості сусідніх елементів до граничного поділу. Якщо рівень яскравості якогось елемента зображення вище максимального рівня його сусідів, то, імовірно, він був підданий впливу шуму. Це - «викид». Його необхідно замінити максимальним рівнем яскравості сусідніх елементів. Аналогічну «фільтруючу» операцію можна застосувати, якщо рівень яскравості елемента нижче мінімального рівня його сусідів. Помітьте, що після такої операції напівтони елементів зображення, що перебувають на границі між об'єктом і тлом, залишаться без зміни [1].

Зниження шуму можна досягти й за допомогою інших фільтруючих операцій. Ми вже згадували операцію усереднення за значеннями сусідніх елементів. Вона приводить до згладжування особливостей і тому зменшує просторову роздільну здатність. В результаті ускладнюються проблеми, що зустрічаються на границях. Інша можливість полягає в заміні кожного значення рівня не середнім, а медіанним значенням сусідніх елементів. (Кількість елементів множини, що лежать нижче й вище медіани, рівна.)

Як і в першій схемі, така операція не змінює значень елементів на границі. Однак обчислити медіану складніше, ніж максимальне або середнє значення [1, 2].

0	0	0	0	0	0	0	0	1	1
0	0	1	0	0	0	1	1	1	1
0	0	0	0	0	1	1	1	1	1
0	1	0	0	0	1	1	0	1	1
0	0	0	1	1	1	1	1	1	1
0	0	1	1	1	0	1	1	1	1
1	1	1	1	1	1	1	1	1	1

Рисунок 3.4 - Крапчастий шум, що виникає в результаті накладення помилок. Останні пов'язані з неточностями вимірів яскравості й сталістю порога, що використовується при отриманні бінарного зображення. Шум особливо значний поблизу границь областей

Кожний з описаних методів дозволяє зменшити кількість елементів зображення, які надалі будуть неправильно класифіковані в процесі бінарізації. Однак вони не гарантують отримання абсолютно надійних результатів. Випадковий шум може призвести до появи несподіваного рівня яскравості не тільки в окремому, але й у декількох сусідніх елементах. При цьому мала ймовірність виникнення значної помилки в одному елементі множиться на малі ймовірності значних помилок у сусідніх. Тому ймовірність того, що середнє значення або медіана містять помилку, порівнянну з помилкою окремого елемента, відносно мала [1].

Тут ми користуємося просторовим взаємозв'язком між елементами. У тих простих випадках, до яких ми зверталися дотепер, передбачалося, що яскравість об'єкта приблизно однакова, в результаті чого сусідні елементи мають близькі за значенням рівні. В таких умовах уводяться

обмеження, що впливають із припущення про те, що саме представлено на зображенні [1].

3.4 Сегментація зображень

Зображення бажано розбити на області, кожна з яких відповідала б поверхні окремого об'єкта сцени, що спостерігається. Наступну обробку можна здійснити незалежно над кожною областю зображення. Дотепер ми розглядали досить простий випадок, коли об'єкти й тло мають рівномірну яскравість. Тепер ми поширимо ці методи на випадки, в яких кожний об'єкт має яскравість, відмінну від яскравостей інших об'єктів [1].

Якщо нам відомий середній рівень яскравості кожного об'єкта, то неважко обрати граничні значення, що знаходяться між цими рівнями, це дозволить нам класифікувати будь-який елемент зображення [1]. Однак відразу ж виникають нові проблеми. Припустимо, що область A примикає до області C . Як було показано вище, рівні яскравості на їхній загальній границі будуть лежати між рівнями, що відносяться до цих областей. Якщо в іншому місці зображення розташована область B , рівень яскравості якої лежить між рівнями яскравості областей A і C , то частина точок границі буде класифікована неправильно як таких, що відносяться до B . Цього можна уникнути, якщо вдатися не до напівтонових, а до інших характеристик зображення [1]. Наприклад, точки, що нас цікавлять, швидше за все утворюють не широку область у вигляді плями, а порізану смугу із зазубреними краями. Якщо ми припустимо, що всі області мають деяку мінімальну ширину, то неважко замінити помилково привласнені мітки. Можна також використати той факт, що ці невірно класифіковані точки примикають до обох областей і A , і C [1].

Іноді необхідні граничні рівні визначають шляхом аналізу гістограми. Однак часто гістограма виявляється занадто порізаною, оскільки лише незначне число точок безпосередньо утворюють кожний

окремий сплеск i , крім того, ці сплески перекриваються значно сильніше, ніж у випадку одного об'єкта й однорідного тла [1].

Згадувані раніше проблеми шуму тут ще більш серйозні. Якщо є кілька областей, то мінімальна різниця між їхніми рівнями не може бути дуже великою. Вона повинна бути малою стосовно загального діапазону зміни рівнів. Якщо N - загальний діапазон, а R - число областей, то мінімальна різниця рівнів повинна бути меншою або дорівнювати $N/(R-1)$ [1]. У підсумку ймовірність неправильної класифікації помітно зростає. Дану проблему можна трохи пом'якшити, якщо врахувати просторовий взаємозв'язок об'єктів.

3.5 Використання кольору

Інший спосіб поліпшити класифікацію складається у використанні інформації про колір. Якщо навіть в результаті вимірів за допомогою єдиного оптичного пристрою виявиться, що поверхні мають однакову яскравість, вони все-таки можуть відрізнитися за кольором. Один зі шляхів використання цієї обставини складається у квантуванні (по рівнях яскравості) декількох зображень, отриманих за допомогою датчиків з різними функціями спектральної чутливості [1]. Цього можна досягти розміщенням перед об'єктивом кольорових світлофільтрів. Такі фільтри вибірково поглинають певні частини спектра й тим самим змінюють власну чутливість датчика до світла різних довжин хвиль. Зазвичай формуються три подібних зображення (через червоний, зелений і синій фільтри). При цьому детальне поводження кривих спектральної чутливості цих фільтрів значення не має [1].

Рівні яскравості трьох зареєстрованих зображень зручно об'єднати в один вектор, що приписується відповідному пікселю. Кожний компонент вектора являє собою рівень яскравості зображення в цьому пікселі,

отриманий за допомогою одного з фільтрів. Окремій області відповідає одна або множина точок у відповідному векторному просторі [1, 2].

Розбити колірний простір на частині, що відповідають різним областям, не просто. Однак: існують методи, що дозволяють здійснити це. Один із них полягає у вимірюванні відстані від представницького або середнього вектора, віднесеного до кожної області. Невідомий вектор можна потім приписати тому класу, який відповідає найближчому вектору-представнику. Це рівнозначно розбивці простору площинами, які перпендикулярні відрізкам між векторами-представниками й ділять ці відрізки навпіл [1].

3.6 Нарощування й розбивка областей

Допустимо, що тим або іншим методом ми здійснили детальну, але грубу сегментацію зображення. Наприклад, просто об'єднали в групи сусідні елементи з однаковими рівнями яскравості. Подальше поліпшення сегментації досягається злиттям сусідніх областей із близькими рівнями. «Близькість» може означати, що рівні відрізняються менше деякої граничної величини. При перевірці на близькість можуть також ураховуватися геометричні фактори, наприклад відношення довжини загальної границі до сумарної довжини границь двох областей [1].

Більш складна перевірка ґрунтується на оцінці середнього рівня яскравості і його варіацій всередині кожної області. За допомогою стандартних методів, запозичених зі статистики, неважко визначити, яка ймовірність того, що різні області випадково мають близькі рівні яскравості. Якщо ймовірність мала, то ми поєднуємо такі області [1]. Злиття - це монотонна операція, що обов'язково завершиться. Звичайно, при цьому можливий і такий результат, що все зображення буде ототожнено з однією областю!

Іноді подібний процес злиття приводить до об'єднання областей із близькими рівнями, які поєднувати не варто було б. Помилки такого роду можна виявити спираючись на геометричні міркування. Наприклад, якщо на границі є два яскраво виражених вирізи, що йдуть приблизно із протилежних сторін, то досить імовірно, що розмежувальну лінію необхідно провести між їхніми кінцями (рис. 3.5). Оцінка середнього й розкиду значень півтонів для виділених областей може підтвердити правильність цього припущення. Подібні методи застосовуються в основному для розбивки, а не нарощування областей. Зокрема, їх можна використати для виділення границь між областями з різними рівнями яскравості [1, 2]. Більш детально проблема виявлення границь буде розглянута пізніше.

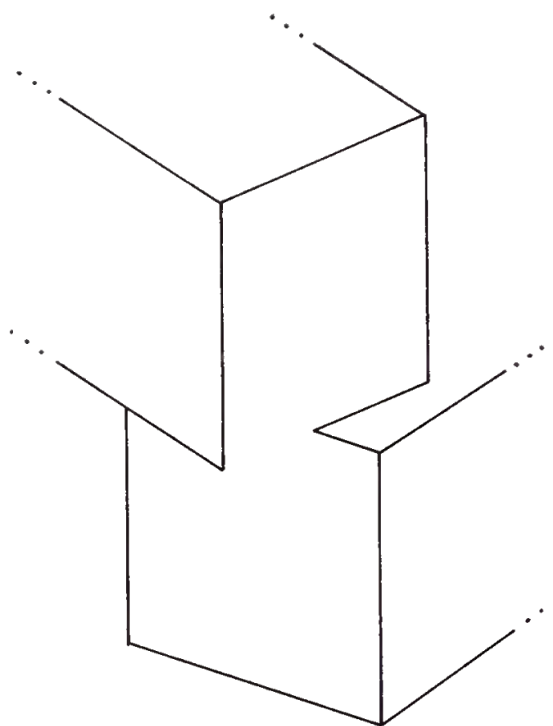


Рисунок 3.5 - Границі областей, які іноді можна визначити з контексту

Одна із проблем, пов'язаних з розглянутими методами сегментації, полягає в тому, що ті припущення, на яких вони ґрунтуються, у реальних

випадках часто не справжуються. Хоча відбивна здатність багатьох поверхонь і справді однакова, їхні зображення зовсім не зобов'язані зберігати цю властивість [1, 2]. На яскравість зображення впливає багато факторів. Тому елементи зображення, що відповідають точкам поверхні різних об'єктів, можуть мати однакові рівні яскравості. Тільки в особливих випадках може виявитися доцільним уявлення про зображення як про сукупність однорідних по яскравості областей [1].

Часто набагато плідотворніше відкласти сегментацію доти, поки не стане відома додаткова інформація, наприклад приблизна форма поверхні об'єкта. Одне з ускладнень, яке виникло при побудові оціночних схем для сегментації зображення, полягає у відсутності чіткої постановки самого завдання. Що таке «правильна» сегментація? Відповідь на це питання залежить від того, яким образом передбачається використати результат [1].

4 ВИКОРИСТАННЯ ОПЕРАТОРІВ ЗГОРТКИ ДЛЯ ВИДІЛЕННЯ ГРАНИЦЬ НА ЗОБРАЖЕННІ

Виділення границь знаходить широке застосування для вирішення наступних практичних задач [11]: розпізнаванні номерів автомобілів, OCR, медицині [3], картографії [4], кластеризації, знаходження тексту [11], фігур або складних об'єктів, наприклад, людей на зображенні [5, 6].

Мета пошуку границь у знаходженні границь (кривих) зміни яскравості або кольору [6]. В подальшому з границь будуються контури об'єктів, а з них – сцени. Вхідними даними для алгоритму є зображення $I_{x,y}$ кожен елемент якого містить три компоненти кольору – червону, зелену і синю, що позначаються $red(I_{x,y})$, $green(I_{x,y})$, $blue(I_{x,y})$. Результатом роботи алгоритму є матриця $R_{x,y}$, кожен елемент якої дорівнює 1, якщо піксель з координатами (x, y) є границею, або 0 в іншому випадку [2, 11].

Розв'язок цієї задачі повинен відповідати вимогам [7, 8] з яких окремо виділимо наступні [11]:

- алгоритм має знайти якомога більше границь присутніх на зображенні;
- знайдені границі мають бути максимально близькі до реальних границь;
- одна реальна границя не повинна створювати декілька паралельних границь на виході;
- шуми не повинні створювати неіснуючих границь.

4.1 Методи виділення границь на зображенні

Домінуючими методами пошуку границь є детектори Робертса [9], Собеля [3], Мар-Хільдрета [3], Кенні [6]. Завдяки тому, що зображення представляє собою матрицю пікселів, дані методи базуються на пошуку залежностей між пікселем і його сусідами.

Наведемо основні етапи базового алгоритму Кенні [11]:

1. Перетворення кольорового зображення у відтінки сірого. Кольорове зображення конвертується у відтінки сірого за формулою

$$I'_{x,y} = 0,212 \cdot \text{red}(I_{x,y}) + 0,715 \cdot \text{green}(I_{x,y}) + 0,072 \cdot \text{blue}(I_{x,y}). \quad (4.1)$$

Константи перед відповідними компонентами кольору $\text{red}(I_{x,y})$, $\text{green}(I_{x,y})$ і $\text{blue}(I_{x,y})$ отримані емпірично, з урахуванням фізіологічних особливостей сприйняття кольору людиною [10].

2. Видалення шумів. Алгоритм чутливий до шумів у зображенні тому, для видалення шуму зазвичай використовується фільтр Гауса [7]

$$S_{x,y} = I'_{x,y} * G_{x,y,\sigma}, \quad (4.2)$$

де $S_{x,y}$ – вихідний піксель; $I'_{x,y}$ – вхідний піксель; x, y – координати пікселя; $G_{x,y,\sigma} = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$ – фільтр Гаусса, пристосований для обробки двохмірних даних; σ – коефіцієнт розсіювання, символом $*$ позначена операція згортки [11].

Крім параметра σ фільтр Гаусса також визначається значенням параметра r , який вказує розмірність ядра. Наприклад, значенням $\sigma = 1,2$ та $r = 5$ відповідає наступна матриця фільтра Гауса [11]

$$G = \begin{bmatrix} 0,00825 & 0,0233 & 0,033 & 0,0233 & 0,00825 \\ 0,0233 & 0,0662 & 0,0937 & 0,0662 & 0,0233 \\ 0,033 & 0,0937 & 0,132 & 0,0937 & 0,033 \\ 0,0233 & 0,0662 & 0,0937 & 0,0662 & 0,0233 \\ 0,00825 & 0,0233 & 0,033 & 0,0233 & 0,00825 \end{bmatrix}. \quad (4.3)$$

Розмір ядра визначає відстань у пікселях, на яку буде проводитись розсіювання.

3. Пошук градієнту зміни яскравості. Градієнт

$$\nabla S_{x,y} = \left[\frac{\partial S_{x,y}}{\partial x}, \frac{\partial S_{x,y}}{\partial y} \right] \quad (4.3)$$

показує напрям і інтенсивність збільшення яскравості. Для кожного пікселя обчислюється величина градієнту $M_{x,y}$ і напрям $\theta_{x,y}$ [11]:

$$M_{x,y} = \sqrt{\left(\frac{\partial S_{x,y}}{\partial x}\right)^2 + \left(\frac{\partial S_{x,y}}{\partial y}\right)^2}; \quad \theta_{x,y} = \arctan\left(\frac{\partial S_{x,y}}{\partial y}\right) / \left(\frac{\partial S_{x,y}}{\partial x}\right). \quad (4.4)$$

Напрямок градієнту округлюється до одного з восьми значень, які визначаються за формулою $i \cdot \pi / 8$, $i = \overline{0, 7}$.

Для наближеного обчислення часткових похідних в формулі для градієнту (4.4) на практиці використовують спеціалізовані оператори згортання. А саме оператори Собеля, Превітта та Робертса. Охарактеризуємо зазначені оператори.

Оператор Собеля [11]:

$$\frac{\partial S_{x,y}}{\partial x} = S_{x,y} * \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}; \quad \frac{\partial S_{x,y}}{\partial y} = S_{x,y} * \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}. \quad (4.5)$$

Такий оператор є типовим для детектору Кенні. Він виконує згортку, беручи яскравість пікселів у напрямі градієнту двічі, тобто приділяючи їм більшої ваги.

Оператор Превітта [11]:

$$\frac{\partial S_{x,y}}{\partial x} = S_{x,y} * \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}; \quad \frac{\partial S_{x,y}}{\partial y} = S_{x,y} * \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}. \quad (4.6)$$

Оператор Превітта однаково враховує пікселі по напрямку градієнту і по діагоналям (горизонтальному, або вертикальному). На практиці оператор Превітта використовується для пошуку величини та орієнтації границі.

Оператор Робертса [11]:

$$\frac{\partial S_{x,y}}{\partial x} = S_{x,y} * \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}; \quad \frac{\partial S_{x,y}}{\partial y} = S_{x,y} * \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \quad (4.7)$$

На відміну від попередніх операторів, оператор Робертса обраховує градієнт у двох діагональних напрямках. Результатом застосування є градієнт від двох верхніх пікселів до двох нижніх. До переваг даного оператору можна віднести швидкість його роботи. Головним недоліком є чутливість до шумів.

4. Пригнічення не-максимумів. Для кожного пікселя, якщо величина його градієнта максимальна у напрямку градієнту, то він залишається без змін, інакше його величина приймається рівною нулю. Наприклад, якщо $\theta_{x,y} = \pi/4$ і $M_{x,y} < M_{x,y-1}$ то $M_{x,y}$ анулюється [11].

5. Відсіювання по граничним значенням. За двома параметрами Low та High анулюються градієнти, величина яких не входить до інтервалу [Low; High]. На завершення у результуючій матриці $R_{x,y}$ границями позначаються такі пікселі, у яких величина градієнту не нульова [11]

$$R_{x,y} = \begin{cases} 0, & \text{якщо } M_{x,y} = 0, \\ 1, & \text{якщо } M_{x,y} \neq 0. \end{cases} \quad (4.8)$$

Результатом роботи алгоритму є бінарна матриця, в якій пікселям, що належать границі об'єкту відповідає значення 1, а всім іншим – 0.

4.2 Приклад застосування оператора Собеля

Застосуємо оператор Собеля по відношенню до наступної матриці яскравості [11]:

$$I' = \begin{bmatrix} 123 & 43 & 67 & 107 & 150 \\ 228 & 80 & 158 & 172 & 167 \\ 43 & 243 & 31 & 154 & 97 \\ 23 & 193 & 237 & 20 & 8 \\ 243 & 100 & 178 & 189 & 120 \end{bmatrix}$$

Пропускаючи через фільтр Гауса будемо спостерігати таку матрицю $\|S_{x,y}\|$

$$S = \begin{bmatrix} 60,154 & 75,081 & 84,027 & 85,440 & 69,798 \\ 86,788 & 115,352 & 126,203 & 119,533 & 91,339 \\ 97,694 & 138,308 & 149,009 & 130,537 & 91,705 \\ 94,007 & 136,451 & 145,900 & 121,515 & 79,778 \\ 75,110 & 105,661 & 112,830 & 93,993 & 61,373 \end{bmatrix}.$$

Використання оператора Собеля призводить до наступного результату

$$\left\| \frac{\partial S_{x,y}}{\partial x} \right\| = \begin{bmatrix} 207,095 & 265,514 & 87,161 & 24,901 & -63,322 \\ 331,424 & 444,094 & 154,019 & 10,95 & -141,262 \\ 376,183 & 528,419 & 193,939 & -26,298 & -215,595 \\ 360,819 & 516,97 & 192,82 & -49,311 & -241,005 \\ 244,228 & 347,772 & 127,331 & -38,271 & -169,036 \end{bmatrix};$$

$$\left\| \frac{\partial S_{x,y}}{\partial y} \right\| = \begin{bmatrix} 86,788 & 288,928 & 443,696 & 487,293 & 456,609 \\ 37,54 & 138,308 & 228,978 & 238,289 & 177,083 \\ 7,219 & 35,537 & 69,112 & 62,473 & 12,099 \\ -22,584 & -77,815 & -124,058 & -141,55 & -139,599 \\ -94,007 & -324,465 & -512,809 & -549,765 & -468,708 \end{bmatrix};$$

$$M = \begin{bmatrix} 224,545 & 392,399 & 452,176 & 487,928 & 460,979 \\ 333,543 & 465,132 & 275,958 & 238,541 & 226,524 \\ 376,252 & 529,613 & 205,885 & 67,782 & 215,934 \\ 361,525 & 522,695 & 229,281 & 149,893 & 278,516 \\ 261,696 & 475,629 & 528,38 & 551,096 & 498,257 \end{bmatrix};$$

$$\theta = \begin{bmatrix} \pi/2 & \pi/4 & 0 & 0 & 0 \\ \pi/2 & \pi/2 & \pi/4 & 0 & 7\pi/4 \\ \pi/2 & \pi/2 & \pi/2 & 0 & 3\pi/2 \\ \pi/2 & \pi/2 & 3\pi/4 & \pi & 5\pi/4 \\ 3\pi/4 & 3\pi/4 & \pi & \pi & \pi \end{bmatrix}.$$

5 СЕГМЕНТАЦІЯ СТРУКТУРОВАНИХ ДОКУМЕНТІВ

В цьому розділі розглядається низхідний підхід до сегментації та представлення документів, що містять табличні структури. Прикладами таких документів є рахунки-фактури та технічні документи з таблицями. Сегментація базується на залученні розширення апарату X-Y дерев. В якості роздільників (розрізів) між областями виступають або лінії, або зони утворені пробілами. Листя X-Y дерев описують області, що містять однорідну інформацію, та роздільники розрізів. Зв'язки суміжності між листям дерева описують локальні зв'язки між відповідними областями [13].

5.1 Загальні відомості

Сегментація документів є базовим завданням в системах зберігання та пошуку, і її виконують методами «знизу вгору» або «зверху вниз» [14]. По суті, при аналізі «знизу вгору» зв'язані компоненти витягуються з зображення та згодом агрегуються у структури вищого рівня, створюючи слова, рядки тексту, абзаци тощо (наприклад, RLSA: Run Length Smoothing Algorithm, алгоритм згладжування довжин серій). При аналізі «зверху вниз» сторінка сегментується від більших компонентів до менших підкомпонентів. Наприклад, при декомпозиції дерева X-Y сторінка рекурсивно розділяється на підчастини шляхом чергування горизонтальних та вертикальних розрізів вздовж проміжків [12, 13, 15].

Методи «зверху вниз» зазвичай швидші та мають перевагу безпосереднього вилучення геометричної ієрархії між регіонами, яка часто тісно пов'язана з логічною структурою документа. Методи «знизу вгору» здаються більш доцільними для макетів «не манхетенського» типу (наприклад, сторінок, де рисунки перемішані із текстом таким чином, що

вони або є окремою частиною документа, або текст розташований з усіх боків, огортаючи рисунки навколо) [12, 13].

В даному розділі основна увага приділяється методу аналізу компоновки документів, що містять табличні структури. Такими документами, зокрема, є рахунки-фактури [13, 16]. Підхід, що ґрунтується на використанні X-Y дерев видається доцільним для сегментації рахунків-фактур, які зазвичай виявляють регулярну компоновку. З іншого боку, декомпозиція на базі X-Y дерев не може відбуватися в тих областях, що утворюються всередині таблиць через наявність ліній. Одним із шляхів вирішення цієї проблеми є модифікація способу розрізання в X-Y підході, коли вводиться більш загальне поняття роздільників, а також враховуються розрізи, що відповідають горизонтальним і вертикальним лініям [13].

Другою властивістю рахунків-фактур, яку необхідно враховувати, є те, що логічні зв'язки між об'єктами здебільшого локальні. Наприклад, поле, що містить загальну суму, зазвичай друкується поруч із ключовим словом «Всього». Для ефективного опису зв'язків суміжності між об'єктами додаються зв'язки суміжності між парами листків дерева [13].

В подальшому описано підхід до сегментації на базі X-Y дерев, аналізується використання модифікованих X-Y дерев (M-X-Y дерев) для представлення макета (компоновки) сторінки [13]. Насамкінець наводяться деякі приклади сегментації, отриманої за допомогою розглянутого підходу.

5.2 Сегментація на базі X-Y дерев

X-Y дерево [15, 17] – це добре відомий низхідний метод аналізу компоновки сторінки. Основним припущенням, що лежить в основі цього підходу, є те, що структурні елементи сторінки зазвичай розташовуються в прямокутних областях, тобто утворюють прямокутні блоки. Крім того, блоки майже завжди можна розділити на групи таким чином, щоб блоки,

що прилягають один до одного в межах групи, мали один спільний вимір (ширина/висота) [15]. Документ послідовно розділяється на менші прямокутні блоки шляхом почергового горизонтального та вертикального «розрізів» вздовж областей, утворених пробілами. Пробіли знаходяться гістограмним методом (враховується кількість чорних пікселів вздовж паралельних ліній, що проходять крізь документ). Результат такої сегментації можна представити у вигляді X-Y дерева, де корінь відповідає всій сторінці, листя – блокам сторінки, тоді як кожен рівень по черзі представляє результати горизонтальної або вертикальної сегментації. Порожній (або пустий) розріз [15] відповідає вузлу з унікальним дочірнім елементом, який потрібен, коли область необхідно розрізати на послідовних рівнях в одному напрямку (наприклад, колонка тексту може бути спочатку розрізаний горизонтально на абзаци, а потім розрізана у тому ж напрямку, щоб розбити абзаци на рядки) [13].

Для визначення належних елементів розрізання потрібні деякі порогові значення, що зменшує ризик «поганих» розрізів (наприклад, розрізів вздовж ліній, що відповідають частинам символів). В розглянутій реалізації всі ці порогові значення розраховуються на основі середньої висоти та ширини символів на сторінці, що обробляється [13].

Визначення 1. *Розрізаючий простір*. Прямокутна область \bar{R} є горизонтальним розрізаючим простором для області R якщо виконуються такі умови (рис. 5.1) [13]:

- 1) $\bar{R}.width = R.width$ (\bar{R} і R мають однакову ширину);
- 2) Область R є білою. Щоб толерувати невелику кількість шуму, область R вважається білою, якщо відсоток чорних пікселів у ній є меншим за невелике порогове значення (наприклад, 2%);
- 3) $\bar{R}.height > T_h$. Встановлюючи $T_h = C_h$, ми дозволяємо сегментацію тексту на окремі рядки. Якщо $T_h = 2 \cdot C_h$, то ми зупиняємо процес на рівні абзацу.

Вертикальні проміжки розрізання можна визначити аналогічним чином.

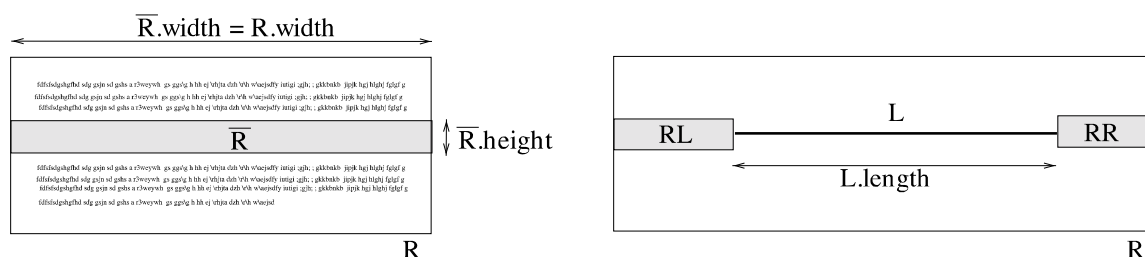


Рисунок 5.1 – Ліворуч показано розрізаючий простір; праворуч: області RL та RR для ліній розрізу

Визначення 2. Лінія розрізу. L є горизонтальною лінією розрізу для області, R якщо виконуються такі умови [13]:

$$1) \frac{L.length}{R.length} > T_L, \text{ де } L.length - \text{довжина лінії, а } T_L - \text{поріг, що}$$

дозволяє різати лише вздовж «довгих» ліній (зазвичай $0,25 \leq T_L \leq 0,5$);

2) два прямокутники RL та RR (див. рис.5.1) є білими, де RL (RR) – це прямокутник висоти $2 \cdot C_h$ що охоплює площу від лівого (правого) краю L до лівого (правого) краю R ;

3) $L.length > 2 \cdot C_w$. За цієї умови ми уникаємо використання частин символів (наприклад, горизонтальної частини літери «L») як ліній розрізу;

4) $L.thickness < C_h$. Занадто товсті лінії не враховуються, щоб уникнути розщеплення деталями, що входять до складу логотипів або графічних об'єктів.

Вертикальні лінії розрізу можна визначити аналогічним чином. Приклад ліній розрізу наведено на рисунку 5.2 [13].

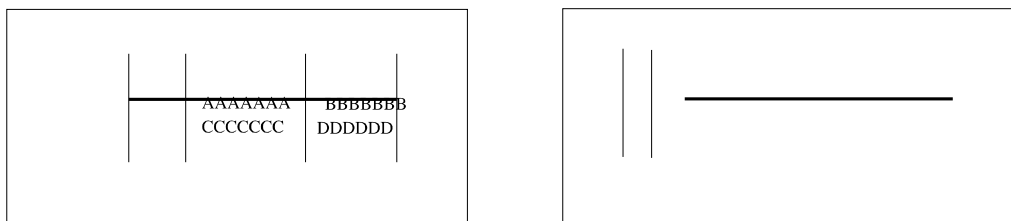


Рисунок 5.2- Приклад лінії розрізу (ліворуч) та лінії, яку не можна використовувати як лінію розрізу (праворуч)

У розглянутому підході [13] сторінка рекурсивно сегментується за допомогою роздільників, введених у визначенні 1 та визначенні 2. В кожній під регіоні (області) алгоритм спочатку шукає лінії розрізу; розрізи виконуються відповідно до всіх ліній розрізу таких, що $L.length > T_{RL} \cdot L_{max}$ (L_{max} - довжина найдовшої лінії розрізу в підрегіоні, а $T_{RL} = 0,7$). Якщо лінії розрізу не знайдено, то шукаються прогалини розрізу, і розрізи виконуються відповідно до всіх прогалин розрізу таких, що $A_S > T_{RA} \cdot A_{max}$ (A_S - площа простору розрізу S , A_{max} - площа найбільшого простору розрізу, а $T_{RA} = 0,7$) [13].

Щойно описаний алгоритм є, по суті, керованим даними, оскільки стратегія розділення використовує мало високорівневої інформації про клас документів. Цей підхід забезпечує гіршу сегментацію, ніж модельно-керований підхід, налаштований на конкретну проблему, проте він дозволяє нам отримати хороші результати для широкого кола документів без зміни підходу [13].

Слід зазначити, що деякі документи містять інші види роздільників. Наприклад, існують документи, де різні елементи розділені різкими розривами кольорів фону або пунктирними лініями. В цих випадках можливості сегментації можна додатково розширити, додавши ці роздільники до набору елементів розрізання [13, 12].

5.3 Модифікація X-Y дерев

У цьому розділі дано опис M-X-Y дерев як структури даних для представлення документів, сегментованих за допомогою алгоритму, викладеного в підрозділі 5.2. Основними особливостями M-X-Y дерева є зберігання ліній розрізу як листків дерева та опис відносин суміжності між листками за допомогою відповідних зв'язків [13].

Визначення 3. *Модифіковане X-Y дерево* $MXY = \{N, A, A_N, A_A\}$ - це дерево з вузлами N , дугами A , сполучними листками, атрибутами вузлів у A_N , атрибутами дуг у A_A [13].

Приклад M-X-Y дерева показано на рис. 5.3, атрибути вузла та дуги визначені наступним чином.

Визначення 4. *Атрибути вузла* – це кортежі з 3 елементів: $A_{N_i} = \{T_i, Pos_i, Child_i\}$, де T_i - тип вузла, Pos_i вказує положення прямокутної області, що відповідає вузлу N_i (що описується його верхнім лівим (x_{l_i}, y_{l_i}) та нижнім правим (x_{r_i}, y_{r_i}) кутами), $Child_i$ – список дочірніх елементів N_i [13].

Якщо N_i - внутрішній вузол, то T_i визначає спосіб розрізання, який генерує його дочірні вузли (HL, VL для горизонтальних та вертикальних ліній, HS, VS для пробілів (незаповнених областей) та V для порожнього розрізу). Якщо N_i - листок (дерева), то T_i визначає тип області, що відповідає вузлу: $T_i = T$ для тексту, $T_i = I$ для зображень та $T_i = L$ для роздільника рядків (зображеного як тонка лінія на рис. 5.3) [13].

Зв'язки суміжності між листям дерева можна розглядати як граф суміжності, де вузли графа відповідають листям дерева. Граф суміжності (див., наприклад, [18]) описує структуру документа, надаючи для кожної області положення найближчих областей у горизонтальному та вертикальному напрямках.

Визначення 5. *Атрибути дуги* (для дуги, що з'єднує N_i з N_j) дорівнюють $A_{A_{i,j}} = \{P_{i,j}, d_{i,j}\}$. Де $P_{i,j} = B$ якщо N_j , розташовано Нижче N_i (див. визначення 6); $P_{i,j} = R$ якщо N_j , розташовано Справа від N_i ; $d_{i,j}$ це відстань між найближчими межами (границями) прямокутників, що відповідають з'єднаним вузлам [13].

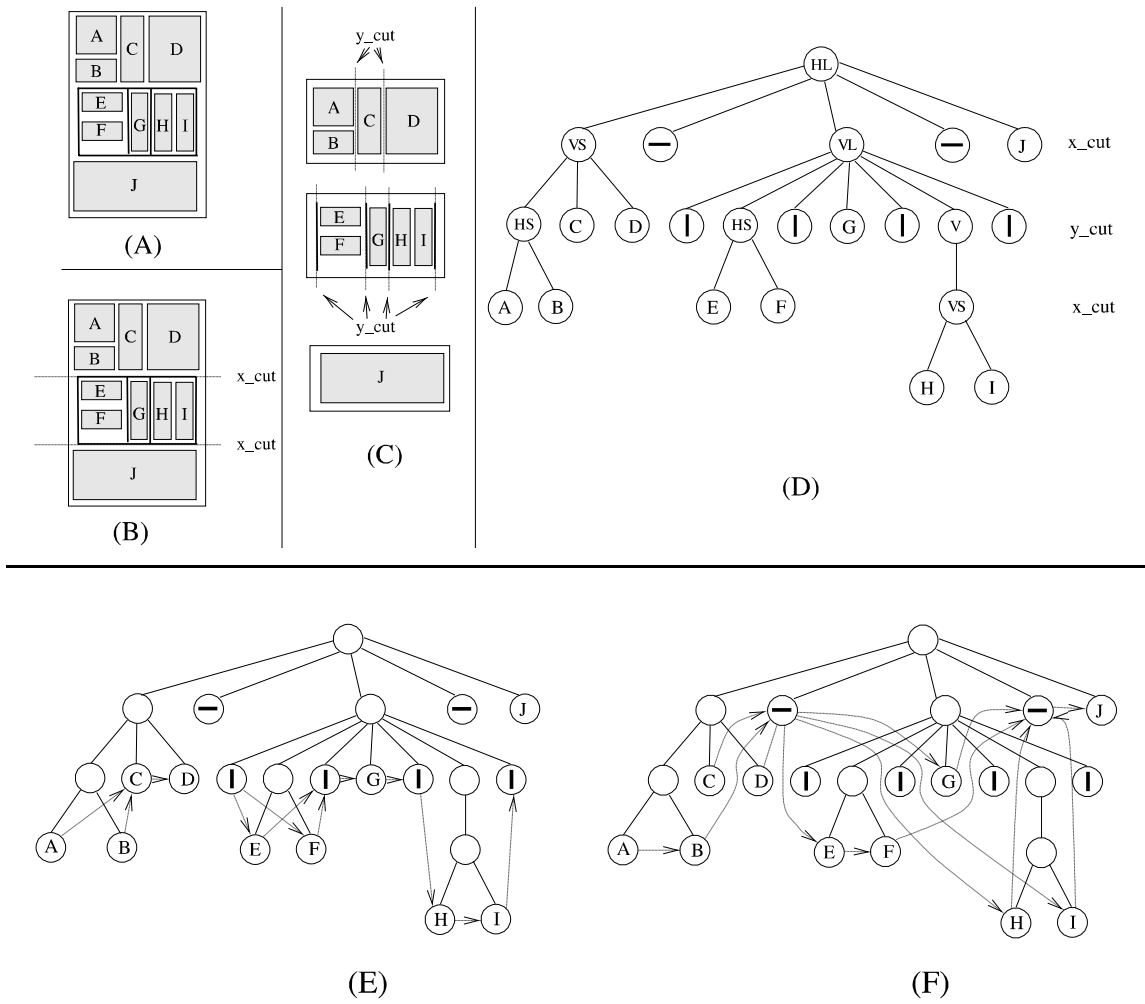


Рисунок 5.3 - Приклад дерева М-Х-У: (А) - зразок сторінки; (В) та (С) - перший та другий рівні сегментації; (D) – М-Х-У дерево; (E), (F) – дуги, що відповідають розташуванню «Праворуч від» та «Нижче від»

Визначення 6. Для двох областей A та B , описаних вузлами N_i та N_j , ми кажемо, що B розташовано Нижче від A якщо виконано такі три умови [13]:

- 1) $yu_i < yu_j$;
- 2) $[xu_i, xb_i] \cap [xu_j, xb_j] \neq \emptyset$;
- 3) $\exists x \in [xu_i, xb_i] \cap [xu_j, xb_j]$ такі, що відрізок, що з'єднує точки (x, yb_i) та (x, yu_i) не перетинає жодної іншої області.

Відношення Справа можна визначити аналогічно.

5.4 Результати експерименту

В цьому розділі ми аналізуємо приклади сегментації, отриманої за допомогою розглянутого підходу [13], для двох класів документів: рахунків-фактур та технічних документів.

Приклад сегментації частини рахунку-фактури, отриманої за допомогою M-X-Y дерева та методу «знизу вгору» (RLSA: Run Length Smoothing Algorithm, алгоритм згладжування довжин серій), показано на рис. 5.4. Реалізація RLSA базується на двох кроках: по-перше, горизонтальні та вертикальні незаповнені/білі області (white runs) з довжиною меншою за відповідні порогові значення (RL_h та RL_v) заповнюються чорними пікселями. По-друге, області знаходяться шляхом визначення місцезнаходження зв'язаних компонентів на заповненому зображенні (зображенні, отриманому після виконання першого етапу) [13].

Використовуючи RLSA для рахунків-фактур, що містять багато рядків, перший етап алгоритму призводить до об'єднання деяких текстових частин разом з рядками, таким чином виявляючи лише кілька областей. Отже, підхід на основі RLSA можна використовувати лише після видалення рядків із зображення, як показано на рисунку 5.4 (зверху).

[REDACTED] SPA		2013	
VIA BERGAMO 21		2013	
COD. CLIENTE	N. VIAGGI FATTURATI	22	FATTURA EMESSA
COORDINATE BANCARIE			PAGAME BANC.
ABI	CAB	C/C	
COD. IVA	DESCRIZIONE	RIF. DT	LEGG.
120	PEDAGGI AUTOSTRADALI	IVA	20%
120	CANONE TELEPASS	IVA	20%
10	QUOTA ASS. I CARTA	ART.	10/1
10	QUOTA ASS. CARTE SUC.	ART.	10/1

(A)

[REDACTED] SPA		2013	
VIA BERGAMO 21		2013	
COD. CLIENTE	N. VIAGGI FATTURATI	22	FATTURA EMESSA
COORDINATE BANCARIE			PAGAME BANC.
ABI	CAB	C/C	
COD. IVA	DESCRIZIONE	RIF. DT	LEGG.
120	PEDAGGI AUTOSTRADALI	IVA	20%
120	CANONE TELEPASS	IVA	20%
10	QUOTA ASS. I CARTA	ART.	10/1
10	QUOTA ASS. CARTE SUC.	ART.	10/1

(B)

Рисунок 5.4 – Результати сегментації: зверху - сегментація з використанням підходу на основі RLSA; знизу - сегментація з використанням - M-X-Y дерева

Після деяких тестів було помічено, що унікальне значення для RL_v неможливо обрати для всього документа, оскільки малі значення розділяють всі текстові рядки, тоді як великі значення об'єднують рядки з різним значенням. На рис. 5.4 (зверху) бачимо подібну ситуацію: два слова з різним значенням («BERGAMO» та «CLIENTE») об'єднані, оскільки вертикальна відстань між ними (після видалення рядків з компоновки) менша, ніж відстань між рядками одного й того ж поля (наприклад, «BERGAMO» та затерте слово, яке є назвою компанії). Рис. 5.4 (знизу)

містить сегментацію, отриману з залученням М-Х-У дерев. В цьому випадку ми маємо сегментацію, більш схожу на «людське сприйняття» документа [13].

Приклад сегментації технічного документа показано на рисунку 5.5: (А) містить оригінальне зображення; (В) – сегментація, отримана за допомогою М-Х-У дерева [13]. Області виглядають однорідними. Наприклад, блоки в текстовій частині описують абзаци тексту, тоді як блоки всередині таблиці розділені. Більше того, рисунок у верхній лівій частині сторінки є унікальною областю; (С) показує сегментацію, отриману шляхом розрізання лише вздовж пробілів (незаповнених областей), враховуючи низький поріг для локалізації цих областей (поріг було визначено гістограмним методом). Більшість областей такі ж, як і в першій частині, але таблиця розглядається як єдина область. Якщо ми бажаємо відокремити області для кожного блоку всередині таблиці, нам потрібно збільшити поріг, щоб робити розрізи також відповідно до рядків таблиці (рис. 5.5 (D)). У цьому випадку області всередині таблиці добре сегментовані, але в абзацах тексту трапляються деякі дивні сегментації, і малюнок розділений на менші підблоки.

Слід зауважити два основних обмеження алгоритму сегментації з використанням М-Х-У дерев. По-перше, як і інші методи, що будуються на основі Х-У дерев, алгоритм вимагає добре вирівняного зображення. По-друге, документ можна розділити (сегментувати) лише вздовж роздільників, що покривають всю ширину (або висоту) зображення, що відповідає вузлу, який потрібно розділити. Отже, таблиці, вкладені в текст, не можна сегментувати [13].

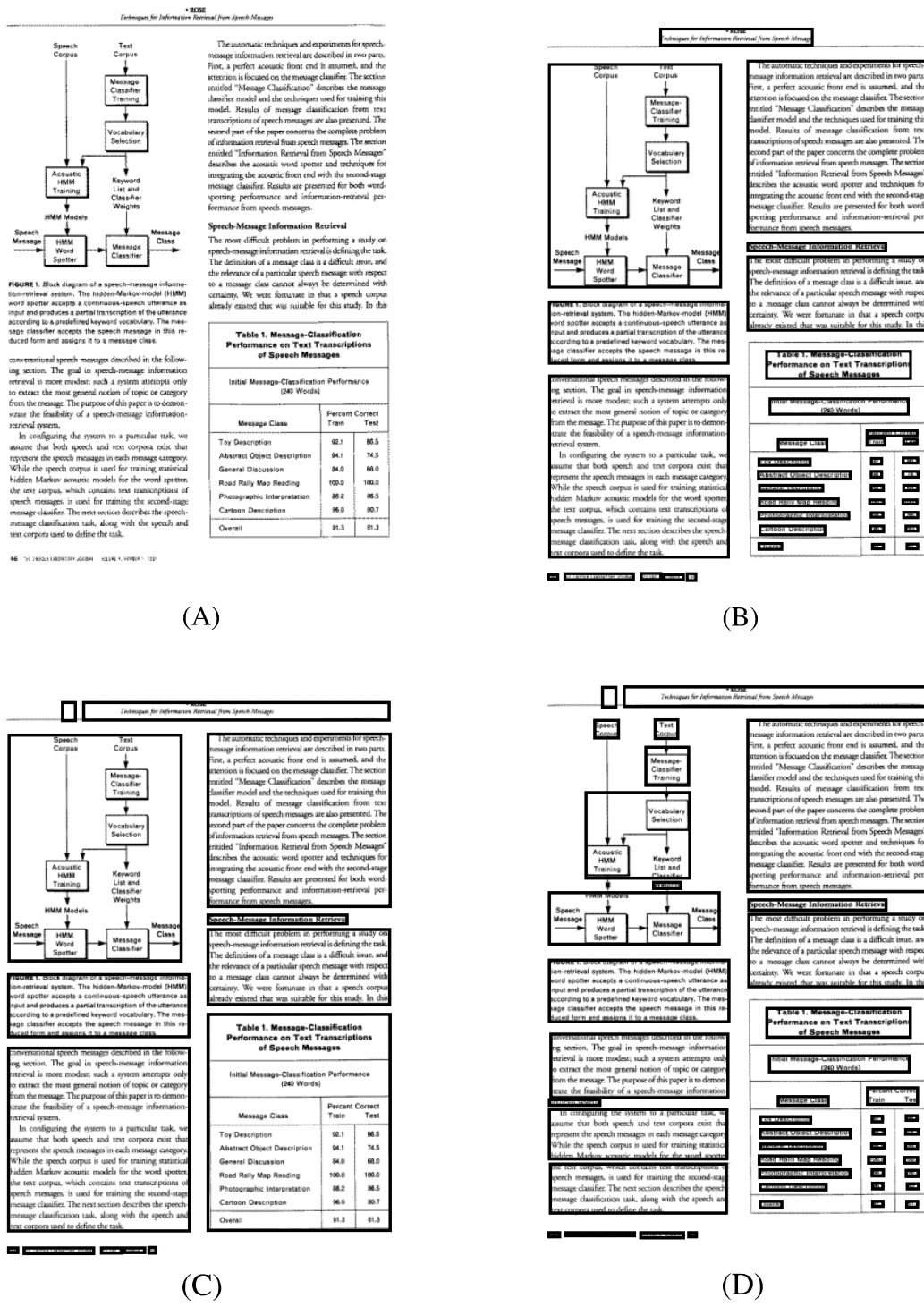


Рисунок 5.5 - Порівняння сегментацій сторінки технічного документа (див. текст для отримання детальнішої інформації)

ВИСНОВКИ

Сегментація документів є базовим завданням в системах зберігання та пошуку, і її виконують методами «знизу вгору» або «зверху вниз». По суті, при аналізі «знизу вгору» зв'язані компоненти витягуються з зображення та згодом агрегуються у структури вищого рівня, створюючи слова, рядки тексту, абзаци тощо. При аналізі «зверху вниз» сторінка сегментується від більших компонентів до менших підкомпонентів. Наприклад, при декомпозиції X-Y дерева сторінка рекурсивно розділяється на підчастини шляхом чергування горизонтальних та вертикальних розрізів вздовж проміжків.

Методи «зверху вниз» зазвичай швидші та мають перевагу безпосереднього вилучення геометричної ієрархії між регіонами, яка часто тісно пов'язана з логічною структурою документа. Методи «знизу вгору» здаються більш доцільними для компоновок «не манхеттенського» типу (наприклад, сторінок, де рисунки перемішані із текстом таким чином, що вони або є окремою частиною документа, або текст розташований з усіх боків, огортаючи рисунки навколо).

Було описано низхідний підхід до сегментації та представлення табличних документів, зокрема рахунків-фактур, на основі даних. Можна виділити дві основні переваги методу сегментації документів на базі M-X-Y дерев: 1) інтеграція більшої кількості стратегій розрізання (розділення) в рамках «класичної» сегментації за допомогою X-Y дерев; 2) врахування структури даних (модифіковане X-Y дерево) для представлення сегментованого документа.

ПЕРЕЛІК ПОСИЛАНЬ

1. Horn, B. (1986). Robot vision. MIT press.
2. Forsyth, D. A., & Ponce, J. (2002). Computer vision: a modern approach. prentice hall professional technical reference.
3. Li, J., & Ding, S. (2011, August). A research on improved Canny edge detection algorithm. In International Conference on Applied Informatics and Communication (pp. 102-108). Berlin, Heidelberg: Springer Berlin Heidelberg.
4. Harris, C., & Stephens, M. (1988, August). A combined corner and edge detector. In Alvey vision conference (Vol. 15, No. 50, pp. 10-5244).
5. Hines, E. L., & Watson, P. A. (1983). Application of edge detection techniques to detection of the bright band in radar data. Image and Vision Computing, 1(4), 221-226.
6. Canny, J. (1986). A computational approach to edge detection. IEEE Transactions on pattern analysis and machine intelligence, (6), 679-698.
7. Panetta, K. A., Wharton, E. J., & Agaian, S. S. (2008). Logarithmic Edge Detection with Applications. J. Comput., 3(9), 11-19.
8. Kaehler, A., & Bradski, G. (2016). Learning OpenCV 3: computer vision in C++ with the OpenCV library. O'Reilly Media, Inc.
9. Steger, C., Ulrich, M., & Wiedemann, C. (2018). Machine vision algorithms and applications. John Wiley & Sons.
10. Johnson, S. (2006). Stephen Johnson on digital photography. O'Reilly Media, Inc.
11. Петров, С. О., Марченко, І. О., & Дібров, Б. О. (2015). Дослідження застосування операторів згортки в задачах виділення границь на зображенні. Восточно-Европейский журнал передовых технологий, 6(4 (78)), 27-31.

12. Safonov, I.V., Kurilin, I., Rychagov, M.N., & Tolstaya, E.V. (2019). Segmentation of Scanned Images of Newspapers and Magazines. *Signals and Communication Technology*, 107-122.

13. Cesarini, F., Gori, M., Marinai, S., & Soda, G. (1999, September). Structured document segmentation and representation by the modified XY tree. In *Proceedings of the Fifth International Conference on Document Analysis and Recognition. ICDAR'99* (Cat. No. PR00318) (pp. 563-566). IEEE.

14. Jain, A. K., & Yu, B. (1998). Document representation and its application to page decomposition. *IEEE Transactions on pattern analysis and machine intelligence*, 20(3), 294-308.

15. Nagy, G., & Seth, S. C. (1984). Hierarchical representation of optically scanned documents.

16. Cesarini, F., Gori, M., Marinai, S., & Soda, G. (2002). INFORMys: A flexible invoice-like form-reader system. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(7), 730-745.

17. Viswanathan, M. (1992). Analysis of scanned documents—A syntactic approach. In *Structured Document Image Analysis* (pp. 115-136). Berlin, Heidelberg: Springer Berlin Heidelberg.

18. Yuan, J., Tang, Y. Y., & Suen, C. Y. (1995, August). Four directional adjacency graphs (FDAG) and their application in locating fields in forms. In *Proceedings of 3rd International Conference on Document Analysis and Recognition* (Vol. 2, pp. 752-755). IEEE.