

Верифікація людини по зображенню її обличчя з можливістю використання в карткових системах контролю доступу до приміщень

Кийко В.М.

Міжнародний науково-навчальний центр інформаційних технологій і систем
НАН та МОН України
kiyko@image.kiev.ua

Abstract

The following face recognition algorithms are considered and tested: 1) face parts template matching algorithm, 2) DCT-based algorithm and 3) algorithm for matching correspondent points using dynamic programming. Experimental results show that DCT-based algorithm performs better than other two algorithms on the same database. This is due mainly to its relatively more high robustness to illumination direction changes, as well as to out of plane rotations of images.

1. Вступ

Задача розробки нових та підвищення надійності існуючих систем контролю доступу до приміщень або інформації є достатньо важливою та актуальною. Ефективно вирішувати цю задачу можна головним чином шляхом введення в систему засобів, які дозволяють розпізнавати людину по її біометричним ознакам. Серед цих ознак зображення обличчя є найбільш зручним та прийнятним для людини у порівнянні, наприклад, з використанням відбитків пальців, які традиційно сприймаються з певною мірою застереження. Процес верифікації по зображенню обличчя в загалом полягає у наступному. Особа, що претендує на доступ, розміщується перед відеокамерою і натискає кнопку пристрою, на якому установлена ця відеокамера. Після цього утворюється зображення обличчя людини, яке у подальшому оброблюється і порівнюється з одним або декількома еталонами. Якщо певним чином обчислена відстань між вхідним зображенням та одним із еталонів є достатньо малою, то приймається рішення про дозвіл людини до приміщення.

Використання верифікації в існуючих карткових системах доступу практично є можливим, коли процес розпізнавання (верифікації) людини по зображенню її обличчя є, по-перше, достатньо швидким та надійним (час розпізнавання : не більше 2-х сек., а надійність не менше 90%), а по-друге, для своєї реалізації потребує порівняно малий об'єм даних (2 – 3 KByte) про еталонне зображення обличчя цієї людини, що повинні бути записані на картку. Важливим є також те, що верифікація повинна виконуватись в умовах змінної освітленості обличчя, а також неминучих певних змін орієнтації обличчя людини, що розпізнається. При цьому фронтальне додаткове освітлення обличчя за допомогою

встановленого на розробленому пристрої освітлювача [1] тільки частково допомагає вирішити проблему зміни освітлення облич людей протягом доби. При таких умовах надійність верифікації можна суттєво збільшити, якщо аналізувати не одне зображення людини перед камерою, а певну відео-послідовність цих зображень, тому що чим з більшої кількості зображень складається ця послідовність, тим більшою є вірогідність того, що вона містить зображення обличчя, яке по орієнтації та умовам освітлення достатньо добре відповідає еталону обличчя людини, що записаний на магнітній картці.

Попередня обробка зображень відео-послідовності виконується за допомогою методів, описаних в [1-4], і складається з таких кроків:

1. Пошук облич на вхідних зображеннях.
2. Пошук центрів зіниць на вхідних зображеннях.
3. Нормалізація зображень облич по розміру та повороту: після нормалізації зображення мають однакові розміри, а також однакові координати центрів зіниць очей.
4. Нормалізація яскравості зображень облич за дисперсією, середнім значенням та за локальним контрастом.

На послідуєчому заключному етапі отримані нормалізовані вхідні зображення облич порівнюються з еталонами. При цьому, якщо обчислена відстань між одним із нормалізованих вхідних зображень обличчя і еталонном є меншою ніж задане значення, то приймається рішення про успішну верифікацію людини перед камерою.

У подальшому розглядаються та порівнюються три алгоритми порівняння зображень. Перший з цих алгоритмів умовно називається "GetSimilarity" і загалом полягає у наступному. Нормалізоване еталонне зображення обличчя спочатку розділяється по вертикалі на три складові частини (очі, ніс та рот), що мають однакову ширину. Потім виконується (шляхом певних зсувів і обчислення кореляції) пошук найкращих положень цих 3-х складових частин на вхідному зображенні, а також обчислення зваженої суми відстаней між цими частинами та вхідним зображенням. Більш детально цей алгоритм розглядається у роботі [2].

Інші два алгоритми ("GetSimilarity_DCT" та "StereoMatch") розглядаються нижче у розділах 2 та 3. Додаткова особливість алгоритму "StereoMatch" полягає у тому, що яскравість вхідних та еталонних зображень не нормалізується за локальним контрастом. Результати експериментальної перевірки алгоритмів та їх аналіз приведені у розділі 4, а головні висновки – у розділі 5.

2. Алгоритм обчислення відстані між двома зображеннями на основі дискретного косинусного перетворення (DCT) цих зображень

Алгоритм складається з наступних 4-х етапів. На першому етапі кожне з двох порівнюваних зображень представляється у вигляді сукупності перекритих між собою блоків 8x8 кліток. Сусідні блоки перекриваються між собою на 50%, тобто на 4 клітки або по вертикалі, або по горизонталі, або ж одночасно по вертикалі та горизонталі. Якщо, наприклад, порівнювані зображення мають розміри 104x104 клітки, то загальна кількість блоків на одному з цих зображень дорівнює $759 = 23 \times 23$.

На другому етапі виконується пряме дискретне косинусне перетворення [5] кожного блока на порівнюваних зображеннях по такій формулі:

$$F_{x,y} = \frac{C(x)C(y)}{4} \sum_{i=0}^7 \sum_{j=0}^7 f_{i,j} \cos\left(\frac{(2i+1)x\pi}{16}\right) \cos\left(\frac{(2j+1)y\pi}{16}\right)$$

де $f_{i,j}$ є яскравості 64-х кліток (i, j) блока, $F_{x,y}$ є 64-и DCT коефіцієнти з індексами (x, y) і $C(x)$, $C(y)$ є константами:

$$C(n) = \begin{cases} \frac{1}{\sqrt{2}} & n = 0, \\ 1, & n \neq 0. \end{cases}$$

Загальна схема 2-D дискретного косинусного перетворення та позначення $f_{i,j}$ і $F_{x,y}$ показана нижче на рис. 1.

Пряме DCT (FDCT) розкладає 2D зображення блока по ортонормованим 2D DCT базисним функціям і в результаті трансформує це зображення у 2D множину DCT коефіцієнтів такого розкладу. Перший з цих коефіцієнтів F_{00} є усереднена яскравість кліток блоку, два інших (F_{01}, F_{10}) є усередненими значеннями градієнту кліток блоку відповідно по горизонталі та вертикалі. Ці три коефіцієнти є найбільш залежними від умов освітлення зображення у порівнянні з іншими коефіцієнтами, серед яких найбільш впливових на результати розпізнавання є напрям освітлення. Пряме DCT є зворотнім (зворотне DCT досить якісно відновлює зображення блока, використовуючи порівняно малу кількість DCT коефіцієнтів), а також має такі дві головні корисні властивості: 1) компактність енергії (концентрація більшої частини енергії зображення у порівняно невеликій кількості DCT коефіцієнтів в околі F_{00}) та 2) декореляція (мінімізація залежності DCT коефіцієнтів між собою). Загалом розміри блоку, а також рівень їх перекриття між собою можуть бути іншими, але найбільш поширені розміри 8x8 є вдалим компромісом між кількістю операцій, що потребує обчислення DCT коефіцієнтів, яка зростає при збільшенні цих розмірів, та інформативністю зображення в межах кожного блока. Збільшення

перекриття блоків, з одного боку, також приводить к збільшенню кількості операцій у квадратичній залежності, але, з другого - суттєво зменшує залежність результатів розпізнавання від змінень орієнтації та помилок нормалізації облич, зумовлених, наприклад не досить точним визначенням положення очей на цих зображеннях

Зображення блока

	0	1	2	3	4	5	6	7
0	f_{00}	f_{01}	f_{02}	f_{03}	f_{04}	f_{05}	f_{06}	f_{07}
1	f_{10}	f_{11}	f_{12}	...				
2	f_{20}	f_{21}	...					
3	f_{30}	...						
4	...							
5								
6								
7								

FDCT
⇒

DCT коефіцієнти

	0	1	2	3	4	5	6	7
0	F_{00}	F_{01}	F_{02}	F_{03}	F_{04}	F_{05}	F_{06}	F_{07}
1	F_{10}	F_{11}	F_{12}	...				
2	F_{20}	F_{21}	...					
3	F_{30}	...						
4	...							
5								
6								
7								

Рис. 1. 2D дискретне косинусне перетворення.

На наступному третьому етапі алгоритму для кожного блока формуються 18 ознак на основі 8x8 DCT коефіцієнтів F_{xy} , відповідних цьому блоку [6]. Для цього спочатку формується список з 15 таких DCT коефіцієнтів: $F_{00}, F_{01}, F_{10}, F_{20}, F_{11}, F_{02}, F_{03}, F_{12}, F_{21}, F_{30}, F_{40}, F_{31}, F_{22}, F_{13}$ і F_{04} , яким відповідає найбільша частина енергії зображення. Цей список можна отримати шляхом зіг-заг сканування матриці 8 x 8 DCT коефіцієнтів F_{xy} , починаючи з F_{00} коефіцієнта. Після цього виконується заміна першого коефіцієнта у цьому списку на два числа. Перше число є різниця F_{00} у двох блоках, що є сусідніми по горизонталі для блока, що розглядається, а друге є різниця F_{00} у двох блоках, що є сусідніми по вертикалі для того ж самого блока. Такі ж самі операції заміни виконуються для наступних двох коефіцієнтів F_{01} та F_{10} .

Таким чином ми отримуємо список із 18 значень. Ці значення далі нормуються і в результаті формується список із 18-и цілих значень – ознак блока, які є порівняно інформативними і менш залежними від напрямку та інших умов освітлення зображення. Відстань (від 0 до 1000) між двома будь-якими блоками може бути обчислена як сума абсолютних різниць відповідних ознак цих блоків, що помножена на число 1000 і поділена на суму абсолютних значень цих ознак.

На наступному четвертому етапі алгоритму обчислюється відстань між ознаками блоків двох порівнюваних зображень як усереднена сума модифікованих відстаней між відповідними блоками цих зображень, тобто блоками, що мають однакові координати на порівнюваних зображеннях. Модифікована відстань між кожними двома відповідними блоками при цьому обчислюється таким чином. Розглядається поточний блок на першому зображенні і 9 блоків на другому зображенні – блок, що є відповідним поточному блоку на першому зображенні, а також додатково ще 8 блоків, що є сусідніми до цього відповідного блока на другому зображенні. Обчислюється відстань по приведеному вище алгоритму між поточним блоком на першому зображенні і кожним із 9-ти розглянутих блоків на другому зображенні і у якості модифікованої відстані для поточного блока береться мінімальна з цих 9-и обчислених відстаней. Остаточне значення відстані обчислюється як усереднена зважена сума відповідних блоків двох порівнюваних зображень. При цьому відстані блоків, що відповідають положенню рота на обличчі, а також певним нижнім лівим та правим частинам зображення, зважуються з меншими коефіцієнтами, тому що зображення цих блоків можуть сильно змінюватись внаслідок змінення міміки людини та змінення фона на зображеннях.

Для прискорення обчислення відстані між двома зображеннями може виконуватись в два етапи. На першому етапі обчислюється не модифікована відстань між зображеннями, без використання перекриття блоків. Якщо обчислена таким чином відстань є більшою певного значення, то приймається рішення, що порівнювані два зображення відповідають різним обличчям. У протилежному разі значення відстані обчислюється більш точно по приведеному вище алгоритму.

3. Алгоритм верифікації на основі пошуку двовимірної деформації між двома фронтальними знімками

Алгоритм “StereoMatch”, який розробив Рябоконт Д.І., базується на результатах роботи [7] і виконує пошук найкращої двовимірної деформації DH , що перетворює одне зображення x в друге x' . При цьому міра близькості (відстань) Q між двома цими зображеннями є така величина:

$$Q = \sum_i \sum_j (DH(i, j) - DH(i, j-1))^2,$$

де $DH(i, j)$ - значення горизонтального зміщення (паралакса) клітки (i, j) першого зображення, що знайдено в результаті пошуку оптимальної двовимірної деформації одного знімка по відношенню до другого. Тобто паралаксом (деформацією) клітки (i, j) першого зображення x є різниця $DH(i, j) = k = i' - i$, де i' і j' – це координати відповідної їй клітки на другому зображенні. Всі припустимі значення паралакса складають множину K .

Для вирішення задачі необхідно серед усіх можливих деформацій знайти таку \bar{k}^* , що має найбільшу вірогідність $p(\bar{k}^* | x, x')$:

$$\bar{k}^* = \arg \max_{\bar{k} \in K^T} p(\bar{k} | x, x') = \arg \max_{\bar{k} \in K^T} \frac{p(\bar{k}) \cdot p(x, x' | \bar{k})}{p(x, x')} \quad (2)$$

$$\arg \max_{\bar{k} \in K^T} p(\bar{k}) \cdot p(x, x' | \bar{k}) = \arg \max_{\bar{k} \in K^T} \prod_{t \in T} g_{it}(k_t, k'_t) \cdot \prod_{t \in T} q_t(k_t)$$

де $g_{it}(k, k')$ є такі значення для всіх пар кліток (i, j) і (i', j') , що є сусідніми по горизонталі:

$$g_{it}(k, k') = g_{it}((k_1, k_2), (k'_1, k'_2)) = \begin{cases} 1 & \text{якщо } (|k_1 - k'_1| \leq \Delta) \& \\ & (|k_2 - k'_2| \leq 1) \\ 0 & \text{інакше} \end{cases}$$

$\Delta \geq 1$ є розмір по вертикалі смуги кліток другого зображення, для яких допустимо зіставлення з клітками деякого одного рядка першого зображення;

$q_t(k)$ - попередньо обчислені значення, що визначають вірогідність відповідності для кожної клітки $t \in T$ першого знімка деякого значення паралакса $k \in K$, а також вірогідність реалізації пари знімків (x, x') при умові відомих паралаксів $k(i, j)$:

$$p(x, x' | \bar{k}) = \prod_{(i,j) \in T} p(x, x' | k(i, j)) \cong \prod_{t \in T} q_t(k_t).$$

Задача (2) вирішується за допомогою метода динамічного програмування. Пошук оптимальної деформації кліток зображення виконується послідовно по рядкам кліток і кожний рядок при цьому оброблюється незалежно від усіх інших рядків [7].

З метою якісної оцінки правильності знаходження відповідних точок на рис. 4 показано чотири приклади зображень, кожний з яких отримано в результаті усереднення яскравості зображень двох різних облич (Рис. 3) згідно знайденої карти двовимірної трансформації.



Рис. 3. Вхідні зображення облич.



Рис. 4. Результат усереднення деформованих зображень.

4. Експериментальні результати

Для експериментальної перевірки алгоритмів використовувалась множина з 356-х вхідних зображень, отриманих у різний час в процесі дослідницької експлуатації у відділі 120 МННЦ ІТiС розробленої системи контролю доступу до приміщень [1]. Ця множина складається з 300 зображень 6-и співробітників відділу, а також 56-и зображень інших 52-х людей, що суттєво відрізняються між собою по умовам освітлення і в меншій мірі по орієнтації облич на цих зображеннях. Приклади вхідних зображень показані на рис. 5.



Рис. 5. Приклади вхідних зображень.

При виконанні експериментів розпізнавались зображення 6-х людей – співробітників відділу. Для кожної з цих людей будувався тільки один еталон за допомогою розробленої процедури, що будує еталони на основі аналізу вибраної оператором сукупності прикладів

зображень облич однієї людини. При цьому, якщо необхідно побудувати тільки один еталон, то виконується пошук нормалізованого зображення обличчя із вказаної множини такого, що при використанні його у якості еталона досягаються порівняно кращі результати розпізнавання всіх інших зображень, а у разі кількох таких еталонних зображень – їх усереднення. Якщо ж кількість еталонів не задана, то виконується пошук найменшої кількості еталонів, що забезпечують безпомилкове розпізнавання вказаної множини зображень облич однієї людини.

В процесі тестування алгоритмів кожен з еталонів порівнювався з усіма зображеннями свого класу для визначення помилок другого типу (FRR - відкази від розпізнавання), а також з усіма зображеннями інших класів для визначення помилки першого типу (FAR – допуск чужого). При цьому при використанні алгоритму “GetSimilarity_DCT” попередньо обчислювались описані вище ознаки порівнюваних зображень і обчислювання відстані виконувалось між сукупностями цих ознак як описано вище у розділі 2. Попередньо для кожного з 3-х алгоритмів експериментально було визначено найкраще значення порога, при якому значення помилки першого типу (FAR) було достатньо малим. Таким чином проведені експерименти відповідають тому, що верифікація виконувалась в найбільш складному режимі на основі розпізнавання лише одного єдиного вхідного зображення, хоча загалом розроблене програмне забезпечення може оброблювати відео-послідовності будь-якого розміру.

Отримані результати розпізнавання при цих підібраних порогах приведені в таблиці 1. Ці результати свідчать про наступне. Порівняно кращі результати розпізнавання має алгоритм “GetSimilarity_DCT”. Цей алгоритм є також більш “швидким” і потребує виконання меншої кількості операцій.

Таблиця 1. Результати тестування 3-х алгоритмів порівняння зображень.

Алгоритм	Кількість помилок 1-го типу	Кількість помилок 2-го типу	FAR (%)	FRR (%)
GetSimilarity_DCT	0	9	0	2.54
GetSimilarity	10	20	2.82	5.64
StereoMatch	11	28	3.1	7.9

Як відмічалось, алгоритм верифікації людини по зображенню обличчя можуть бути використаний в системах контролю доступу на основі використання магнітних карток, якщо: цей алгоритм 1) забезпечує достатньо високу надійність розпізнавання, 2) виконує верифікацію за час менший, ніж 2 сек., а також 3) потребує достатньо малої пам'яті (2 – 3 KByte) для запису на магнітній картці даних про еталон її власника. Розглянемо по черзі кожний з цих 3-х пунктів.

Із приведених вище результатів експериментів, а також інших результатів тестування, частина яких приведена в роботах [1,2], можна сподіватись, що за допомогою розроблених алгоритмів забезпечується не менша, ніж 90%, надійність правильної верифікації людини при використанні всього одного її еталона: FRR менше 10% при значно меншому значенні FAR (менше 1%). Час

повної обробки одного вхідного зображення на сучасному комп'ютері не перевищує 0.5 сек.. Тобто можна розраховувати на можливість обробки відео-послідовності із кількох вхідних зображень при верифікації, що як правило забезпечує подальше збільшення надійності розпізнавання у порівнянні з використанням лише одного вхідного зображення людини, що розпізнається.

Розглянемо тепер виконання останньої 3-ї умови – використання порівняно малої пам'яті для запису на магнітній картці даних про еталон її власника. При реалізації приведених вище алгоритмів використовувались еталони розміром 104 x 104 кліток. Збереження цих еталонів в форматі “Windows Bitmap” (*.bmp) потребує 11.6 КByte пам'яті, що є неприйнятним в наших умовах. Ось чому для вирішення задачі розглядалися всі інші відомі стандартні формати представлення даних про еталонні зображення, а також оцінювались та досліджувались інші засоби стиснення даних, зокрема на основі відомого метода кодування Хаффмана [8]. При цьому перевірка всіх відомих форматів зберігання зображень показала, що порівняно найбільш прийнятним з них є формат “JPEG” (*.jpg) на основі використання DCT перетворення та послідовного економічного стиснення сукупності отриманих DCT коефіцієнтів. При використанні цього формату нормалізоване еталонне зображення з розмірами 104 x 104 кліток потребує 6.8 КByte без будь-яких втрат даних про це зображення, а також 2.1 КByte при зниженні якості цього зображення на 20%. При цьому експериментально досліджувалось, наскільки погіршуються результати розпізнавання зображень обличчя при такому погіршенні (на 20%) якості еталона в результаті його стиснення до 2.1 КByte. Проведені експерименти на тій же сукупності зображень показують, що при використанні в алгоритмі “GetSimilarity_DCT” таких стиснених еталонів значення помилки FRR збільшується на 0.9% до FRR = 3.44% при такому ж значенні помилки FAR (0%), що є також прийнятним для нашої задачі.

Досліджувався також алгоритм стиснення еталонних зображень, що складається з таких двох етапів. На першому етапі початкове еталонне зображення трансформується у так зване “різницеве” зображення. Це виконується в процесі сканування еталонного зображення уздовж його рядків, починаючи з клітки з координатами (0,0), тобто з лівого-верхнього кута цього зображення. При скануванні послідовно розглядаються клітки еталонного зображення і значення яскравості кожної з цих кліток замінюється на різницю між значенням цієї яскравості та усередненим значенням (пів-сумою) яскравості двох кліток, що є сусідніми зліва та зверху до поточної клітки, що розглядається.

На другому етапі виконується кодування отриманого “різницевого” зображення за допомогою метода Хаффмана. Отримані оцінки такого метода стиснення еталонних зображень свідчать про те, що цей метод дозволяє стискувати нормалізоване за яскравістю еталонне зображення розміром 104 x 104 клітки у середньому до 3-х КByte без будь-яких втрат його якості.

5. Висновки

На основі приведених вище даних можна зробити такий головний висновок: алгоритм “GetSimilarity_DCT” для обчислення відстані між двома порівнюваними зображеннями, є кращим у порівнянні з двома іншими алгоритмами, що розглядалися, і може бути використаним для верифікації людини по зображенню її обличчя в системах контролю доступу з використанням магнітних карток. Порівняно кращі результати розпізнавання за цим алгоритмом досягаються завдяки тому, що ці результати є менш залежними, по-перше, від умов та напряму освітлення зображень завдяки використанню додаткової локальної нормалізації зображень за яскравістю шляхом використання різницевих значень перших трьох DCT коефіцієнтів, а по – друге, - менш залежним від змінень орієнтації та помилок нормалізації обличчя, що зумовлені головним чином не досить точним визначенням положення очей на цих зображеннях.

Література

1. В.М.Кийко, К.В.Кийко, В.В.Мацелло, Ю.Л.Провалов, В.М.Шарипанов. Система доступу до приміщення на основі розпізнавання людських обличчя, Восьма всеукраїнська міжнародна конференція УкрОБРАЗ'2006, Київ. Збірник праць. с. 123-126.
2. В.М.Кийко, К.В.Кийко, А.В.Недашковский, Д.И.Рябоконе. Распознавание изображений лиц на основе сравнения с 2D и 3D эталонами, Управляющие системы и машины, № 3, 2006, с. 3-12, 21.
3. В.М.Кийко, К.В.Кийко, А.В.Недашківський, Шлезінгер М.І., Мацелло В.В. Патент України на винахід “Спосіб комп'ютерної ідентифікації особи за зображенням її обличчя” № а2006 00777 від 30.01. 2006.
4. P.Viola and M.Jones, “Robust real-time object detection,” In Proc. of IEEE Workshop on Statistical and Computational Theories of Vision, pp. 1-25, 2001.
5. N. Ahmed, T. Natarajan, and K. R. Rao. Discrete Cosine Transform, IEEE Trans. on Computers, C-23. 1974.
6. C.Sanderson and K.K.Paliwal, “Fast features for face authentication under illumination direction changes”, Pattern Recognition Letters, Vol. 24, No. 14, 2003, pp. 2409 – 2419.
7. Рябоконе Д. Восстановление пространственной конфигурации объектов и сцен по их стереозображениям // Управляющие системы и машины. – 2005. – №1. – С. 22-31.
8. Huffman D.A. A method for the Construction of Minimum-Redundancy Codes. Proc. Of the I.R.E., Sept. 1952, pp. 1098-1102.