

14. Метод на собствените лица (собствените обекти) за разпознаване на физиономии (обекти). Линеен дискриминант на Фишер при разпознаване на лица.

Метод на собствените лица (собствените обекти) за разпознаване на физиономии (обекти)

Компютърният подход, използван от Сирович и Кърби (виж 3 и 4 от литературата), за кодиране и декодиране на изображения с лица ни помага да вникнем в информационното съдържание на изображенията с лица, подчертавайки по-значителните локални и глобални характеристики. Макар че тези характеристики могат да бъдат пряко свързани с интуитивното ни човешко разбиране за лицеви черти като очи, нос, устни и коса, това не е задължително.

На езика на компютърната теория, ние искаме да извлечем съществената информация от изображение с лице, след което да я кодираме в база данни по възможно най-ефективен начин за последващото сравнение с кодираната информация за други изображения с лица. Прост подход към извличането на информация, съдържаща се в лицево изображение, е да се улови вариацията (изменението) на постъпилото изображение спрямо колекцията от събрани лицеви изображения, в зависимост от съответните характеристики, и да използваме тази информация за кодиране и сравнение на индивидуални снимки на лица.

Тези собствени вектори могат да се разглеждат като множество от характеристики, които заедно описват вариацията между изображенията с лица. Всяко изображение от базата допринася (много или малко) за формирането на всеки собствен вектор, така че собственият вектор може да се характеризира като абстрактно (ефирно, неосезаемо, призрачно, ghostly) лице, което се нарича „*собствено лице*”. На местата, където лицевите характеристики на кодираните лица се различават, се наблюдават отклонения от еднородно сивите нюанси в собственото лице. Те представляват своеобразна карта на измененията между лицата в базата.

Всяко отделно лице може да се представи *точно* чрез линейна комбинация на *всички* собствените лица. От друга страна, на база на собствените лица с най-големи собствени стойности (т.е. тези които улавят най-голямото изменение между всички лица), могат да се получат *приближения* на всяко лице в базата. Най-добрите M собствени лица формират M -мерно пространство – „пространство на лицата” – от всички възможни изображения.

Идеята за използване на собствените лица е мотивирана от техника, разработена от Сирович и Кърби (1987, 1990), за ефективно представяне на изображения с лица, която използва метода на главните компоненти (PCA, Principal Component Analysis). Разполагайки с набор от изображения с лица, те пресмятат най-добрата координатна

система за компресиране на изображения, където всяка координата представлява изображение, което те наричат „*собствено изображение*” (картинка). Те защитават тезата, че всяка лице от колекция от изображения с лица може да бъде приблизително реконструирано от малка колекция от теглови характеристики за всяко лице и набор от стандартни изображения (собствените изображения). Теглата, които описват всяко лице се получават чрез проектиране на изображението на лицето върху всяко едно от собствените изображения.

След като набор от изображения с лица могат да бъдат реконструирани чрез сума от претеглени характеристики (т.нар. собствени изображения), то ефективен подход за запомяне и разпознаване на лица включва определяне на характеристики на база опита извлечен в течение на времето и последващо разпознаване на лица чрез сравняване на теглата им за съответните определени характеристики, необходими за приблизителното им възстановяване, с теглата на вече известните индивиди в системата. Така лицето на всеки индивид ще бъде описано от теглата на малък набор от характеристики (или собствени изображения), необходими за тяхното описание и възстановка. Това е изключително компактно представяне за съхранение в сравнение със съхраняването на целите изображения.

За подходът за разпознаване на лица са необходими следните инициализации:

1. Сдобиване с начално множество от изображения с лица (обучаващото множество);
2. Изчисляване на собствените лица от обучаващото множество, като се съхраняват само M -те изображения с най-високи собствени стойности. Тези M на брой изображения определят *пространството от лица*. С добавянето на нови лица собствените лица се обновяват и преизчисляват;
3. Изчисляване на съответното разпределение в M -мерно пространство от тегла за всеки познат индивид чрез проектиране на изображението с лицето му в „*пространството от лица*”.

Стъпки 2 и 3 се повтарят от време на време, когато се освободи процесорно време.

Разполагайки с инициализирана система, следните стъпки се прилагат при разпознаване на новопостъпило изображение с лице:

1. Изчислява се множеството от тегла на входното изображение чрез проектирането му върху всяко от M -те собствени лица;
2. Определяне на това дали входното изображение е с лице чрез проверка на това дали то е достатъчно близо до „*пространството от лица*”;
3. Ако изображението е на лице, пристъпваме към класификация на база теглата изчислени в 1. Тук определяме дали лицето е на познат човек или не;

4. (По избор) Обновяваме собствените лица и/или теглата на изображенията в системата;
5. (По избор) Ако едно и също непознато лице се забележи няколко пъти, изчисляваме неговите тегла на характеристиките в системата и ги добавяме в системата.

Подходът разглежда изображения на лица $I(x, y)$ под формата на матрица с размери N на N . За кодирането на интензитета всеки елемент от изображението (пиксел) се използват 8 бита (монохромни изображения). Изображението може също да се разглежда и като вектор с измерение N^2 , така че типично изображение с размер 256 на 256 се превръща във вектор с размер размерност 1 на 65 536, или точка в пространство с 65 536 размерности. Така дадена колекция от изображения сочи към точки в това огромно пространство.

Изображенията на лица, които като цяло имат подобна конфигурация, няма да бъдат произволно разпределени в това огромно пространство, което означава, че те могат да се опишат с относително по-малко подизмерение на разглежданото. Основната идея зад метода на главните компоненти (РСА – или разложение на Карунен - Льов) е да се намерят вектори, които най-добре описват разпределението на изображенията от лица в цялото пространство от изображения. Тези вектори дефинират подпространство от изображенията на лица, т. нар. „*пространство от лица*”. Всеки вектор е с дължина N^2 , описва изображение с размери N на N и е линейна комбинация на оригиналните изображения на лица. Тъй като тези вектори са собствените вектори на ковариационната матрица, която кореспондира на първоначалните изображения с лица, и понеже техните изображения наподобяват лица, те са наречени „*собствени лица*”.

Линеен дискриминант на Фишер при разпознаване на лица.

Линейният дискриминант на Фишер се опитва да улови „формата” на разсейването на елементите от наблюдаваното множество около средната стойност с цел по-надеждна класификация. Този метод избира W (проекцията, която най-добре отделя точките от даден клас) по такъв начин, че коефициентът на междукласовото разсейване към вътрешнокласовото разсейване да бъде максимален.

Нека дефинираме матрицата на междукласовото разсейване така:

$$S_B = \sum_{i=1}^c N_i (\mu_i - \mu)(\mu_i - \mu)^T ,$$

а матрицата на вътрешнокласовото разсейване така:

$$S_W = \sum_{i=1}^c \sum_{x_k \in X_i} N_i (x_k - \mu_i)(x_k - \mu_i)^T ,$$

където μ_i е средното изображение от клас X_i , а N_i е броят представители от клас X_i . Ако S_W не е изродена ($\det(S_W) \neq 0$), като оптимална проекция W_{opt} се избира матрицата с ортонормални¹ колони, която максимизира коефициента на детерминантата на матрицата на междукласовото разсейване на проектираните представители към детерминантата на матрицата на вътрешнокласовото разсейване на проектираните представители, т.е.

$$W_{opt} = \arg \max_W \frac{|W^T S_B W|}{|W^T S_W W|} = [w_1 w_2 \dots w_m] \quad (1),$$

където $w_i \mid i=1,2,\dots, m$ е множеството от обобщени собствени вектори на S_B и S_W , отговарящо на m -те най-големи обобщени собствени стойности $\lambda_i \mid i=1,2,\dots, m$, т.е.

$$S_B w_i = \lambda_i S_W w_i, i=1,2,\dots, m$$

Трябва да се отбележи, че има най-много $c - 1$ ненулеви обобщени собствени стойности, като така горната граница за m е $c - 1$, където c е броят класове.

За да се илюстрират ползите от специфичните за всеки клас линейни проекции (W), ще конструираме двумерен аналог на задачата за класификация, в която представителите от всеки клас лежат близо до линейно подмножество. На графиката са сравнени методът на главните компоненти (МГК) и линейният дискриминант на Фишер (ЛДФ) за задача за класификация на два класа, в които представителите на всеки клас са произволно разбъркани в перпендикулярна на линейното подпространство посока. В примера $N=20$, $n=2$ и $m=1$. Така представителите от всеки клас лежат близо до линия, която минава през началото на двуизмерното пространство на характеристиките. И двата подхода се използват за проектиране на точки от двумерното към едномерното пространство. Сравнявайки двете проекции на фигурата, можем да забележим, че методът на главните компоненти струпва класовете заедно и така те вече не са линейно разделими в пространство на проекцията. МГК постига по-голямо общо разсейване, докато ЛДФ постига по-голямо междукласово разсейване, а то улеснява класификацията.

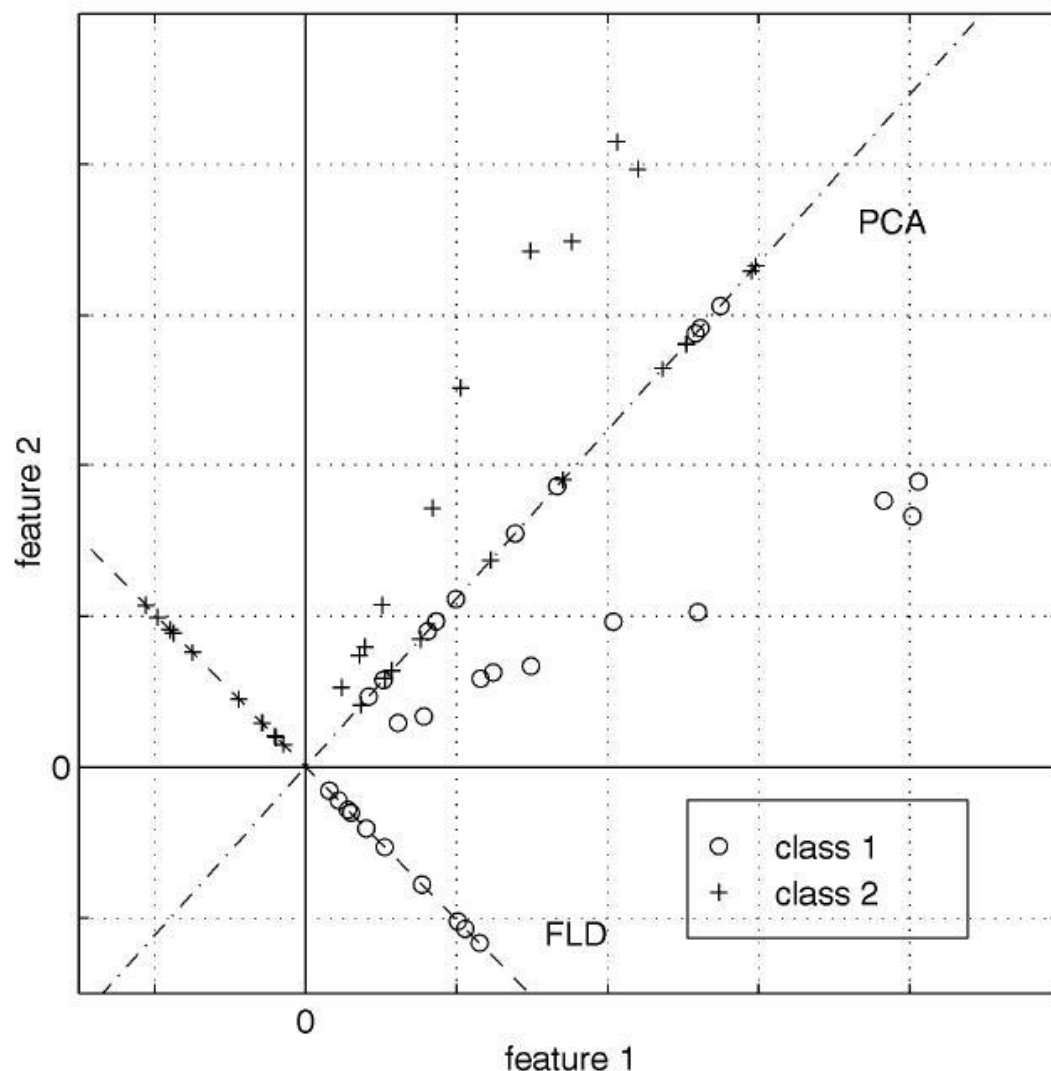
При разпознаването на лица сме изправени пред проблема, че матрицата на вътрешнокласовото разсейване $S_W \in \mathbb{R}^{n \times n}$ е винаги изродена (т.е. $\det(S_W) = 0$ и не съществува S_W^{-1}). Това произлиза от факта, че рангът на S_W е най-много $N - c$ и, в общия случай, броят на изображенията в обучаващото множество N е много по-малък от броя пиксели във всяко изображение n от множеството. Това означава, че е възможно да се

¹ Два вектора са ортонормални, когато са ортогонални ($\theta = 90^\circ$ или $\cos(\theta) = \frac{x \cdot y}{\|x\| \|y\|} = 0$, където θ е ъгълът

между двата вектора) и с единична (unit) дължина. Единичните собствени вектори на дадена матрица са ортонормални.

избере матрица W , такава че вътрешнокласовото разсейване на проектираните представители да бъде точно нула (0).

С цел да бъде избегнато усложнението, произтичащо от изродена S_W , се прилага алтернатива на (1). Този метод, наречен „лица на Фишер“, избягва проблема като проектира множеството от изображения върху пространство с по-ниско измерение така, че резултантната матрица на вътрешнокласовото разсейване S_W да не е изродена. Това се постига чрез **прилагане на метода на главните компоненти, който намалява размерността на пространството на характеристиките до $N - c$. След това се прилага** стандартната линейна дискриминанта на Фишер, дефинирана в (1), за да се намали размерността до $c - 1$.



Сравнение между метода на главните компоненти и линейния дискриминант на Фишер за задача с два класа, където данните за всеки клас лежат близо до линейното подпространство

Казано по-формално, W_{opt} се получава по следния начин: $W_{opt}^T = W_{fld}^T W_{pca}^T$, където

$$W_{pca} = \arg \max_W |W^T S_T W|, \text{ а } W_{fld} = \arg \max_W \left| \frac{W^T W_{pca}^T S_B W_{pca} W}{W^T W_{pca}^T S_W W_{pca} W} \right|.$$

Оптимизацията на W_{pca} се прави по $n \times (N-c)$ матрици с ортонормални вектор-стълбове, докато оптимизацията за W_{fld} се прави по $(N-c) \times m$ матрици с ортонормални колони. При изчисляването на W_{pca} се премахват само най-малките $c - 1$ основни (главни) компоненти.

Съществуват и други методи за намаляване на вътрешнокласовото разсейване при запазване на междукласовото разсейване. Например, при друг метод се избира W , което да максимизира междукласовото разсейване на проектираните представители след като първо се намали вътрешнокласовото разсейване. В краен случай може да максимизираме междукласовото разсейване на проектираните представители, които са обект на ограничението, че вътрешнокласовото разсейване е нула, т.е. $W_{opt} = \arg \max_{W \in M} |W^T S_B W|$, където M е множеството от $n \times m$ матрици с ортонормални вектор-стълбове, съдържащи се в ядрото (множеството от всички вектори x , за които е в сила $A\vec{x} = \vec{0}$) на матрицата на вътрешнокласовото разсейване S_W .

Литература:

1. Belhumeur, P.N., J.P. Hespanha, and D.J. Kriegman, Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection, IEEE Trans. on PAMI, vol. 19, no. 7, 1997, pp.711-720;
2. Turk M., A. Pentland: Eigen faces for Recognition, Journal of Cognitive Neuroscience, Vol.3, No.1, MIT, 1991, pp.71-86;
3. Sirovich, L. & Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. Journal of the Optical Society of America A, 4(3), 519-524.
4. Kirby, M., & Sirovich, L. (1990). Application of the Karhunen-Loeve procedure for the characterization of human faces. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(1).
5. <http://dailyburrito.com/projects/facerecog/FaceRecReport.html>;
6. <http://www.scholarpedia.org/article/Fisherfaces>;
7. http://en.wikipedia.org/wiki/Principal_component_analysis;
8. http://en.wikipedia.org/wiki/Linear_discriminant_analysis.

От автора: Темата е развивана предимно от т. 1 и т. 2. Съветвам да се прочетат и двете статии. За разбирането им е необходимо да се припомнят метода на главните компоненти (PCA) и линейния дискриминантен анализ (LDA or FDA). За разяснение на някои от термините на български съм ползвал Wikipedia (не задължително само от т. 7 и 8) и каквото съм открил в интернет.

Моля за извинение за допуснати грешки и неточности, но темата е развивана предимно по малките часове на нощта, а и не е редактирана.