

ОБРАБОТКА ИЗОБРАЖЕНИЙ, РАСПОЗНАВАНИЕ ОБРАЗОВ

МЕТОДЫ ДВУМЕРНОЙ ПРОЕКЦИИ ЦИФРОВЫХ ИЗОБРАЖЕНИЙ В СОБСТВЕННЫЕ ПОДПРОСТРАНСТВА: ОСОБЕННОСТИ РЕАЛИЗАЦИИ И ПРИМЕНЕНИЕ

Г.А. Кухарев¹, Н.Л. Щеголева¹

¹Федеральное государственное автономное образовательное учреждение высшего образования
«Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)»,
Санкт-Петербург, Россия

Аннотация

Рассматриваются алгоритмы проекции цифровых изображений в собственные подпространства в рамках линейных методов PCA, LDA, PLS и CCA. Приводится история развития этих методов за последние 100 лет на фоне появления новых областей их применения и меняющихся в связи с этим требований к ним. Показано, что развитие было инициировано четырьмя основными требованиями, вытекающими из современных задач и практики цифровой обработки изображений и, в первую очередь, изображений лиц. Первым является требование использования методов PCA, LDA, PLS и CCA в условиях как малой, так и чрезвычайно большой выборки изображений лиц в исходных наборах. Второе требование связано с критерием, определяющим собственный базис, который должен обеспечить, например, минимум ошибки аппроксимации изображений лиц, улучшение кластеризации в собственном подпространстве или максимум корреляции (ковариации) между наборами данных в подпространстве. Третье – связано с возможностью приложения рассматриваемых методов к задачам обработки двух и более наборов изображений с различных сенсорных источников или нескольких наборов любых числовых матриц. Именно эти три требования обусловили появление, развитие и применение методов двумерной проекции в собственные подпространства – 2DPCA, 2DLDA, 2DPLS и 2DCCA. В статье рассмотрены несколько основных ветвей алгоритмической реализации этих методов (итерационные, не итерационные, на основе SVD и т.д.), оценены их достоинства и недостатки, а также показаны примеры их использования на практике. Наконец, четвертое требование – возможность реализации двумерных проекций изображений лиц (или других числовых матриц) непосредственно в слоях сверточных нейронных сетей (CNN/Deep NN) и/или интеграции их функций в состав NN отдельными блоками. В настоящей статье обсуждается это требование и рассматриваются примеры решений. Приводятся оценки вычислительной сложности для представленных алгоритмов и примеры решения конкретных задач обработки изображений.

Ключевые слова: наборы изображений лиц и числовых матриц, собственный базис и собственные подпространства, анализ главных компонент (PCA), линейный дискриминантный анализ (LDA), частичный метод наименьших квадратов (PLS), канонический корреляционный анализ (CCA), преобразование Карунена–Лоэва (KLT), 2DPCA/2DKLT, 2DPLS/2DKLT, 2DCCA/2DKLT, CNN, Deep NN.

Цитирование: Кухарев, Г.А. Методы двумерной проекции цифровых изображений в собственные подпространства: особенности реализации и применение / Г.А. Кухарев, Н.Л. Щеголева // Компьютерная оптика. – 2018. – Т. 42, № 4. – С. 637-656. – DOI: 10.18287/2412-6159-2018-42-4-637-656.

Введение

Истоки методов проекции в собственные подпространства. Основные идеи методов проекций в собственные подпространства как одного из инструментов математической обработки наблюдений и инструмента выявления и/или установления связей в наблюдениях были заложены в работах [1–4]. Так, в [1] было показано решение задачи аппроксимации экспериментальных данных, представленных на плоскости набором точек, с использованием оригинального критерия, при котором всегда достигается минимум ошибки аппроксимации. Опираясь на этот результат, в [2] был представлен метод поиска **главных компонент** в наборе переменных, где каждая переменная

была представлена отдельным вектором эмпирических данных. И именно здесь введено понятие собственных векторов и собственных чисел (eigenvector, eigenvalue), алгоритм выбора главных компонент (Principal Component Analysis – PCA) и проекция исходных данных в собственное подпространство. В [3] обсуждается задача линейного дискриминантного анализа – нахождения функции, определяющей наилучшее разделение популяций данных в собственном подпространстве. Для решения этой задачи предлагается критерий, позволяющий минимизировать внутриклассовое и максимизировать межклассовое расстояние в подпространстве, что можно рассматривать как улучшение кластеризации в нем. Описан алгоритм построения собственного

базиса, основанный на этом критерии, выбор главных компонент и процедура проекции в собственное подпространство. Решения [3] создали возможность более эффективного решения задач распознавания образов при упрощенной структуре классификаторов, поскольку разделение данных в подпространстве выполнено наилучшим образом из всех возможных. В [4] представлен критерий максимизации взаимной корреляции в собственном подпространстве для двух независимых (в общем случае) наборов данных и необходимый для этого алгоритм поиска собственного базиса, общего для двух наборов исходных данных. Решения [4] создали возможность детальной разработки канонического корреляционного анализа (Canonical Correlation Analysis – CCA) и частичных наименьших квадратов (Partial Least Squares – PLS), как методов проекции в общее собственное подпространство для двух и более наборов исходных данных.

Все упомянутые выше методы проекции в собственное подпространство реализуются в два последовательных этапа. Первый этап включает построение собственного базиса для исходных данных и выбор главных компонент. Наибольшие вычислительные затраты здесь связаны с решением задач на собственные значения. И если размерность исходного пространства признаков равна D , то сначала вычисляются необходимые матрицы ковариации и общие матрицы рассеяния (в соответствии с требуемым критерием), порядок которых равен D , а потом вычисляются D собственных чисел и D соответствующих им собственных векторов, а также определяются d главных компонент ($d \ll D$). В общем случае вычислительные затраты здесь составляют $O_1 \cong D^3$.

На втором этапе реализуется проекция исходных данных в собственное подпространство (как преобразование в собственном базисе или преобразование Карунена–Лоэва) и выполняется редукция размерности пространства признаков до значений d . При этом вычислительные затраты здесь составляют $O_2 \leq D^2$.

1. Проблемы реализации PCA для изображений

В задачах построения собственных базисов для наборов цифровых изображений учитываются пять базовых параметров: $\{M, N, K, L, Q\}$. Здесь: M и N – число строк и столбцов исходных изображений; K и L определяют число классов изображений в наборе исходных данных и число изображений в каждом классе; Q – число наборов изображений. В общем случае $K > 1$, $L \geq 1$ и $Q \geq 1$. Если $Q = 1$, то все изображения размещены в одном наборе исходных данных. Варианты, когда $Q \geq 2$, представлены в параграфе 5.

Параметры M и N задают размерность D исходного пространства признаков так, что $D = MN$ и определяют вычислительную сложность реализуемых алгоритмов. При этом задаче реализации проекции изображений в собственные подпространства сопутствуют две проблемы. Первая из них – проблема малого числа изображений в наборе по сравнению с размерностью D , что определяется как **проблема малой выборки**, когда

$D \gg KL$. Вторая – **проблема чрезвычайно большой выборки**, то есть когда $KL \gg D \gg 1000$ (например, KL может составлять от нескольких тысяч до миллиона изображений, что характерно для современных баз изображений лиц или мультимедийных баз данных).

Напомним, что в подходах [2–4] все эти операции выполнялись для наборов исходных данных, представленных в векторной форме. Поэтому первые попытки применения, например, методов PCA и LDA для цифровых изображений появились почти под конец XX века [5–7], а методов CCA и PLS уже только в XXI веке, что было связано как с невозможностью прямого переноса методов [2–4] на изображения (рассматриваемые как двумерные структуры данных), так и с возникающим при этом большим объемом вычислений.

С учетом этого, методология использования PCA в приложении к обработке изображений строилась, с одной стороны, с необходимостью использования базового метода PCA [2], изначально ориентированного на обработку векторных данных, а с другой стороны, с учетом условия

$$D < KL, \quad (1)$$

для обеспечения устойчивости решения задач на собственные значения. При этом каждое исходное изображение преобразовывалось путем конкатенации его столбцов (или строк) в общий вектор размером $MN \times 1$, и тогда весь набор исходных данных уже содержал KL таких векторов. Далее к исходному набору данных применялись несколько подходов, «позволяющих лавировать» между отмеченными выше ограничениями и условием (1).

Если, например, возникает проблема малой выборки (то есть не выполняется условие (1)), а значение $KL \ll 1000$, то весь набор $\{X\}$ исходных изображений представляется матрицей размера $MN \times KL$. В этом случае вместо обычных матриц ковариации используются матрицы Грамма G порядка KL так, что $G = X^T X$. При этом данные в матрице X должны быть центрированы относительно среднего. Далее для матрицы G решалась задача на собственные значения, что обеспечивало вычисление KL собственных значений и соответствующих им собственных векторов (каждый размером $KL \times 1$), которые могли быть пересчитаны в весь состав собственных векторов [5]. При этом существенно сокращался объем вычислений. Вычислительные затраты на этапе решения задачи на собственные значения в этом случае составляют $O \cong (KL)^3$. Проекция в собственное подпространство реализуется как одномерное преобразование Карунена–Лоэва (1D KLT) векторизованных исходных данных путем их умножения на матрицу проекции с затратами $O \cong (KL)^2$.

Если выполнялось условие (1), но при этом размерность D исходного пространства признаков была велика (например, $D \gg 1000$), то решение основывалось на уменьшении каждого исходного изображения до размера $m < M$ и $n < N$, а далее выполнялось его преобразование в общий вектор размером $mn \times 1$. По набору таких векторов (центрированных относительно среднего

вектора по набору) формировалась матрица ковариации порядка mn . Матрица проекции (также порядка mn) в подпространство определялась далее на основе решения задачи на собственные значения для матрицы ковариации. Строки матрицы проекции (размером $mn \times 1$) являются искомыми собственными векторами матрицы ковариации. При этом проекция в собственное подпространство реализуется как одномерное преобразование Карунена–Лозва (1D KLT) векторов (представляющих центрированные исходные данные) путем их умножения на матрицу проекции.

Значения m и n необходимо выбирать из условия устойчивости решения задачи на собственные значения. Алгоритмически задача уменьшения исходного пространства признаков $MN \rightarrow mn$ для исходных изображений решается с использованием процедур уменьшения размера изображения или путем перехода из исходного пространства яркостных признаков в другое пространство – например, в гистограмму яркости, градиенты яркости [6], спектр дискретного преобразования Фурье [3] и т.д. Вычислительные затраты на задачи собственных значений в этих случаях составляют $O \cong (mn)^3$, а на одномерное преобразование Карунена–Лозва составляют $O \cong (mn)^2$.

При этом надо помнить, что уменьшение разрешения изображений, связанное с его физическим размером до значений $m \times n$, может приводить к потере информации в них. Поэтому если возникает проблема чрезвычайно большой выборки, а предыдущее решение с уменьшением исходного пространства признаков недопустимо из-за потери важной информации в исходных изображениях, то практического решения задачи построения собственного базиса для цифровых изображений в рамках идей [2] не достичь! Наконец, следует отметить, что при «искусственной векторизации» изображений в векторах не сохраняются пространственные отношения соседних элементов. При этом утрачивается и возможность полного использования корреляции, имеющейся в исходных матрицах между соседними элементами.

Естественно, что эти проблемы – невозможность использования PCA для изображений при граничных значениях базовых параметров, потеря информации и утрата полного использования корреляции между исходными элементами – касается также методов LDA, PLS и CCA.

Однако решение этой проблемы неожиданно появилось в 1998 году при новом подходе к реализации преобразования Карунена–Лозва в приложении к изображениям лиц. Подход был назван «Vector Based Approximation of KLT» – VKLT [8]. Основная его идея состоит в том, что в отличие от [2, 5, 6] в VKLT вычисляются две матрицы ковариации C_0 и R_0 , которые определяются отдельно по столбцам и строкам всех исходных изображений без их конкатенации в вектор. Далее вычисляются две матрицы проекции Φ_1 и Φ_2 , определяемые собственными векторами матриц ковариации C_0 и R_0 . Проекция исходных изображений в новое пространство признаков реализуется как дву-

мерное (по строкам и столбцам) преобразование Карунена–Лозва так, что

$$Y_k = \Phi_1^T X_k \Phi_2, \text{ для всех } k = 1, 2, \dots, K, \quad (2)$$

где X_k , Y_k – изображение и его проекция в собственном подпространстве, что и отличает подход [8] от [2, 5–6]. Заметим, что в (2) исходное изображение X_k и его проекция Y_k в подпространстве представлены матрицами размера $M \times N$.

Описанные в [8] эксперименты по распознаванию изображений лиц были выполнены на изображениях лиц базы ORL [24]. При этом авторами была отмечена более высокая робастность метода VKLT (в сравнении с KLT) по отношению к качеству исходных изображений, в том числе их «зашумлению», циклическому сдвигу и изменениям яркости. Также отмечалось существенное снижение требуемого объема памяти и вычислительных затрат при реализации задач на собственные значения (то есть при вычислении матриц проекции Φ_1 и Φ_2).

Идеи работы [8] представлены в монографиях [9–11] в форме PCArc (PCA по строкам и столбцам изображений) и исследованы в приложении к распознаванию изображений лиц в условиях малой выборки (Small Sample Size – SSS), а также в условиях низкого разрешения и аддитивного шума тестовых изображений. В [9] впервые было показано, что при любом $K \geq 1$ проблема SSS в рамках PCArc не возникает, а робастные свойства, отмеченные в [8], сохраняются. Также были рассмотрены примеры отображения изображений лиц на главные компоненты в 3D-подпространствах для PCArc и (PCArc+LDA). В [10] также был представлен пакет «Моделирование систем распознавания лиц» (Face Recognition System Modeler/FaRes-MOD), в рамках которого реализованы итерационные алгоритмы двумерной проекции в собственные подпространства PCArc и LDArc. Пакет FaRes-MOD реализован на платформе Borland® C++Builder™, работает под управлением ОС Windows и переносим на все его современные версии. С 2003 года пакет используется в Политехническом университете города Щецин в Польше и Санкт-Петербургском государственном электротехническом университете (ЛЭТИ) в учебных целях по курсам «Основы биометрии» и «Распознавание изображений». В [11] представлены векторно-матричные процедуры, позволяющие непосредственно реализовать методы двумерной проекции и редукции размерности пространства признаков в алгоритмических языках, поддерживающих матричные операции. Здесь рассмотрены методы улучшения кластеризации в собственном подпространстве при использовании каскада методов проекции PCArc+LDA.

Дальнейшее развитие идей [8–11] было представлено в публикациях уже как методы 2DLDA/2DKLT и 2DPCA/2DKLT [12–14]. В такой форме записи впервые в технической литературе подчеркивается два этапа выполнения этих методов: первый этап – анализ исходных данных (включающий формирование общих матриц рассеяния по заданному критерию, вычисление

матриц проекции и определение главных компонент) и следующий за ним второй этап – двумерное преобразование Карунена–Лозва. Кроме того, эта форма записи отличается от других методов 2DLDA и 2DPCA, реализованных по итерационным алгоритмам. В целом, в этих публикациях представлен сравнительный анализ методов 1D- и 2D-проекции в собственные подпространства и показаны примеры решения задач кластеризации, компрессии и распознавания изображений лиц на различных бенчмарковых базах. При этом в [13] впервые были представлены все характеристики метода 2DPCA/2DKLT. В [14] была представлена система распознавания изображений лиц, использующая метод 2DLDA/2DKLT при наличии **только одного (!) изображения** в каждом классе базы обучения. Представленное здесь решение на практике показывает возможность «обхода проблемы SSS», обычно сопутствующей задачам обработки изображений лиц. Далее нами были описаны неитерационные алгоритмы (двумерной проекции в собственные подпространства) в параллельной и каскадной формах их реализации.

2. Формальное описание метода 2D PCA/2D KLT

Пусть нам задан набор X , состоящий из K матриц, где каждая матрица представляет изображение размером $M \times N$, причем $MN \gg K$:

$$X = [X^{(1)} X^{(2)} \dots X^{(K)}], \quad \forall k = 1, 2, \dots, K. \quad (3)$$

Целью 2DPCA является определение двух матриц проекции W_1 и W_2 , трансформирующих исходные данные (3) в собственное подпространство признаков при выполнении условия минимальной ошибки аппроксимации, что можно записать следующим образом:

$$\|X^{(k)} - \hat{X}^{(k)}\| \rightarrow \min, \quad \forall k, \quad (4)$$

где $\hat{X}^{(k)}$ – результат аппроксимации k -го изображения по главным компонентам.

Версии параллельного и каскадного алгоритмов 2DPCA/2DKLT приведены в табл. 1.

Редукция размерности пространства признаков реализуется как «усеченное 2DKLT», при котором в дальнейших преобразованиях участвуют только те собственные векторы, которые соответствуют d главным компонентам. Для этого из матрицы W_1^T выбираем d строк, а из матрицы W_2 выбираем d столбцов, соответствующих d наибольшим собственным числам, и на их основе формируем две матрицы редукции F_1 и F_2 . При этом $d < \min(M, N)$ или $d_1 < M$; $d_2 < N$, если $d_1 \neq d_2$.

Верхние границы параметра d определяются исходя, например, из критерия энергетической значимости собственных значений, а нижние границы выбираются с учетом желаемого качества аппроксимации исходных данных. В этом случае «усеченное двумерное преобразование Карунена–Лозва» реализуется следующим образом:

$$\hat{Y}^{(k)} = F_1 \bar{X}^{(k)} F_2, \quad \forall k. \quad (5)$$

Матрицы F_1 и F_2 в (5) имеют размеры $(d \times M)$ и $(N \times d)$ соответственно или $(d_1 \times M)$ и $(N \times d_2)$ в общем случае. Знак « \wedge » определяет отличие результата ап-

проксимации от результата реконструкции $Y^{(k)}$. При этом результат (5) – матрица размера $d_1 \times d_2$ представляет исходные изображения в собственном подпространстве признаков. Теперь можно представить условие (4) в новой форме:

$$\|X^{(k)} - F_1^T \hat{Y}^{(k)} F_2^T\| \rightarrow \min, \quad \forall k = 1, 2, \dots, K. \quad (6)$$

Схематически основные этапы выполнения 2DPCA для центрированных (относительно среднего) исходных данных: вычисление двух матриц ковариации, решение двух задач на собственные значения, выбор двух границ для главных компонент и формирование левой F_1 и правой F_2 матриц проекции, реализация процедуры 2DKLT в параллельном алгоритме и организация вычислений в каскадном алгоритме 2DPCA/2DKLT показаны в [53].

Характеристики метода 2DPCA/2DKLT

Размерности пространства признаков при решении задач на собственные значения в методе 2DPCA/2DKLT определяются числом строк и столбцов изображений (то есть значениями M и N). Поэтому наибольшая размерность исходного пространства признаков определяется как $\max(M, N)$.

При представлении K исходных изображений как совокупности строк и столбцов общее число получаемых при этом векторов составляет величину $K(M+N)$. Поэтому при любых значениях $\{M, N, K\}$ соотношение «размерность исходного пространства признаков/число векторов» всегда будет отвечать условию $\max(M, N) \ll K(M+N)$, «обходя» таким образом проблему SSS.

Порядок матриц ковариаций составляет M и N , что предопределяет практическую возможность решения задачи на собственные значения и стабильность этого решения даже для изображений относительно больших размеров. Для матриц ковариации, вычисляемых отдельно по строкам и столбцам одного исходного изображения, потребуется $2MN(M+N)$ операций, а по всем K изображениям – $2KMN(M+N)$ операций.

Объем вычислений на решение задач на собственные значения для базы изображений с параметрами $\{M, N, K, L\}$ составит не более $((M)^3 + (N)^3)$ при нахождении всех собственных чисел и соответствующих им собственных векторов. Отметим, что операционная сложность для метода 1DPCA определяется значениями $(MN)^3$ или $(KL)^3$. Если, например, положить, что $MN = KL$ и $M = N$, то сокращение вычислений для метода 2DPCA в сравнении с методом 1DPCA определится отношением $M^3/2$.

На реализацию одного полного 2D KLT потребуется $MN(M+N)$ операций, а на реализацию процедуры редукции размерности пространства признаков (РПП) в рамках «усеченного» 2D KLT потребуется $Md(d+N)$ операций (d – число главных компонент). При этом сокращение вычислений можно приблизительно оценить как величину $(M+N)/d$, поскольку:

$$\frac{NM^2 + MN^2}{Md^2 + MNd} = \frac{MN(M+N)}{M(d+N)d} \approx \frac{M+N}{d},$$

при условии $d \ll N$.

Табл. 1. Параллельный и каскадный алгоритмы для 2DPCA/2DKLT

Параллельный алгоритм 2DPCA/2DKLT	Каскадный алгоритм 2DPCA/2DKLT
Вход: матрицы $X^{(k)} \in \mathbb{R}^{M \times N}, \forall k=1, 2, \dots, K$.	Вход: матрицы $X^{(k)} \in \mathbb{R}^{M \times N}, \forall k=1, 2, \dots, K$.
Выход: матрицы $\bar{X}, W_1, W_2, \Lambda_1, \Lambda_2, Y^{(k)}$.	Выход: матрицы $\bar{X}, W_1, W_2, \Lambda_1, \Lambda_2, Y^{(k)}$.
<p>1. Вычислить средний образ из всех данных</p> $\bar{X} = \frac{1}{K} \sum_{k=1}^K X^{(k)}.$ <p>2. Центрировать исходные данные:</p> $\bar{X}^{(k)} = X^{(k)} - \bar{X}, \forall k.$ <p>3. Вычислить две матрицы ковариации относительно <u>строк</u> и <u>столбцов</u> матриц-результатов, полученных в п. 2:</p> $C^{(r)} = \sum_{k=1}^K \bar{X}^{(k)} (\bar{X}^{(k)})^T, C^{(c)} = \sum_{k=1}^K (\bar{X}^{(k)})^T \bar{X}^{(k)}.$ <p>4. Решить две задачи на собственные значения:</p> $C^{(r)} W_1 = \Lambda_1 W_1 \text{ и } C^{(c)} W_2 = \Lambda_2 W_2,$ <p>определив при этом диагональные матрицы собственных значений (Λ_1, Λ_2) и матрицы W_1 и W_2 собственных векторов (или левую и правую матрицы проекций).</p> <p>5. Упорядочить собственные значения по убыванию и переставить в соответствии с новым порядком столбцы матриц W_1 и W_2.</p> <p>6. Выполнить проекцию исходных данных в собственное пространство, реализовав двумерное преобразование Карунена–Лоэва:</p> $Y^{(k)} = W_1^T \bar{X}^{(k)} W_2, \forall k.$ <p>7. Для реконструкции исходных данных, по результату $Y^{(k)}$, выполнить обратное двумерное преобразование Карунена–Лоэва:</p> $\bar{X}^{(k)} = W_1 Y^{(k)} W_2^T, \forall k.$ <p><i>Конец</i></p> <p>Параметры всех вычисленных выше матриц</p> $\bar{X} \in \mathbb{R}^{M \times N}; Y^{(k)} \in \mathbb{R}^{M \times N};$ $C^{(r)} \in \mathbb{R}^{M \times M}; C^{(c)} \in \mathbb{R}^{N \times N};$ $\{W_1, \Lambda_1\} \in \mathbb{R}^{M \times M}; \{W_2, \Lambda_2\} \in \mathbb{R}^{N \times N}.$ <p>Матрицы Λ_1, Λ_2 – диагональные.</p> <p>Матрицы W_1, W_2 – ортогональные так, что:</p> $W_1^T W_1 = I_1; W_2^T W_2 = I_2, \text{ где } I_1, I_2 - \text{единичные матрицы}$ <p>порядков M и N.</p> <p>P.S. Матрицы, вычисленные в каскадном алгоритме, также соответствуют этим параметрам.</p>	<p>1. Вычислить средний образ из всех данных</p> $\bar{X} = \frac{1}{K} \sum_{k=1}^K X^{(k)}.$ <p>2. Центрировать исходные данные:</p> $\bar{X}^{(k)} = X^{(k)} - \bar{X}, \forall k.$ <p>3. Вычислить ковариационную матрицу относительно <u>строк</u> матриц-результатов п. 2:</p> $C^{(r)} = \frac{1}{MN} \sum_{k=1}^K \bar{X}^{(k)} (\bar{X}^{(k)})^T.$ <p>4. Решить первую задачу на собственные значения:</p> $C^{(r)} W_1 = \Lambda_1 W_1,$ <p>определив при этом диагональную матрицу собственных значений Λ_1 и матрицу W_1 собственных векторов (или левую матрицу проекций).</p> <p>5. Упорядочить собственные значения по убыванию и переставить в соответствии с новым порядком столбцы матрицы W_1.</p> <p>6. Выполнить проекцию исходных данных в промежуточное подпространство, реализовав преобразование Карунена–Лоэва:</p> $X_1^{(k)} = W_1^T \bar{X}^{(k)}, \forall k.$ <p>7. Вычислить матрицу ковариации по <u>столбцам</u> матрицы результата (6):</p> $C^{(c)} = \sum_{k=1}^K (X_1^{(k)})^T X_1^{(k)}.$ <p>8. Решить вторую задачу на собственные значения:</p> $C^{(c)} W_2 = \Lambda_2 W_2$ <p>и таким образом вычислить матрицу собственных значений Λ_2 и матрицу собственных векторов – правую матрицу проекции W_2.</p> <p>9. Упорядочить собственные значения по убыванию и переставить в соответствии с новым порядком столбцы матрицы W_2.</p> <p>10. Выполнить проекцию исходных данных в собственное подпространство, реализовав KLT по одномерной или двумерной форме:</p> $Y^{(k)} = X_1^{(k)} W_2 \equiv W_1^T \bar{X}^{(k)} W_2, \forall k.$ <p><i>Конец</i></p>

Например, для $M=112$ и $N=92$ (база лиц ORL [24]) и $d=10$ сокращение вычислений составит примерно 20 раз (!) на каждое изображение. С учетом параметра K – числа изображений, ускорение вычислений составит величину, равную примерно $K(M+N)/d$ на все исходные данные. Результат (РППП) содержит $(d \times d)$ или $(d_1 \times d_2)$ элементов, поэтому РППП определяется соотношением $MN/(dd)$ или $MN/(d_1 d_2)$, если $d_1 \neq d_2$. Например, для $M=112$, $N=92$ и $d=10$ пространство признаков сократится более чем 100 раз (!) на каждое изображение.

3. Алгоритм реализации 2DLDA/2DKLT

Отдельно рассмотрим алгоритм реализации 2DLDA/2DKLT для набора изображений, впервые представленный в мае 2005 года [12].

В этом случае исходные данные должны быть структурированы (разбиты на классы) и в каждом классе должно быть не менее двух изображений.

Пусть нам задан набор X , состоящий из KL матриц $X^{(k,l)}$, где каждая матрица представляет изображение размером $M \times N$, а $k=1, 2, \dots, K$ и $l=1, 2, \dots, L$.

Целью 2DLDA является определение левой и правой матриц проекции W_1 и W_2 , трансформирующих исходные данные в собственное подпространство признаков так, что $X^{(k,l)} \rightarrow Y^{(k,l)}, \forall k \text{ и } \forall l$, при выполнении условия минимизации внутриклассового и максимизации межклассового расстояния в собственном подпространстве. Ниже, в табл. 2, приводится псевдокод для алгоритма 2DLDA/2DKLT.

Табл. 2. Псевдокод для алгоритма 2DLDA/2DKLT

<p>Вход: матрицы $X^{(k,l)} \in \mathbb{R}^{M \times N}$, $\forall k=1, 2, \dots, K; \forall l=1, 2, \dots, L$. Выход: матрицы \bar{X}, W_1, W_2, Λ_1, Λ_2, $Y^{(k,l)}$.</p> <ol style="list-style-type: none"> 1. Вычислить средний образ в каждом классе $\bar{X}^{(k)} = \frac{1}{L} \sum_{l=1}^L X^{(k,l)}, \forall k.$ 2. Вычислить средний образ для всех данных: $\bar{X} = \frac{1}{K} \sum_{k=1}^K \bar{X}^{(k)}.$ 3. Вычислить «внутриклассовую (Within-class) и межклассовую» (Between-class) матрицы ковариаций, определенные относительно строк: $W^{(r)} = \sum_{k=1}^K \sum_{l=1}^L (X^{(k,l)} - \bar{X}^{(k)}) (X^{(k,l)} - \bar{X}^{(k)})^T;$ $B^{(r)} = \sum_{k=1}^K (\bar{X}^{(k)} - \bar{X}) (\bar{X}^{(k)} - \bar{X})^T.$ 4. Вычислить «внутриклассовую (Within-class) и межклассовую» (Between-class) матрицы ковариаций, определенные относительно столбцов: $W^{(c)} = \sum_{k=1}^K \sum_{l=1}^L (X^{(k,l)} - \bar{X}^{(k)})^T (X^{(k,l)} - \bar{X}^{(k)});$ $B^{(c)} = \sum_{k=1}^K (\bar{X}^{(k)} - \bar{X})^T (\bar{X}^{(k)} - \bar{X}).$ 5. Выполнить регуляризацию матриц $W^{(r)}$ и $W^{(c)}$ перед их обращением и регуляризацию обеих матриц рассеяния: $S^{(r)} = [W^{(r)}]^{-1} B^{(r)} \text{ и } S^{(c)} = [W^{(c)}]^{-1} B^{(c)}.$ 6. Решить две задачи на собственные значения: $S^{(r)} W_1 = \Lambda^{(r)} W_1; \quad S^{(c)} W_2 = \Lambda^{(c)} W_2.$ <p>Выполнить проекцию исходных данных в собственное пространство, реализовав двумерное преобразование Карунена–Лоэва: $Y^{(k,l)} = W_1^T \bar{X}^{(k,l)} W_2, \quad \forall k.$</p> <p>Для реконструкции исходных данных по результату $Y^{(k,l)}$ выполнить обратное двумерное преобразование Карунена–Лоэва: $\bar{X}^{(k,l)} = W_1 Y^{(k,l)} W_2^T, \quad \forall k.$</p> <p><i>Конец</i></p>	<p>Редукция размерности пространства признаков</p> <p>Из матрицы W_1^T выбираем d строк, а из матрицы W_2 выбираем d столбцов, соответствующих d наибольшим собственным числам, и на их основе формируем две матрицы редукции F_1 и F_2. В этом случае «усеченное двумерное преобразование Карунена–Лоэва» реализуется следующим образом: $\hat{Y}^{(k,l)} = F_1 \bar{X}^{(k,l)} F_2, \quad \forall k,$ что аналогично преобразованиям в алгоритме 2DPCA/2DKLT.</p> <p>Матрицы F_1 и F_2 имеют размеры $(d \times M)$ и $(N \times d)$ соответственно или $(d_1 \times M)$ и $(N \times d_2)$ в общем случае. Знак «\wedge» отличает результат аппроксимации от результата реконструкции $Y^{(k,l)}$.</p> <p>В общем случае исходные данные могут представлять K наборов изображений, где каждый набор состоит из L_k изображений так, что $L_k \geq 2$, а $K \geq 3$. Эти два условия позволяют вычислить среднее изображение в классе, а также определить не менее двух признаков для отображения собственного подпространства на плоскости. В условиях обработки изображений достаточно иметь только два изображения на класс. Для $K=4$ и $L=2$ собственное подпространство показано в [53].</p> <p>В этой работе максимальное расстояние в классах равно 0,02, минимальное расстояние между классами составило 0,21, поэтому параметр качества кластеризации составил 4,78. Каждая выделенная точка представляет отдельное изображение в трехмерном собственном подпространстве, а пара связанных точек представляет отдельный класс. И получить такое представление для 8 изображений, каждое из которых имеет размер 320×240 пикселей, можно только в рамках двумерных методов проекции.</p>
<p>Параметры всех вычисленных выше матриц $\{\bar{X}, \bar{X}^{(k)}, Y^{(k,l)}\} \in \mathbb{R}^{M \times N}; \{C^{(r)}, W_1, \Lambda_1\} \in \mathbb{R}^{M \times M}; \{C^{(c)}, W_2, \Lambda_2\} \in \mathbb{R}^{N \times N}.$ Матрицы Λ_1, Λ_2 – диагональные; матрицы W_1, W_2 – близкие к ортогональным.</p>	

4. Развитие методов двумерной проекции ИЛ в собственные подпространства

Продолжая историю методов проекции в собственные подпространства, отметим, что некоторые «новые подходы к реализации 2DPCA» рассматривались в работе [15], вышедшей в свет после работ [8–11]. Однако в работе [15] допущены принципиальные ошибки в представлении базовой процедуры 2DPCA. Во-первых, в [15] анализ главных компонент выполнялся *только по одному направлению* – а именно по столбцам исходных изображений (!). Во-вторых, по другому направлению (по строкам) анализ главных компонент не выполнялся, но изменялся размер изображения. Естественно, что и KLT здесь реализуется также только по одному направлению, как и редукция размерности пространства признаков. В-третьих, для возможности практического применения подхода [15] для задач рас-

познавания изображений лиц здесь применялось предварительное уменьшение размера исходных изображений (что не требуется в методах [8–14]). Именно поэтому подходы, представленные в [15], не относятся к методам двумерной проекции в собственные подпространства и не могут называться 2DPCA.

Развивая идеи работы [15], David Zhang опубликовал в декабре 2005 года статью [16], в которой представил *mix* из понятий «Two-Directional» и «Two-Dimension», что, однако, еще больше усложнило понимание идей двумерной проекции в собственные подпространства в приложении к обработке изображений.

Тем не менее, подходы [15] имели и других последователей, которые их развивали. Суть их сводилась к предварительной декомпозиции исходного изображения в набор подобластей. Именно это и позволяло «напрямую» реализовать метод [15] для таких подобластей. Число публикаций, описывающих формы таких подоб-

ластей, способы их получения и использования, сегодня уже не охватить ссылками в рамках библиографии настоящей статьи. Все эти статьи доступны в интернете по ключевым словам, например, «Diagonal PCA», «Matrix PCA», «Extension PCA», «Mixture of Bilateral-Projection 2DPCA». Отметим также, что практически во всех этих работах одним из соавторов является David Zhang, идеи [15] которого мы исследовали выше. И перечисление всех работ Davida Zhanga было бы здесь не совсем корректным по отношению к другим оригинальным идеям.

Начиная с 2005 года алгоритмы двумерной проекции в собственные подпространства (в рамках PCA) были представлены и другими группами авторов, но уже как «Generalized Low Rank Approximations of Matrices», «Generalized 2DPCA», и даже как «2DSVD» [17–19]. В этих работах есть ссылка на статью [15], не представляющую идеи 2DPCA, но нет ссылок ни на работу [8], ни на работы [9–14], что довольно странно для научных публикаций.

Что же касается «метода 2DSVD» из [19], то тут следует отметить два важных факта: если параметры базы изображений определены как $L = 1$ и $K = 1$, то метод 2DPCA/2DKLT «вырождается» в метод SVD; если $K > 1$ и $L \geq 1$, метод 2DPCA/2DKLT не становится методом 2DSVD, поскольку в общем случае матрицы ковариации, вычисляемые в методах 2DPCA, имеют полный ранг и поэтому не являются сингулярными. И этот второй факт не позволяет назвать 2DPCA/2DKLT, как 2DSVD. Некорректность применения названия 2DSVD в [19] возникла по ассоциации с тем, что PCA можно реализовать через SVD. Очевидно также, что решение задач на собственные значения в 2DPCA может быть реализовано через два независимых SVD, но это совсем не одно и то же, поскольку в состав всего алгоритма 2DPCA/2DKLT входят не только задачи на собственные значения.

Интересно отметить, что «метод 2DSVD» используется в основном как инструмент обработки карт местности для их компрессии, кодирования видеоданных, обработки рукописных текстов для задач их классификации и числовых матриц для задач их низкоранговой аппроксимации [20–22]. Методы 2DPCA традиционно применяются как инструмент в задачах лицевой биометрии – для редукции исходного пространства признаков, распознавания изображений лиц и построения их моделей. Более подробно это будет представлено в следующем параграфе.

5. Методы двумерных проекций для двух наборов изображений лиц

Казалось бы, очень странно, что методы двумерных проекций в собственные подпространства получили наибольшее развитие и применение в задачах обработки изображений лиц. Однако этому факту есть объяснение, представленное ниже. Особенностью обработки изображений лиц (ИЛ) является то, что в рамках соответствующих систем могут использоваться различные сенсорные датчики ИЛ, различные методы предобработки ИЛ и различные формы их представления. В общем случае ИЛ могут быть представлены в форме

2D-изображений в видимом (VIS), тепловом (NIR) и инфракрасном свете (IR), в виде композиционных фотороботов, рисованных скетчей и популяций из них; в форме «range image», в форме контурных моделей области лица (Active Shape Model – ASM), моделей внешнего вида (Active Appearance Model – AAM), определяющих текстуру ИЛ, и, наконец, в форме 3D-моделей ИЛ.

При этом исходные данные на входе системы могут быть одновременно представлены несколькими наборами ИЛ, связанными «в пары», «тройки» или даже «группы наборов ИЛ более трех». Примером последнего варианта является шесть наборов ИЛ, представляющих лицо одного и того же человека, из которых первый набор содержит ИЛ(VIS), второй – ИЛ(NIR), а третий – ИЛ(IR), четвертый – ИЛ в форме скетчей, пятый набор содержит ИЛ в форме 3D и, наконец, шестой представлен набором «IR - Range Images». В отечественной биометрии такие наборы относят к данным «мультисенсорной природы». В зарубежной литературе их относят к группе «гетерогенных данных». Наиболее часто встречающиеся в технической литературе связи в рамках взаимных трансформаций ИЛ показаны в [53].

С одной стороны, наличие разнообразных способов представления ИЛ существенно расширяет возможности и области применения систем поиска и распознавания людей по ИЛ. С другой стороны, это разнообразие значительно усложняет структуру соответствующих систем распознавания ИЛ, алгоритмы их функционирования и саму реализацию таких систем. Например, чтобы реализовать поиск ИЛ, оперировать подобными данными и интерпретировать их, необходимо связать их между собой (например, два изображения разной сенсорной или разной семантической природы). Далее, если предположить, что между этими данными есть взаимное соответствие, то при запросе на поиск данных, входящих в один из наборов, можно легко найти соответствующий ему результат из другого набора. Необходимое взаимное соответствие (подобное взаимному индексированию) достигается на этапах обучения по выборкам мультисенсорных (или гетерогенных) данных.

С учетом этого в последние годы в обработке ИЛ сформировались новые классы задач – «Heterogeneous Face Recognition and Matching», «Cross-Modal Face Matching», «Face Image Indexing and Retrieval», а также и более общие подходы для поиска информации, например, «Cross-Modal Multimedia Retrieval». Главная особенность этих задач – работа с двумя или более наборами ИЛ (или другими формами данных).

Минусом же всех цитируемых выше методов проекции в собственные подпространства являлось то, что они изначально были ориентированы на обработку только одного набора исходных данных, хотя параметры $K \geq 1$ и $L \geq 1$ этих данных могли определять их сложную структурную организацию. Однако, несмотря на то, что эти методы не соответствуют современным классам задач обработки изображений,

некоторые их решения удалось получить в рамках как одномерных, так и двумерных методов PCA и LDA. В этом случае каждый отдельный набор исходных данных проецировался в отдельное собственное подпространство, а далее использовалась взаимная трансформация изображений между этими подпространствами. Например, в [23] эксперименты выполнены на базе 1DPCA и двух наборах ИЛ – наборе фото и наборе соответствующих им скетчей [24]. При этом в [23] показана возможность построения моделей взаимной реконструкции для пар фото–скетч, а также решение задачи распознавания фото и скетчей с помощью таких моделей. Позднее на идеях, заложенных в [23], удалось решить и задачи суперразрешения между фото и скетчами (часто эти задачи в приложении к ИЛ называют задачами «галлюцинации») и задачи генерации популяций ИЛ. Так, в [25] эксперименты выполнены на основе 1DPCA и на основе 2DPCA/2DKLT для двух наборов, содержащих смешанные группы изображений. Эти группы включают: фото и скетчи, изображения VIS и NIR, фото и соответствующие им контурные изображения, а также изображения, семантически не связанные в соответствующих парах. Источники этих изображений приведены в [25]. Пример результата взаимной реконструкции изображений для таких наборов приведен в [53]: в рядах 1 и 2 представлены обучающие наборы данных; ряд 3 – результат реконструкции, полученный по изображениям ряда 2; ряд 4 – изображения ряда 2 в низком разрешении с добавлением шума; ряд 5 – результат реконструкции ИЛ из изображений ряда 4. В процессе обучения исходные данные трансформируются в собственные подпространства, а также формируются две матрицы взаимной трансформации изображений, которые далее используются для взаимного преобразования изображений в парах. Методы, представленные в [25], были использованы в работе [53] также для генерации популяций фото по заданным скетчам.

И все же решение отмеченных выше задач наиболее полно реализуется только на идеях ССА (заложенных еще в начале 20 века [4]) и идеях PLS-регрессии, впервые реализованных в эконометрии и развитых в хемометрии. В приложении к цифровой обработке многомерных сигналов (а по факту изображений) алгоритмы PLS, например, можно рассматривать как «расширение методов PCA/SVD на два и более набора исходных данных» [26].

Методы ССА и PLS позволяют отобразить два набора исходных данных (часто не коррелированных, или внешне подобных, но семантически не связанных между собой) в общее собственное подпространство признаков, в котором они сильно коррелируют. Это позволяет связать пары исходных данных в собственном подпространстве с помощью общей регрессионной модели, чтобы представлять и понимать одни наблюдения, действия (или явления) через другие, устраняя тем самым полностью или частично сенсорный или семантический разрыв между ними.

Первые применения ССА и PLS для обработки ИЛ появились в начале текущего столетия. А первыми задачами были: реконструкции одних ИЛ по другим (например, в [27] рассмотрена реконструкция вида $VIS(rgb) \rightarrow 3D$ и $VIS(rgb) \rightarrow NIR$); индексация, распознавание и реконструкция ИЛ в гетерогенных наборах ИЛ (например, в [28] использованы формы «Range Image» и текстура цветного ИЛ); согласование параметров моделей внешнего вида для ИЛ (Active Appearance Model [29]). При этом следует отметить: на начальных стадиях применения методов ССА и PLS задачи обработки гетерогенных (или мультисенсорных) данных решались на основе одномерных алгоритмов их реализации, что требовало предварительного представления ИЛ в форме векторов. И здесь снова появилась проблема SSS, сопутствующая задачам обработки ИЛ. Для ее пояснения рассмотрим основные идеи одномерных алгоритмов реализации ССА и PLS в соответствии с [26].

Пусть нам заданы два набора исходных данных, состоящие из K числовых матриц, а каждая матрица представляет ИЛ размером $M \times N$, причем $MN \gg K$, что является обычным для задач обработки ИЛ. Каждое ИЛ «развернем в вектор» путем конкатенации его строк (или столбцов) и центрируем относительно среднего значения в этом векторе. В этом случае мы получаем два набора исходных данных X и Y , состоящие из K векторов размером $MN \times 1$ каждый так, что:

$$\begin{aligned} X &= [X^{(1)} X^{(2)} \dots X^{(K)}]; \\ Y &= [Y^{(1)} Y^{(2)} \dots Y^{(K)}], \quad \forall k=1, 2 \dots K. \end{aligned} \quad (7)$$

Целью ССА является нахождение двух матриц проекции W_x и W_y , трансформирующих исходные данные X и Y в собственное подпространство – подпространство канонических переменных

$$\begin{aligned} U &= W_x^T X, \\ V &= W_y^T Y \end{aligned} \quad (8)$$

так, чтобы выполнялось условие: $\|U - V\| \rightarrow \min$.

Пары исходных данных $X^{(k)}$ и $Y^{(k)}$ при этом могут быть не связаны между собой корреляцией, в то время как канонические переменные $U^{(k)}$ и $V^{(k)}$, определяемые как

$$\begin{aligned} U^{(k)} &= W_x^T X^{(k)}, \\ V^{(k)} &= W_y^T Y^{(k)}, \end{aligned} \quad (9)$$

связываются между собой устойчивой корреляцией, максимум которой достигается при решении двух совместных задач на собственные значения [26]:

$$\left\{ \begin{aligned} (C_{xx}^{-1} C_{xy} C_{yy}^{-1} C_{yx}) W_x &= \Lambda_x W_x \\ (C_{yy}^{-1} C_{yx} C_{xx}^{-1} C_{xy}) W_y &= \Lambda_y W_y \end{aligned} \right\}, \quad (10)$$

где C_{xx} , C_{yy} , C_{xy} , C_{yx} – матрицы ковариации порядков MN , причем

$$\begin{aligned} C_{xx} &= XX^T; \quad C_{yy} = YY^T; \\ C_{xy} &= XY^T; \quad C_{yx} = C_{xy}^T; \end{aligned}$$

Λ_x, Λ_y – диагональные матрицы собственных значений; W_x и W_y – матрицы собственных векторов (матрицы проекций).

Целью PLS также является нахождение двух матриц проекции W_x и W_y , трансформирующих исходные данные в собственное подпространство так, чтобы выполнялось условие максимизации ковариации между переменными в этом подпространстве. При этом пары исходных данных $X^{(k)}$ и $Y^{(k)}$ также могут быть не связаны между собой корреляцией, в то время как пары переменных $U^{(k)}$ и $V^{(k)}$ в собственном подпространстве связываются между собой устойчивой корреляцией, максимум которой достигается при решении двух следующих совместных задач на собственные значения:

$$\begin{cases} (C_{xy}C_{yx})W_x = \Lambda_x W_x; \\ (C_{yx}C_{xy})W_y = \Lambda_y W_y, \end{cases} \quad (11)$$

где матрицы взаимной ковариации определены аналогично (10).

Исходные данные, структуру вычислений в ССА и PLS, а также результаты U и V схематически представлены в [53].

Вычисления в ССА и PLS заканчиваются выбором значения параметра d – редукции размерности пространства признаков (причем $d < MN$). Далее реализуется одномерное преобразование Карунена–Лоэва (ПКЛ), в результате чего получаем полное представление переменных U и V в собственном подпространстве или их представление только по главным компонентам. В последнем случае ПКЛ реализуется с использованием только тех собственных функций, которые соответствуют главным компонентам.

Обратим внимание на соотношение (10), в котором общие матрицы рассеяния получаются в результате перемножения четырех симметричных (по определению) матриц порядков MN . Две из них – матрицы автоковариации – требуют предварительного обращения, что для матриц, **формируемых по ИЛ** (и в условиях $MN \gg K$), не является тривиальным, поскольку ранг этих матриц может быть меньше их порядка. Общие матрицы рассеяния также имеют порядки MN , но не являются симметричными. Решение обобщенной задачи на собственные значения в этом случае будет либо неустойчивым, либо невозможным в принципе уже при значениях $M > 100$ и $N > 100$ (хотя, например, в стандартах биометрии используются ИЛ размером 320×240). Очевидно, что эти характеристики ограничивали широкое применение методов одномерного ССА в задачах обработки изображений. Два подхода для реализации ССА и PLS с использованием процедур SVD представлены в работах [30, 31].

В статье [30] рассмотрено решение задачи слежения за областью лица в процессе формирования модели 3D-формы ИЛ. Здесь область интересов на ИЛ представлена не всеми MN яркостными признаками, а небольшим (≤ 100) набором антропометрических координат на области лица. В этом случае можно реализовать ССА с использованием трех процедур SVD, что позволяет избежать вычисления исходных кова-

риационных матриц из (10), требуемых здесь обращений матриц, а также обойтись без решения задач на собственные значения.

В статье [31] рассмотрено решение задач кросс-модального распознавания ИЛ различной сенсорной природы, включая скетчи, фотооригиналы, ИЛ с высоким и низким разрешением и различной позой и т.д. Однако в этом случае ССА через SVD может быть применено только для небольшого числа пар ИЛ, причем на размеры изображений также накладываются ограничения. Последнее связано с тем, что порядки ортогональных матриц, формируемые в рамках SVD, определяются значениями MN и K из выражения (7). Кроме того, все решения в этом подходе могут быть реализованы только в вариантах выполнения проекций по одному направлению, что требует представления пар ИЛ в векторной форме (7).

«Революция» в решениях задач обработки двух наборов ИЛ началась в 2007 году с появлением первых статей, посвященных методам 2DCCA [32, 33].

В [32] представлен итерационно-каскадный алгоритм *вычисления двух матриц проекции*. Основной недостаток этих решений – сложность понимания и реализации итерационно-каскадного алгоритма 2DCCA в условиях априорного выбора числа итераций, параметра сходимости и последующего обеспечения сходимости. И, хотя авторы отмечают существенное снижение вычислительных затрат в рамках 2DCCA (в сравнении с 1DCCA), в выполненных ими экспериментах размер ИЛ не превышал размеров 50×50 . А это также косвенно подчеркивает сложность реализации и применения итерационно-каскадного алгоритма 2DCCA.

В [33] представлена реализация метода «2DCCA», основанная на идеях работы [15]. И, естественно, что все особенности (а точнее недостатки) из работы [15] перешли на этот метод: «2DCCA» здесь реализуется по одному направлению, в то время как по другому направлению выполняется только уменьшение размера исходного изображения. Но даже и в этом подходе (при выполнении экспериментов) исходные ИЛ необходимо было предварительно уменьшать до 28×23 , что также свидетельствует о несовершенстве этого метода.

Несмотря на отмеченное несовершенство подходов [32, 33], они определили две независимые ветви применения ССА в задачах обработки ИЛ. Хотя в публикациях, опирающихся на эти подходы, часто так и остаются неясными параметры реализации ССА (и/или размеры исходных и выходных данных) в проводимых экспериментах.

Фузия признаков, популяции признаков и каскады методов проекции в собственные подпространства

В параграфе 1 мы отмечали проблему SSS, которую можно условно связать с малой репрезентативностью исходных данных. В работах [34–36] предложены решения, позволяющие «обходить» эту проблему. Предлагаемые здесь решения основаны на создании популяций по исходным данным.

Так, в работе [34] на примере базы YALE-B исследуется решение задачи распознавания ИЛ в нестабильных условиях их освещения и малого числа эталонов. Уменьшение влияния изменения яркости изображений достигается в рамках предварительного их логарифмирования и использования четырех специальных процедур выделения признаков из результатов логарифмирования. Этим достигается создание четырех популяций признаков из каждого ИЛ. При этом блоками системы распознавания являются: четыре блока PCA и шесть блоков CCA. Четыре блока PCA (1–4) формируют вторичные признаки (по главным компонентам) от каждой популяции и далее образуют шесть возможных пар наборов признаков (1 и 2; 1 и 3; 1 и 4; 2 и 3; 2 и 4; 3 и 4). Эти пары наборов поступают на входы шести блоков CCA, а результат выбирается по методу ближайшего соседа с эталоном. Реализованные в [40] решения основаны на идеях работ [15] и [33] и сохраняют все их недостатки.

В работе [35] представлен алгоритм, названный 3DCCA, предназначенный для куба векторных данных. Предварительной фильтрацией исходных данных по каждому направлению (X, Y, Z) формируются по два набора вторичных признаков. Эти наборы поступают на вход трех блоков CCA, а результаты проекции в собственное подпространство объединяются (фузируются) в общий вектор. Отмечаются три достоинства предложенного решения: сокращение вычислительных затрат; логичность и обоснованность объединения признаков и высокая результативность распознавания. К минусам можно отнести то, что решения 3DCCA основаны на одномерных CCA итерационного типа.

В работе [36] на примерах данных «The Texas 3D Face Recognition Database» рассмотрена задача взаимного распознавания 2D-изображений лиц и их аналогов, представленных в форме «range images». Последние являлись картами глубин для формирования 3D-форм лиц, но в их текстуре не различимы не только примитивы лица, но даже и границы их областей. В статье представлено несколько вариантов архитектуры соответствующих систем распознавания. Первая основана на фузии 2D-изображений лиц и «range images», что обеспечивает улучшение качества исходных данных, используемых для обучения. Вторая использует предварительную процедуру PCA для уменьшения объема признакового пространства на входе блока CCA (в статье она названа как «архитектура CCA-PCA»). Третья и четвертая архитектура систем основана на вариантах каскадирования блоков CCA (архитектура «CCADouble»). Авторы отмечают высокую результативность взаимного распознавания 2D-изображений лиц и «range images». Однако во всех случаях исходные изображения представлялись в форме векторов, что снижает интерес к этой работе.

Подводя итог работам [34–36], можно отметить, что они являются «промежуточным мостиком» между идеями [15] и идеями двумерных методов проекции в собственные подпространства, не требующими

уменьшения размеров исходных изображений и включающих их векторизованное представление.

Первый неитерационный алгоритм 2DCCA/2DKLT, продолжающий идеи [8–11], представлен нами в статье [37]. В записи «2DCCA/2DKLT» подчеркивается два этапа выполнения – анализ исходных данных по двум направлениям (по столбцам и строкам ИЛ) с выбором главных компонент и двумерное преобразование Карунена–Лоэва.

6. Неитерационные алгоритмы реализации 2DCCA/2DKLT и 2DPLS/2DKLT

Пусть заданы два набора исходных данных X и Y , состоящие из K матриц размером $M \times N$, причем $MN \gg K$:

$$X = [X^{(1)} X^{(2)} \dots X^{(K)}] \text{ и } Y = [Y^{(1)} Y^{(2)} \dots Y^{(K)}], \quad \forall k = 1, 2, \dots, K. \quad (12)$$

Матрицы $X^{(k)}$ и $Y^{(k)}$ образуют связанные пары – например, ИЛ одного человека, но полученные от разных сенсорных источников.

Целью алгоритмов 2DCCA/2DKLT и 2DPLS/2DKLT является нахождение двух пар матриц $\{W_{x1}, W_{y1}\}$ и $\{W_{x2}, W_{y2}\}$, используемых как левые и правые матрицы проекции в реализации двумерного преобразования Карунена–Лоэва исходных данных $X^{(k)}$ и $Y^{(k)}$, $\forall k$. Проекция исходных данных в общее собственное подпространство $X^{(k)} \rightarrow U^{(k)}$ и $Y^{(k)} \rightarrow V^{(k)}$ реализуется так, чтобы выполнялось условие:

$$\|U^{(k)} - V^{(k)}\| \rightarrow \min, \quad (13)$$

которое в [4] определено как критерий максимума корреляции между переменными в собственном подпространстве:

$$\operatorname{argmax} = \operatorname{Cov}(U, V). \quad (14)$$

$$\{W_{x1}, W_{x2}, W_{y1}, W_{y2}\}$$

В общем случае исходные данные $X^{(k)}$ и $Y^{(k)}$ могут быть не связаны между собой корреляцией, в то время как канонические переменные $U^{(k)}$ и $V^{(k)}$ связываются между собой устойчивой корреляцией.

Именно это свойство CCA и PLS создает основу для решения задач «Heterogeneous Face Recognition and Matching», «Cross-Modal Face Matching», «Face Image Indexing and Retrieval», задач «Face Hallucination» и «Face Super-resolution» и даже задач «Cross-Modal Multimedia Retrieval».

Версии алгоритмов 2DCCA/2DKLT и 2DPLS/2DKLT, записанные в форме псевдокодов, представлены ниже, в табл. 3.

Редукция размерности исходного пространства признаков реализуется как «усеченное 2DKLT», при котором в дальнейших преобразованиях участвуют только те собственные векторы, которые соответствуют d главным компонентам. Для этого из левых матриц проекции $\{W_{x1}^T, W_{y1}^T\}$ выбираем d строк, а из правых матриц проекции $\{W_{x2}^T, W_{y2}^T\}$ выбираем d столбцов, соответствующих d наибольшему собственным числам, и на их основе формируем четыре матрицы

редукции $\{Fx_1, Fx_2\}$ и $\{Fy_1, Fy_2\}$. При этом $d < \min(M, N)$ или $d_1 < M$; $d_2 < N$, если $d_1 \neq d_2$. Верхние границы параметра d определяются исходя, например,

из критерия энергетической значимости собственных значений, а нижние границы выбираются с учетом желаемого качества аппроксимации исходных данных.

Табл. 3. Псевдокоды для 2DCCA/2DKLT и 2DPLS/2DKLT

Алгоритм 2DCCA/2DKLT	Алгоритм 2DPLS/2DKLT
Вход: $\{X^{(k)}, Y^{(k)}\} \in \mathbb{R}^{M \times N}, \forall k = 1, 2, \dots, K$	Вход: $\{X^{(k)}, Y^{(k)}\} \in \mathbb{R}^{M \times N}, \forall k = 1, 2, \dots, K$
Выход: $\{\bar{X}, \bar{Y}\}, \{\Lambda_{x1}, \Lambda_{x2}, \Lambda_{y1}, \Lambda_{y2}\}, \{W_{x1}, W_{x2}, W_{y1}, W_{y2}, U^{(k)}, V^{(k)}\}, \forall k.$	Выход: $\{\bar{X}, \bar{Y}\}, \{\Lambda_{x1}, \Lambda_{x2}, \Lambda_{y1}, \Lambda_{y2}\}, \{W_{x1}, W_{x2}, W_{y1}, W_{y2}, U^{(k)}, V^{(k)}\}, \forall k.$
<p>1. Вычислить средние образы из данных:</p> $\bar{X} = \frac{1}{K} \sum_{k=1}^K X^{(k)}; \bar{Y} = \frac{1}{K} \sum_{k=1}^K Y^{(k)}.$ <p>2. Центрировать исходные данные $\forall k$:</p> $\bar{X}^{(k)} = X^{(k)} - \bar{X}; \bar{Y}^{(k)} = Y^{(k)} - \bar{Y}.$ <p>3. Вычислить ковариационные матрицы относительно <u>строк</u> матриц из п. 2:</p> $C_{xx}^{(r)} = \sum_{k=1}^K \bar{X}^{(k)} (\bar{X}^{(k)})^T; C_{yy}^{(r)} = \sum_{k=1}^K \bar{Y}^{(k)} (\bar{Y}^{(k)})^T;$ $C_{xy}^{(r)} = \sum_{k=1}^K \bar{X}^{(k)} (\bar{Y}^{(k)})^T; C_{yx}^{(r)} = (C_{xy}^{(r)})^T.$ <p>4. Вычислить ковариационные матрицы относительно <u>столбцов</u> матриц из п. 2:</p> $C_{xy}^{(c)} = \sum_{k=1}^K (\bar{X}^{(k)})^T \bar{Y}^{(k)}; C_{yx}^{(c)} = (C_{xy}^{(c)})^T.$ $C_{xx}^{(c)} = \sum_{k=1}^K (\bar{X}^{(k)})^T \bar{X}^{(k)}; C_{yy}^{(c)} = \sum_{k=1}^K (\bar{Y}^{(k)})^T \bar{Y}^{(k)}.$ <p>5. Вычислить общие матрицы рассеяния:</p> $\begin{cases} S^{(1,r)} = [C_{xx}^{(r)}]^{-1} C_{xy}^{(r)} [C_{yy}^{(r)}]^{-1} C_{yx}^{(r)}, \\ S^{(2,r)} = [C_{yy}^{(r)}]^{-1} C_{yx}^{(r)} [C_{xx}^{(r)}]^{-1} C_{xy}^{(r)}, \\ S^{(1,c)} = [C_{xx}^{(c)}]^{-1} C_{xy}^{(c)} [C_{yy}^{(c)}]^{-1} C_{yx}^{(c)}, \\ S^{(2,c)} = [C_{yy}^{(c)}]^{-1} C_{yx}^{(c)} [C_{xx}^{(c)}]^{-1} C_{xy}^{(c)}. \end{cases}$ <p>6. Решить 4 задачи на собственные значения:</p> $\begin{cases} S^{(1,r)} W_{x1} = \Lambda_{x1} W_{x1}; & S^{(1,r)} W_{x2} = \Lambda_{x2} W_{x2}; \\ S^{(2,r)} W_{y1} = \Lambda_{y1} W_{y1}; & S^{(2,r)} W_{y2} = \Lambda_{y2} W_{y2} \end{cases}$ <p>и определить матрицы собственных значений (Λ_{*1} и Λ_{*2}) и матрицы собственных векторов, а по факту – левые (W_{x1}, W_{y1}) и правые (W_{x2}, W_{y2}) матрицы проекций.</p> <p>7. Выполнить проекцию исходных данных в собственное пространство, реализовав двумерное преобразование Карунена–Лозва:</p> $U^{(k)} = W_{x1}^T \bar{X}^{(k)} W_{x2}, \quad \forall k,$ $V^{(k)} = W_{y1}^T \bar{Y}^{(k)} W_{y2}, \quad \forall k.$ <p>8. Для реконструкции исходных данных выполнить обратное двумерное преобразование Карунена–Лозва:</p> $\bar{X}^{(k)} = W_{x1} U^{(k)} W_{x2}^T, \quad \forall k;$ $\bar{Y}^{(k)} = W_{y1} V^{(k)} W_{y2}^T, \quad \forall k.$ <p>Конец алгоритма Размеры всех вычисленных выше матриц $\{\bar{X}^{(k)}, \bar{Y}^{(k)}\} \in \mathbb{R}^{M \times N}; C^{(r)} \in \mathbb{R}^{M \times M}; C^{(c)} \in \mathbb{R}^{N \times N};$ $\{W_{x1}, W_{y1}, \Lambda_{*1}\} \in \mathbb{R}^{M \times M};$ $\{W_{x2}, W_{y2}, \Lambda_{*2}\} \in \mathbb{R}^{N \times N}.$ Матрицы: C – симметричные; $\Lambda_{*1}, \Lambda_{*2}$ – диагональные; W_{*1}, W_{*2} – «почти ортогональные».</p>	<p>1. Вычислить средние образы из данных:</p> $\bar{X} = \frac{1}{K} \sum_{k=1}^K X^{(k)}; \bar{Y} = \frac{1}{K} \sum_{k=1}^K Y^{(k)}.$ <p>2. Центрировать исходные данные $\forall k$:</p> $\bar{X}^{(k)} = X^{(k)} - \bar{X}; \bar{Y}^{(k)} = Y^{(k)} - \bar{Y}.$ <p>3. Вычислить ковариационные матрицы относительно <u>строк</u> матриц по п. 2:</p> $C_{xy}^{(r)} = \sum_{k=1}^K \bar{X}^{(k)} (\bar{Y}^{(k)})^T;$ $C_{yx}^{(r)} = (C_{xy}^{(r)})^T.$ <p>4. Вычислить ковариационные матрицы относительно <u>столбцов</u> матриц из п. 2:</p> $C_{xy}^{(c)} = \sum_{k=1}^K (\bar{X}^{(k)})^T \bar{Y}^{(k)};$ $C_{yx}^{(c)} = (C_{xy}^{(c)})^T.$ <p>5. Вычислить общие матрицы рассеяния:</p> $\begin{cases} S^{(1,r)} = C_{xy}^{(r)} C_{yx}^{(r)}, \\ S^{(2,r)} = C_{yx}^{(r)} C_{xy}^{(r)}, \\ S^{(1,c)} = C_{xy}^{(c)} C_{yx}^{(c)}, \\ S^{(2,c)} = C_{yx}^{(c)} C_{xy}^{(c)}. \end{cases}$ <p>6. Решить 4 задачи на собственные значения:</p> $\begin{cases} S^{(1,r)} W_{x1} = \Lambda_{x1} W_{x1}; & S^{(1,r)} W_{x2} = \Lambda_{x2} W_{x2}; \\ S^{(2,r)} W_{y1} = \Lambda_{y1} W_{y1}; & S^{(2,r)} W_{y2} = \Lambda_{y2} W_{y2} \end{cases}$ <p>и определить матрицы собственных значений (Λ_{*1} и Λ_{*2}) и матрицы собственных векторов, а по факту – левые (W_{x1}, W_{y1}) и правые (W_{x2}, W_{y2}) матрицы проекций.</p> <p>7. Выполнить проекцию исходных данных в собственное пространство, реализовав двумерное преобразование Карунена–Лозва:</p> $U^{(k)} = W_{x1}^T \bar{X}^{(k)} W_{x2}, \quad \forall k,$ $V^{(k)} = W_{y1}^T \bar{Y}^{(k)} W_{y2}, \quad \forall k.$ <p>8. Для реконструкции исходных данных выполнить обратное двумерное преобразование Карунена–Лозва:</p> $\bar{X}^{(k)} = W_{x1} U^{(k)} W_{x2}^T, \quad \forall k;$ $\bar{Y}^{(k)} = W_{y1} V^{(k)} W_{y2}^T, \quad \forall k.$ <p>Конец алгоритма Размеры всех вычисленных выше матриц $\{\bar{X}^{(k)}, \bar{Y}^{(k)}\} \in \mathbb{R}^{M \times N};$ $C^{(r)} \in \mathbb{R}^{M \times M}; C^{(c)} \in \mathbb{R}^{N \times N};$ $\{W_{x1}, W_{y1}, \Lambda_{*1}\} \in \mathbb{R}^{M \times M};$ $\{W_{x2}, W_{y2}, \Lambda_{*2}\} \in \mathbb{R}^{N \times N}.$ Матрицы: C – симметричные; $\Lambda_{*1}, \Lambda_{*2}$ – диагональные; W_{*1}, W_{*2} – ортогональные.</p>

Усеченное двумерное преобразование Карунена-Лоэва реализуется следующим образом:

$$\hat{U}^{(k)} = F_{x1} \bar{X}^{(k)} F_{x2}; \quad \hat{V}^{(k)} = F_{y1} Y^{(k)} F_{y2}; \quad \forall k. \quad (13)$$

Знак «^» определяет отличие результата аппроксимации от результата реконструкции $U^{(k)}$ и $V^{(k)}$. Матрицы $\{F_{x1}, F_{y1}\}$ и $\{F_{x2}, F_{y2}\}$ имеют размеры $(d \times M)$ и $(N \times d)$ соответственно или $(d_1 \times M)$ и $(N \times d_2)$ в общем случае. При этом результат (5) – матрица размера $d \times d$ (или $d_1 \times d_2$) представляет исходные изображения в собственном подпространстве признаков.

7. Практическая реализация CCA и PLS

а) ИЛ, являясь данными, полученными с различных источников, могут содержать как пропуски данных (области, состоящие из нулевых по значениям строк и столбцов), так и области с повторяющимися (одинаковыми по значениям) данными. В этом случае ранг матриц ковариаций (пункт 4 приведенного выше псевдокода) при $K < 10$ может быть меньше их порядка. Для устранения этого эффекта достаточно будет на исходные изображения наложить шум с амплитудой $\leq 5\%$ от максимума по используемому диапазону яркости изображений. Иногда в алгоритме 2DCCA вместо центрирования исходных изображений X и Y необходимо выполнить их нормирование (например, $\hat{X}^{(k)} = X^{(k)} / \|X\|$). В этом случае норма каждого изображения в наборах X и Y будет равна 1, а среднее значение близко к единице, что будет соответствовать их «выбеливанию».

б) Как показано в пункте 5 приведенного выше псевдокода, при вычислении общих матриц рассеяния в алгоритме 2DCCA используются операции обращения матриц автоковариаций, которые могут быть плохо обусловлены. Поэтому перед обращением этих матриц необходимо выполнить их регуляризацию.

с) Общие матрицы рассеяния S в алгоритме 2DCCA необходимо также подвергнуть регуляризации ввиду их несимметричности. Это во многом упростит решение задач на собственные значения (СЗ).

д) В теории алгоритмы 2DCCA и 2DPLS требуют решения четырех задач на СЗ (см. пункт 6 псевдокода). Однако, как в алгоритме 2DCCA/2DKLT, так и в алгоритме 2DPLS/2DKLT можно сократить число решаемых задач на СЗ до двух (!).

Покажем, как это сделать для алгоритма 2DCCA/2DKLT. Рассмотрим левую пару задач на СЗ (в пункте 6 псевдокода), которая определена относительно строк изображений наборов X и Y . Решаем первую задачу на СЗ и находим матрицу проекции W_{x1} . Вторую матрицу W_{y1} вычисляем без решения задачи на СЗ следующим образом [26]:

$$W_{y1} = [C_{yy}^{(r)}]^{-1} C_{yx}^{(r)} W_{x1}. \quad (14)$$

А далее вычисленные собственные векторы (то есть столбцы матрицы W_{y1}) необходимо нормировать

к единице, поделив их на «собственную норму». Также следует поступить со второй (правой) парой задач на СЗ, то есть решить одну задачу на СЗ для получения матрицы проекции W_{x2} и пересчитать из нее матрицу W_{y2} .

Для алгоритма 2DPLS/2DKLT поступаем аналогично, однако без использования матриц автоковариаций в произведении (14). Обоснование и выяснение этих решений можно найти, например, в работах [26, 37]. А теперь перечислим и оценим вычислительные затраты на оставшиеся операции, которые необходимо выполнить при реализации параллельного алгоритма 2DCCA для наборов изображений X и Y с параметрами $\{M, N, K\}$, где $MN \gg K$ и когда 2DCCA выполняется по всем яркостным признакам:

- 1) формирование шести матриц ковариаций: $O(3K(M^2N + N^2M))$;
- 2) обращение 4 матриц автоковариаций: $O(2(M^3 + N^3))$;
- 3) вычисление матрицы проекции W_{x1} (решение задачи на СЗ): $O(M^3)$;
- 4) формирование матрицы W_{y1} по (14) вместе с её нормировкой: $O(2M^3 + M^2)$;
- 5) вычисление матрицы проекции W_{x2} (решение задачи на СЗ): $O(N^3)$;
- 6) формирование матрицы W_{y2} по (14) вместе с её нормировкой: $O(2N^3 + N^2)$.

Тогда общие затраты составят не более

$$O(3K(M^2N + N^2M) + 5(M^3 + N^3) + M^2 + N^2).$$

Если положить, что $M = N$ и $M^3 \gg M^2$, то можно записать:

$$3K(M^3 + M^3) + 5(M^3 + M^3) = 6KM^3 + 10M^3. \quad (15)$$

Теперь, если $K \gg 10$, то основные вычислительные затраты приходятся на формирование шести матриц ковариаций. Их вычисление основано на перемножении матриц $X^{(k)}$ и $Y^{(k)}$ и суммировании результата $\forall k$. Заметим, что и в итерационных алгоритмах 2DCCA также формируются шесть матриц ковариаций по всему набору исходных данных и еще по параметру числа итераций. При этом каждая матрица ковариации здесь основана на перемножении 4 числовых матриц, из которых две представляют исходные данные, а две другие – формируемые в процессе итераций левые и правые матрицы проекции. Получаемые при этом вычислительные затраты итерационных алгоритмов будут еще больше того, что показано в (15). Поэтому если исходными данными будут не яркостные признаки исходных изображений, а результат какого-либо их преобразования (градиентного, спектрального, по главным компонентам (через PCA, например)) так, что $M \rightarrow m$ и $N \rightarrow n$, и при этом $m \ll M$, а $n \ll N$, то затраты будут существенно снижены. Именно это (уменьшение размеров исходных изображений, переход к другому пространству признаков и т.д.) мы и видим при использовании итерационных алгоритмов 2DCCA в приложении к ИЛ. Но надо помнить, что такое преобразование ИЛ не всегда возможно и не во всех сценариях допустимо. А неитерационные алгоритмы 2DPCA/2DKLT и 2DPLS/2DKLT ра-

ботаю как при холистическом представлении ИЛ, так и их предварительном преобразовании в промежуточное подпространство признаков. А простые вычисления, реализуемые в них, проще и программировать, и выполнять с аппаратной поддержкой!

В [53] показаны параллельная и каскадная форма реализации алгоритмов 2DCCA/2DKLT и 2DPLS/2DKLT. Структура параллельного алгоритма соответствует описанию псевдокода, приведенному выше. Структура каскадного алгоритма состоит из двух каскадов, при этом первый каскад реализует алгоритм CCA (или PLS) относительно строк исходных данных $X(k)$ и $Y(k)$, а второй каскад – относительно столбцов результата, полученного в первом каскаде.

8. Применение PLS и CCA для наборов ИЛ

Наши собственные результаты применения 2DCCA/2DKLT и 2DPLS/2DKLT в задачах обработки ИЛ приведены в статьях [37–39]. В этих работах представлены следующие результаты: сравнительный анализ решения задач распознавания ИЛ (различной сенсорной и гендерной природы); взаимная индексация и распознавание изображений, не связанных семантически; особенности параллельной и каскадной схем реализации 2DCCA/2DKLT и 2DPLS/2DKLT с оценкой их характеристик; исследование систем распознавания ИЛ с различными вариантами фьюзии признаков в собственном признаковом подпространстве и различными типами классификаторов.

Представленные в этих работах эксперименты выполнены на широко известных в практике обработки цифровых изображений бенчмарковых базах: «FERET», «Equinox», «Люди и собаки», «Photo-Sketch CUNH» [24, 40, 41].

Если, например, пара ИЛ содержала гендерно не связанные между собой лица, а пары этих ИЛ подбирались случайным образом, корреляция между ними в исходном пространстве признаков была близка к нулю. В базе «Equinox» связанные в пары изображения включали лица, полученные в видимом свете и инфракрасном – тепловом излучении (VIS и Th). Между этими изображениями имелись существенные различия по текстуре и размеру компонентов лица, что также сводило корреляцию между ними до нуля. База «Люди и Собаки» (“People and Dogs” [40]) представляет случай, когда исходные изображения не относятся к одному глобальному классу. При этом изображения в классах имели некоторое внешнее сходство (между ИЛ – «хозяином» и его собакой), которое определяется следующими факторами: близким по форме выражением лица хозяина и «морды» собаки; одинаковым ракурсом двух портретов; близкой по форме прически хозяина и экстерьером собаки; цветовой гаммой – одинаковый цвет волос хозяина и окраски шерсти собак, а также текстуры.

В рамках этих исходных данных мы ставили следующие цели: проверить, насколько субъективная оценка подобия изображений в парах (и наборов в целом) сравнима с формальными оценками их подобия;

какие методы формальной оценки подобия можно считать достоверными; какова мера подобия этих изображений в пространстве канонических переменных.

Во всех экспериментах было показано, что в пространстве канонических переменных между переменными U и V возникает устойчивая корреляционная связь, которую можно описать линейной регрессией. При этом корреляция позволяет решать задачи взаимного распознавания, индексации и замещения переменных, а регрессия – задачи взаимной реконструкции переменных как в собственном подпространстве, так и в исходном пространстве на основе обратного преобразования KLT.

Из других решений и приложений 2DCCA и 2DPLS можно отметить работы [42–44].

В статье [42] представлен алгоритм решения задач *суперразрешения ИЛ* – реконструкции ИЛ с высоким разрешением (High-Resolution/HR) по ИЛ с низким разрешением (Low-Resolution/LR). При этом обучение реализовано в рамках 2DCCA на наборах изображений X с высоким (HR) и Y с низким разрешением (LR), а реконструкция LR→HR реализована в собственном подпространстве.

Однако 2DCCA здесь был реализован по итерационному алгоритму [32], что требовало ограничивать размеры исходных данных и дополнительно требовало декомпозиции исходных изображений на три части. В итоге, итерационный алгоритм обучения в рамках 2DCCA, реконструкция LR→HR, сшивка результата HR и компенсация высокочастотных составляющих выполнялись здесь три раза по наборам декомпозированных частей изображений. Тем не менее, подход [42] для задач суперразрешения ИЛ представляет практический интерес, как методически проработанный и достаточно ясно представленный в статье [42] поэтому интересно было бы реализовать идеи [42] также и в рамках неитерационных алгоритмов 2DCCA\2DKLT, не требующих уменьшения размеров ИЛ и их декомпозиции на части.

В работе [43] представлена задача *Face Hallucination*, аналогичная по своей сути задаче суперразрешения в приложении к ИЛ. Основное отличие идей [43] состоит в том, что обучающая выборка в [43] представляет собой комбинацию ИЛ с низким и высоким разрешением. Это сокращает объем вычислений на этапе обучения и дает возможность использовать (на этапе обучения и реконструкции) вместо процедур 2DCCA процедуры 2DPCA.

В работе [43] представлено решение задач распознавания изображений отпечатков ладони (Palmprint) и ИЛ базы FERET. Авторы работы оценивают вычислительную сложность итерационного подхода к реализации 2DPLS в сравнении с неитерационным подходом к реализации 2DPLS (со ссылкой на свою же работу 2005 года). При этом показывается превосходство итерационного подхода в случаях предварительной трансформации исходных изображений в набор новых признаков, размерность которых меньше исходного пространства яркостных признаков изобра-

жений. Экспериментально доказывается также превосходство (по точности и сложности вычислений) подхода 2DPLS в сравнении с 1DPLS.

Следует отметить, что алгоритмы 2DPLS развивались независимо от методов 2DCCA, часто опережая своими решениями идеи реализации 2DCCA. Одним из «примеров такого опережения» является независимая от [32, 33] ветвь реализации PLS и CCA на основе SVD (смотри, например, ранние работы Magnus'a Borga и David'a Weenink).

9. Применение методов проекции в подпространства в многомерной биологии и многомерной медицине

Одним из революционных достижений XXI века является появление многомерной биологии [45] (high dimensional biology). Ее направления для краткости называются OMICS (genomics, transcriptomics и т. д.). Современные высокопроизводительные технологии «омики» позволяют эффективно генерировать большие экспериментальные наборы данных для задач биологии и медицины. По факту эти данные представляются как многомерные данные (2D, 3D и большей размерности). Из этих многомерных исходных данных необходимо извлекать новые данные, содержащие основную изучаемую информацию, что достижимо, например, в рамках методов редукции исходного пространства данных. В работе [46] изучаются методы сокращения избыточности исходных данных, а также подходы к их интеграции и выявления и/или установления связей в этих данных. В этой связи в статье обсуждаются все рассмотренные нами методы проекции в подпространства (PCA, LDA, CCA и PLS), а также оцениваются их характеристики в приложении к многомерным данным, обрабатываемым в многомерной биологии и медицине.

10. Применение методов проекции в подпространства в нейронных сетях

На границе XX и XXI веков в машинном обучении и компьютерном зрении произошла методологическая (и технологическая) революция, начатая сверточными нейронными сетями (Convolutional Neural Network – CNN). В задачах обработки изображений (классификация, распознавание, кодирование и реконструкция, и т.д.) CNN работают с оригинальными (необработанными) данными, поэтому каждый пиксель исходного изображения должен быть связан с независимым входным каналом CNN. С ростом размеров исходных изображений и ростом числа обучающей выборки (доходящей до нескольких миллионов изображений) обработка такого объема данных требует большого числа слоев CNN. И это связано, с одной стороны, с избытком ненужной информации в исходных изображениях, а с другой стороны – с переизбытком подобной же информации в обучающих данных. При очень большом числе слоев такие сети называются глубокими нейронными сетями (Deep NN). Так, например, в 2016 году число слоев Deep NN перевалило за 150! При этом чем больше слоев со-

держит Deep NN, тем дольше и труднее ее обучить. Также при большом числе слоев можно попасть и в переобучение (!), когда сеть теряет способность к обобщению и не распознает скрытые процессы и важную информацию в обучающей выборке или, наоборот, принимает шум в исходных данных за важную информацию.

Именно поэтому перед исследователями и пользователями Deep NN встала задача уменьшения объема исходных данных и объема обучающей выборки при условии сохранения важной информации в них. Этого можно достичь, например, сокращением размерности пространства признаков в исходных изображениях и/или предварительным очищением их от шума, представлением исходных изображений только по главным компонентам. Уменьшение объема обучающих данных можно достичь, например, индексированием их в обучающей выборке, предварительной ее классификацией и также представлением данных по главным компонентам. А эти задачи хорошо решаются в рамках методов проекций изображений в собственные подпространства (ПИСП), что было показано выше.

Далее будут представлены несколько публикаций последних лет [47–52], в которых показано использование методов ПИСП в Deep NN.

В одной из первых обзорных работ [47] рассмотрены различные варианты NN, построенные с включением в них линейных и нелинейных процедур обработки на основе одномерных и двумерных методов SVD, PCA, LDA и CCA. Представлены также алгоритмы их обучения для этих вариантов сетей. Авторы отмечают, что такие NN легко реализуемы и эффективно используются в решении задач адаптивной обработки изображений, слепого разделения данных (полезной информации от шума), распознавания образов и их сжатии. А введение процедур SVD, PCA, LDA и CCA в NN сокращает объем информации на их входе, оставляя только полезную информацию.

В статье [48] предложена очень простая процедура обучения Deep NN для классификации изображений, основанная на методе ПИСП/PCA и получившая название PCANet. При этом PCANet включает следующие базовые компоненты обработки: банк фильтров, реализованный на каскаде процедур PCA; процедуру бинарного хеширования; вычисление блочных гистограмм. Отмечается, что комбинирование методов ПИСП с методами бинарного хеширования, гистограммного представления изображений и пирамидального объединения промежуточного результата в слоях позволяет упростить и ускорить обучение Deep NN, а также уменьшить число слоев в них.

В статье [49] предлагается вариант Deep NN, названной авторами как «Multiple scales PCANet», предназначенной для распознавания ИЛ. Сеть использует глубокое обучение на нескольких шкалах (multiple scales) из наборов представлений ИЛ, реализованных с помощью PCA в первых двух слоях сверточной нейронной сети. Фактически на этом уровне

формируется *популяция новых представлений* исходных данных, что повышает их репрезентативность и приводит к улучшению задачи распознавания. При построении и обучении сети авторы исходят из предположения, что чем выше уровень признакового пространства (чем больше популяция), тем более точно и полно он может представлять семантику исходных данных. Остальная часть NN строится аналогично рассмотренной в работе [48]. Статья сопровождается уникальными экспериментами моделирования работы «Multiple scales PCANet», выполненными на нескольких новых бенчмарковых базах ИЛ, и в том числе на базах ИЛ (2D и 3D), полученных в неконтролируемых условиях съемки и различных сценариях. Полученные результаты сравниваются с известными по этой теме. В заключении отмечается, что «Multiple scales PCANet» может быть применена и в других задачах компьютерного зрения – распознавания эмоций и трекинга и также покажет свою высокую эффективность.

В статье [50] предлагается комбинированная архитектура Deep NN, в которой архитектура PCANet [48] объединена с возможностями идей PLS-регрессии, что дает новую методику классификации изображений, которую авторы назвали «методом PLSNet». Здесь в первых слоях NN редукция размерности пространства признаков реализуется по архитектуре PCANet, а в следующих слоях – в рамках PLS-регрессии. Развивая эту идею дальше, авторы применяют новые фильтры для экстракции признаков в следующих слоях NN и получают новую архитектуру Deep NN, названную ими как «Improved PLSNet». При этом достигается более высокая точность классификации изображений, чем в PCANet. Эксперименты выполнены на наборах данных MNIST (цифры, написанные от руки) и CIFAR-10 (наборы изображений).

В предыдущих решениях основу Deep NN создают процедуры редукции пространства признаков, используемые на входе – в первых слоях Deep NN. В статьях [51, 52] рассматриваются новые архитектуры Deep NN, названные как «Deep Canonical Correlation Analysis» и «Deep generalized canonical correlation analysis». Здесь каждый набор исходных данных (два набора в архитектуре [51] и несколько наборов в архитектуре [52]) поступает на вход «собственной CNN», в которой реализуется нелинейное преобразование исходных данных без выполнения каких-либо операций умножения. При этом улучшается репрезентативность этих данных, что свойственно каждой CNN. Выходные слои «собственной CNN» подключаются к блоку CCA, где обрабатываются совместно, чем достигается высокая корреляция между ними. Блоки CCA имеют обратную связь со слоями каждой «собственной CNN», которая, очевидно, используется в процессе обучения.

Таким образом, можно констатировать, что идеи проекции цифровых изображений в собственные подпространства вышли на уровень архитектур CNN, включая Deep NN.

Заключение

В статье представлена история появления методов проекции в собственные подпространства (МПСП) как инструмента обработки экспериментальных данных и инструмента выявления и/или установления связей в этих данных. Детально описан метод PCA, основанный на идеях «eigenfaces», и через него представлены методы LDA, PLS и CCA. Показаны особенности и проблемы применения МПСП в приложении к обработке изображений лиц. В рамках анализа публикаций исследованы этапы развития новых подходов к МПСП, приведших к появлению идей и методов двумерной проекции, ориентированных на обработку изображений как двумерных объектов. В технической литературе эти методы представлены как вариации от сокращений 2DPCA, 2DLDA, 2DPLS и 2DCCA. Рассмотрены алгоритмы практической реализации двумерных МПСП, и выявлены четыре основные ветви их развития: две ветви итерационных алгоритмов – в двух независимых вариантах и две ветви неитерационных алгоритмов – на основе прямого решения задач на собственные значения (ЗСЗ) и решения ЗСЗ с использованием процедур SVD. Оценены их достоинства и недостатки, а также показаны примеры их использования в практике распознавания ИЛ. Отмечено также, что методы проекции в подпространства широко используются в «многомерной биологии и многомерной медицине», а для цифровых изображений МПСП вышли на уровень архитектур CNN, включая Deep NN. Показаны примеры соответствующих решений, и дан их краткий анализ.

Особое внимание в статье уделено группе методов, в которых двумерная проекция в собственные подпространства определена как двухэтапная процедура: 2DPCA/2DKLT, 2DLDA/2DKLT, 2DPLS/2DKLT и 2DCCA/2DKLT. На первом этапе реализуется анализ исходных данных, включающий: формирование общих матриц рассеяния в соответствии с заданным критерием; решение задач на собственные значения с вычислением матриц проекции; определение главных компонент. На втором этапе реализуется полное (или усеченное только по главным компонентам) **двумерное преобразование Карунена–Лоэва**. Такая запись более точно отражает все процессы, реализуемые в рамках двумерных проекционных методов, не смешивая их и не подменяя одни процессы и понятия другими (что часто можно встретить в технической литературе). И эта же запись выделяет ветвь методов двумерной проекции с прямым (неитерационным) решением задач на собственные значения от трех других ветвей. Наконец, такая запись позволяет легко представить (скомпоновать, реализовать) обобщенный алгоритм (generalized algorithm) для проблемы двумерных МПСП. В этом случае по исходным данным и в соответствии с выбранным критерием формируется каноническая ковариационная матрица, блоки которой используются для построения матриц рассеяния в отдельных методах, а в процедуре 2DKLT изменяются только матрицы проек-

ции, поскольку ее реализация является общей для всех методов.

Дальнейшие исследования будут связаны с реализацией алгоритмов взаимной реконструкции изображений лиц на основе двумерных методов проекции в собственные подпространства.

Литература

1. **Pearson, K.** On lines and planes of closest fit to systems of points in space / K. Pearson // The London, Edinburgh and Dublin Philosophical Magazine and Journal of Sciences. – 1901. – Vol. 6, Issue 2. – P. 559-572. – DOI: 10.1080/14786440109462720.
2. **Hotelling, H.** Analysis of complex variables into principal components / H. Hotelling // Journal of Educational Psychology. – 1933. – Vol. 24, Issue 6. – P. 417-441. – DOI: 10.1037/h0071325.
3. **Fisher, R.A.** The use of multiple measurements in taxonomic problems / R.A. Fisher // Annals of Eugenics. – 1936. – Vol. 7, Issue 2. – P. 179-188. – DOI: 10.1111/j.1469-1809.1936.tb02137.x.
4. **Hotelling, H.** Relations between two sets of variates / H. Hotelling // Biometrika. – 1936. – Vol. 28, No. 3/4. – P. 321-377. – DOI: 10.2307/2333955.
5. **Sirovich, L.** Low-dimensional procedure for the characterization of human faces / L. Sirovich, M. Kirby // Journal of the Optical Society of America A: Optics, Image Science and Vision. – 1987. – Vol. 4, Issue 3. – P. 519-524. – DOI: 10.1364/JOSAA.4.000519.
6. **Turk, M.** Eigenfaces for recognition / M. Turk, A. Pentland // Journal of Cognitive Neuroscience. – 1991. – Vol. 3, Issue 1. – P. 71-86. – DOI: 10.1162/jocn.1991.3.1.71.
7. **Belhumeur, P.N.** Eigenfaces vs. fisherfaces: Recognition using class specific linear projection / P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 1997. – Vol. 19, Issue 7. – P. 711-720. – DOI: 10.1109/34.598228.
8. **Tsapatoulis, N.** A vector based approximation of KLT and its application to face recognition / N. Tsapatoulis, V. Alexopoulos, S. Kollias // Proceedings of the IX European Signal Processing Conference (EUSIPCO-98). – 1998. – Vol. III. – P. 1581-1584. – DOI: 10.5281/zenodo.36612.
9. **Кухарев, Г.А.** Биометрические системы: Методы и средства идентификации личности человека / Г.А. Кухарев. – СПб.: Политехника, 2001. – 240 с. – ISBN: 5-7325-0623-3.
10. **Kukharev, A.** Techniki Biometryczne: Część 1. Metody Rozpoznawania Twarzy / G. Kukharev, A. Kuźmiński. – Szczecin: Pracownia Poligraficzna WI PS, 2003. – 310 p.
11. **Kukharev, G.** Data dimensionality reduction for face recognition / G. Kukharev, P. Forczmański // Machine GRAPHICS & VISION. – 2004. – Vol. 13, No. 1/2. – P. 99-121.
12. **Kukharev, G.** Face recognition by means of two-dimensional direct linear discriminant analysis / G. Kukharev, P. Forczmański // Proceedings of the 8th International Conference Pattern Recognition and Information Processing (PRIP'2005). – P. 280-283.
13. **Кухарев, Г.А.** Системы распознавания человека по изображению лица / Г.А. Кухарев, Н.Л. Щеголева. – СПб.: Изд-во СПбГЭТУ «ЛЭТИ», 2006. – 156 с. – ISBN: 5-7629-0665-5.
14. **Kukharev, G.** System of face recognition using LDA with one training image per person / G. Kukharev, A. Tujaka, N. Binh // Metody Informatyki Stosowanej. – 2008. – No. 3(16). – P. 167-185.
15. **Yang, J.** Two-dimensional PCA: A new approach to appearance-based face representation and recognition / J. Yang, D. Zhang, A.F. Frangi, J.-Y. Yang // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2004. – Vol. 26, Issue 1. – P. 131-137. – DOI: 10.1109/TPAMI.2004.1261097.
16. **Zhang, D.** (2D)2PCA: Two-directional two-dimensional PCA for efficient face representation and recognition / D. Zhang, Z.H. Zhou // Neurocomputing. – 2005. – Vol. 69, Issues 1-3. – P. 224-231. – DOI: 10.1016/j.neucom.2005.06.004.
17. **Ye, J.** Generalized low rank approximations of matrices / J. Ye // Machine Learning. – 2005. – Vol. 61, Issues 1-3. – P. 167-191. – DOI: 10.1007/s10994-005-3561-6.
18. **Kong, H.** Generalized 2D principal component analysis for face image representation and recognition / H. Kong, L. Wang, E.K. Teoh, X. Li, J.-G. Wang, R. Venkateswarlu // Neural Networks. – 2005. – Vol. 18, Issues 5-6. – P. 585-594. – DOI: 10.1016/j.neunet.2005.06.041.
19. **Ding, Ch.** Two-dimensional singular value decomposition (2DSVD) for 2D maps and images / Ch.H.Q. Ding, J. Ye // Proceedings of the 2005 SIAM International Conference on Data Mining. – 2005. – DOI: 10.1137/1.9781611972757.4.
20. **Gu, Zh.** Two-dimensional singular value decomposition (2D-SVD) based video coding / Zh. Gu, W. Lin, B.-S. Lee, Ch.T. Lau, M. Paul // 2010 IEEE International Conference on Image Processing. – 2010. – P. 181-184. – DOI: 10.1109/ICIP.2010.5650998.
21. **Gurumoorthy, K.S.** A method for compact image representation using sparse matrix and tensor projections onto exemplar orthonormal bases / K.S. Gurumoorthy, A. Rajwade, A. Banerjee, A. Rangarajan // IEEE Transactions on Image Processing. – 2010. – Vol. 19, Issue 2. – P. 322-334. – DOI: 10.1109/TIP.2009.2034991.
22. **Inoue, K.** Equivalence of non-iterative algorithms for simultaneous low rank approximations of matrices / K. Inoue, K. Urahama // 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). – 2006. – DOI: 10.1109/CVPR.2006.112.
23. **Tang, X.** Face sketch recognition / X. Tang, X. Wang // IEEE Transactions on Circuits and Systems for Video Technology. – 2004. – Vol. 14, Issue 1. – P. 50-57. – DOI: 10.1109/TCSVT.2003.818353.
24. **CUHK Face Sketch Database (CUFS)** [Electronical Resource]. – URL: <http://mmlab.ie.cuhk.edu.hk/archive/facesketch.html> (access date: 02.07.2018).
25. **Kukharev, G.** Face photo-sketch transformation and population generation / G. Kukharev, A. Oleinik. – In: International conference on computer vision and graphics. ICCVG 2016: Computer vision and graphics // ed. by L. Chmielewski, A. Datta, R. Kozera, K. Wojciechowski. – 2016. – P. 329-340. – DOI: 10.1007/978-3-319-46418-3_29.
26. **Borga, M.** Learning multidimensional signal processing. – Linköping, Sweden: Linköpings Universitet, 1998. – 193 p. – ISBN: 91-7219-202-X.
27. **Reiter, M.** 3D and infrared face reconstruction from rgb data using canonical correlation analysis / M. Reiter, R. Donner, G. Langs, H. Bischof // Proceedings of the 18th International Conference on Pattern Recognition (ICPR 2006). – 2006. – Vol. 1. – P. 425-428. – DOI: 10.1109/ICPR.2006.24.
28. **Reiter, M.** Estimation of face depth maps from color textures using canonical correlation analysis / M. Reiter, R. Donner. – In: Proceedings of the Computer Vision Winter Workshop 2006 (CWW' 06) / ed. by O. Chum, V. Franc. – Telč: Czech Society for Cybernetics and Informatics, 2006. – ISBN: 80-239-6530-1.

29. **Donner, R.** Fast active appearance model search using canonical correlation analysis / R. Donner, M. Reiter, G. Langs, P. Peloschek, H. Bischof // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2006. – Vol. 28, Issue 10. – P. 1960-1964. – DOI: 10.1109/TPAMI.2006.206.
30. **Alonso, J.** Face tracking using canonical correlation analysis / J. Alonso Y. Zepeda, F. Davoine, M. Charbit // Proceedings of the 2nd International Conference on Computer Vision Theory and Applications (VISAPP 2007). – 2007. – Vol. 2. – P. 396-402.
31. **Sharma, A.** Bypassing synthesis: PLS for face recognition with pose, low-resolution and sketch / A. Sharma, D.W. Jacobs // IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). – 2011. – P. 593-600. – DOI: 10.1109/CVPR.2011.5995350.
32. **Lee, S.H.** Two-dimensional canonical correlation analysis / S.H. Lee, S. Choi // IEEE Signal Processing Letters. – 2007. – Vol. 14, Issue 10. – P. 735-738. – DOI: 10.1109/LSP.2007.896438.
33. **Zou, C.-R.** 2DCCA: A novel method for small sample size face recognition / C.-R. Zou, N. Sun, Zh.-H. Ji, Zh. Li // IEEE Workshop on Application of Computer Vision (WACV'07). – 2007. – DOI: 10.1109/WACV.2007.1.
34. **Shao, M.** Joint features for face recognition under variable illuminations / M. Shao, Y. Wang // Fifth International Conference on Image and Graphics. – 2009. – P. 922-927. – DOI: 10.1109/ICIG.2009.128.
35. **Gong, X.** Application to three-dimensional canonical correlation analysis for feature fusion in image recognition / X. Gong // Journal of Computers. – 2011. – Vol. 6, Issue 11. – P. 2427-2433.
36. **Kamencay, P.** 2D-3D face recognition method based on a modified CCA-PCA algorithm / P. Kamencay, R. Hudec, M. Benco, M. Zachariasova // International Journal of Advanced Robotic Systems. – 2014. – Vol. 11, Issue 3. – 9 p. – DOI: 10.5772/58251.
37. **Kukharev, G.** Two-dimensional canonical correlation analysis for face image processing and recognition / G. Kukharev, E. Kamenskaya // Metody Informatyki Stosowanej. – 2009. – No. 3(20). – P. 103-112.
38. **Kukharev, G.** Face recognition using two-dimensional CCA and PLS / G. Kukharev, A. Tujaka, P. Forczmański // International Journal of Biometrics. – 2011. – Vol. 3, Issue 4. – P. 300-321. – DOI: 10.1504/IJBM.2011.042814.
39. **Кухарев, Г.** Методы обработки и распознавания изображений лиц в задачах биометрии / Г. Кухарев, Е. Каменская, Ю. Матвеев, Н. Щеголева; под ред. М.В. Хитрова. – СПб: Политехника, 2013. – 388 с. – ISBN: 978-5-73251-028-7.
40. Почему собаки похожи на хозяев [Электронный ресурс]. – URL: http://www.house-dog.ru/about_391.html (дата обращения 02.07.2018).
41. **Gupta, S.** Texas 3D face recognition database [Electronical Resource] / S. Gupta, M.K. Markey, K.R. Castleman, A.C. Bovik. – URL: <http://live.ece.utexas.edu/research/texas3dfr/index.htm> (access date: 23.04.2017).
42. **Tu, Ch.-T.** A new approach for face hallucination based on a two-dimensional direct combined model / Ch.-T. Tu, M.-Ch. Ho, M.-Y. Lin // Pattern Recognition. – 2017. – Vol. 62. – P. 1-20. – DOI: 10.1016/j.patcog.2016.07.020.
43. **An, L.** Face image super-resolution using 2D CCA / L. An, B. Bhanu // Signal Processing. – 2014. – Vol. 103. – P. 184-194. – DOI: 10.1016/j.sigpro.2013.10.004.
44. **Hou, Sh.** A two-dimensional partial least squares with application to biological image recognition / Sh. Hou, Q. Sun, D. Xia // 2010 Sixth International Conference on Natural Computation (ICNC 2010). – 2010. – P. 57-61. – DOI: 10.1109/ICNC.2010.5583135.
45. **Вельков, В.В.** Многомерная биология и многомерная медицина / В.В. Вельков // Химия и Жизнь. – 2007. – № 3. – С. 10-15.
46. **Meng, C.** Dimension reduction techniques for the integrative analysis of multi-omics data / C. Meng, O.A. Zeleznik, G.G. Thallinger, B. Kuster, A.M. Gholami, A.C. Culhane // Briefings in Bioinformatics. – 2016. – Vol. 17, Issue 4. – P. 628-641. – DOI: 10.1093/bib/bbv108.
47. **Qiu, J.** Neural network implementations for PCA and its extensions / J. Qiu, H. Wang, J. Lu, B. Zhang, K.-L. Du // ISRN Artificial Intelligence. – 2012. – Vol. 2012. – 847305. – DOI: 10.5402/2012/847305.
48. **Chan, T.-H.** PCANet: A simple deep learning baseline for image classification? / T.-H. Chan, K. Jia, Sh. Gao, J. Lu, Z. Zeng, Y. Ma // IEEE Transactions on Image Processing. – 2015. – Vol. 24, Issue 12. – P. 5017-5032. – DOI: 10.1109/TIP.2015.2475625.
49. **Tian, L.** Multiple scales combined principle component analysis deep learning network for face recognition / L. Tian, Ch. Fan, Y. Ming // Journal of Electronic Imaging. – 2016. – Vol. 25, Issue 2. – 023025. – DOI: 10.1115/1.JEI.25.2.023025.
50. **Hasegawa, R.** PLSNet: A simple network using Partial Least Squares regression for image classification / R. Hasegawa, K. Hotta // Proceedings of the 23rd International Conference on Pattern Recognition (ICPR). – 2016. – P. 1601-1606. – DOI: 10.1109/ICPR.2016.7899865.
51. **Andrew, G.** Deep canonical correlation analysis / G. Andrew, R. Arora, J. Balmes, K. Livescu // Proceedings of the 30th International Conference on Machine Learning: Proceedings of Machine Learning Research. – 2013. – Vol. 28(3). – P. 1247-1255.
52. **Benton, A.** Deep generalized canonical correlation analysis [Electronical Resource] / A. Benton, H. Khayrallah, B. Gujral, D. Reisinger, Sh. Zhang, R. Arora. – URL: [arXiv:1502.02519v2](https://arxiv.org/abs/1502.02519v2) (request date: 02.07.2018).
53. **Kukharev, G.A.** Algorithms of two-dimensional projection of digital images in eigensubspace: History of development, implementation and application / G.A. Kukharev, N.L. Shchegoleva // Pattern Recognition and Image Analysis. – 2018. – Vol. 28, Issue 2. – P. 185-206. – DOI: 10.1134/S1054661818020116.

Сведения об авторах

Кухарев Георгий Александрович, профессор кафедры математического обеспечения СПбГЭТУ (ЛЭТИ) и профессор на факультете информатики Щецинского технологического университета в Польше (Szczecin, West Pomeranian University of Technology, Poland). В 2001–2003 годах был приглашённым профессором на кафедре информатики и математики в Ecole Centrale de Lyon во Франции. В 2005–2006 годах был приглашённым профессором в Ханойском университете (Вьетнам). Является автором и соавтором более 100 научных работ, в том числе 44 авторских свидетельств на изобретения и 10 монографий, посвящённых методам и техническим средствам цифровой обработки сигналов и изображений. E-mail: gkukharev@wi.zut.edu.pl.

Щеголева Надежда Львовна, доктор технических наук, доцент кафедры математического обеспечения ЭВМ Санкт-Петербургского государственного электротехнического университета (ЛЭТИ). Является автором и соавтором более 70 научных работ, соавтором 4 патентов, двух монографий, посвящённых методам и цифровой обработке сигналов и изображений. Научные интересы лежат в области биометрии, в том числе: обнаружение и распознавание изображений лиц, биометрические системы идентификации и контроля доступа, построение и моделирование систем распознавания. E-mail: NLShchegoleva@etu.ru.

ГРПТИ: 28.23.15

Поступила в редакцию 30 ноября 2017 г. Окончательный вариант – 17 июля 2018 г.

METHODS OF TWO-DIMENSIONAL PROJECTION OF DIGITAL IMAGES INTO EIGEN-SUBSPACES: PECULIARITIES OF IMPLEMENTATION AND APPLICATION

G.A. Kukharev¹, N.L. Shchegoleva¹

¹ Electrotechnical University "LETI" (ETU), Saint-Petersburg, Russia

Abstract

The history of development of algorithms for projection of digital images into their eigensubspaces using linear methods based on PCA (principal component analysis), LDA (linear discriminant analysis), PLS (partial least squares), and CCA (canonical correlation analysis) is considered. We show that the emergence of new application areas has changed the requirements for the methods, with major changes involving (i) the use of PCA, LDA, PLS and CCA methods for both small and extremely large face image (FI) samples in the initial sets; (ii) a criterion for determining the eigen-basis, which also should provide the solution of a particular problem (the minimum error of face image approximation, etc.); (iii) the applicability of the methods under consideration to the processing of two or more image sets from different sensors or several sets of any number of matrices; and (iv) the possibility of realizing two-dimensional projections of face images (or other numerical matrices) directly into the layers of convolutional neural networks (NN) and/or integrating their functions into the NN as separate blocks. Estimates of the computational complexity and examples of solving image processing problems are also given.

Keywords: face image (FI) sets and numeric matrices, an eigen-basis and eigensubspaces, principal components analysis (PCA), linear discriminant analysis (LDA), partial least squares (PLS), canonical correlation analysis (CCA), Karhunen-Loeve transformation (KLT), 2DPCA/2DKLT, 2DPLS/2DKLT, 2DCCA/2DKLT, CNN, Deep NN.

Citation: Kukharev GA, Shchegoleva NL. Methods of two-dimensional projection of digital images into eigen-subspaces: peculiarities of implementation and application. Computer Optics 2018; 42(4): 637-656. DOI: 10.18287/2412-6159-2018-42-4-637-656.

References

- [1] Pearson K. On lines and planes of closest fit to systems of points in space. The London, Edinburgh and Dublin Philosophical Magazine and Journal of Sciences 1901; 6(2): 559-572. DOI: 10.1080/14786440109462720.
- [2] Hotelling H. Analysis of complex variables into principal components. Journal of Educational Psychology 1933; 24(6): 415-441. DOI: 10.1037/h0071325.
- [3] Fisher RA. The use of multiple measurements in taxonomic problems. Annals of Eugenics 1936; 7(2): 159-188. DOI: 10.1111/j.1469-1809.1936.tb02137.x.
- [4] Hotelling H. Relations between two sets of variates. Biometrika 1936; 28(3/4): 321-377. DOI: 10.2307/2333955.
- [5] Sirovich L, Kirby M. Low-dimensional procedure for the characterization of human faces. J Opt Soc Am A 1987; 4(3): 519-524. DOI: 10.1364/JOSAA.4.000519.
- [6] Turk M, Pentland A. Eigenfaces for recognition. Journal of Cognitive Neuroscience 1991; 3(1): 71-86. DOI: 10.1162/jocn.1991.3.1.71.
- [7] Belhumeur PN, Hespanha JP, Kriegman DJ. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence 1997; 19(7): 711-720. DOI: 10.1109/34.598228.
- [8] Tsapatsoulis N, Alexopoulos V, Kollias S. A vector based approximation of KLT and its application to face recognition. Proc EUSIPCO-98 1998; III: 1581-1584. DOI: 10.5281/zenodo.36612.
- [9] Kukharev G. Biometric systems: Methods and means for people recognition [In Russian]. Sankt-Petersburg: "Politehnika" Publisher; 2001. ISBN: 5-7325-0623-3.
- [10] Kukharev G, Kuźmiński A. Techniki biometryczne: Część 1. Metody rozpoznawania twarzy. Szczecin: „Pracownia Poligraficzna WI PS” Publisher, 2003.
- [11] Kukharev G, Forczmański P. Data dimensionality reduction for face recognition. Machine GRAPHICS & VISION 2004; 13(1/2): 99-121.
- [12] Kukharev G, Forczmański P. Face recognition by means of two-dimensional direct linear discriminant analysis. Proc PRIP'2005: 280-283.
- [13] Kukharev GA, Shchegoleva NL. Human face recognition systems [In Russian]. Saint-Petersburg: Saint-Petersburg Electrotechnical University "LETI" Publisher; 2006. ISBN: 5-7629-0665-5.
- [14] Kukharev G, Tujaka A, Binh N. System of face recognition using LDA with one training image per person. Metody Informatyki Stosowanej 2008; 3(16): 167-185.
- [15] Yang J, Zhang D, Frangi AF, Yang J-Y. Two-dimensional PCA: A new approach to appearance-based face represen-

- tation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2004; 26(1): 131-137. DOI: 10.1109/TPAMI.2004.1261097.
- [16] Zhang D, Zhou ZH. (2D)2PCA: Two-directional two-dimensional PCA for efficient face representation and recognition. *Neurocomputing* 2005; 69(1-3): 224-231. DOI: 10.1016/j.neucom.2005.06.004.
- [17] Ye J. Generalized low rank approximations of matrices. *Machine Learning* 2005; 61(1-3): 167-191. DOI: 10.1007/s10994-005-3561-6.
- [18] Kong H, Wang L, Teoh EK, Li X, Wang J-G, Venkateswarlu R. Generalized 2D principal component analysis for face image representation and recognition. *Neural Networks* 2005; 18(5-6): 585-594. DOI: 10.1016/j.neunet.2005.06.041.
- [19] Ding CHQ, Ye J. Two-dimensional singular value decomposition (2DSVD) for 2D maps and images. *Proc 2005 SIAM International Conference on Data Mining* 2005.
- [20] Gu Z, Lin W, Lee B-S, Lau CT, Paul M. Two-dimensional singular value decomposition (2D-SVD) based video coding. *IEEE International Conference on Image Processing* 2010: 181-184. DOI: 10.1109/ICIP.2010.5650998.
- [21] Gurumoorthy KS, Rajwade A, Banerjee A, Rangarajan A. A method for compact image representation using sparse matrix and tensor projections onto exemplar orthonormal bases // *IEEE Transactions on Image Processing* 2010; 19(2): 322-334. DOI: 10.1109/TIP.2009.2034991.
- [22] Inoue K, Urahama K. Equivalence of non-iterative algorithms for simultaneous low rank approximations of matrices. *CVPR'06* 2006. DOI: 10.1109/CVPR.2006.112.
- [23] Tang X, Wang X. Face sketch recognition. *IEEE Transactions on Circuits and Systems for Video Technology* 2004; 14(1): 50-57. DOI: 10.1109/TCSVT.2003.818353.
- [24] CUHK Face Sketch Database (CUFS). Source: <http://mmlab.ie.cuhk.edu.hk/archive/facesketch.html>.
- [25] Kukharev G, Oleinik A. Face photo-sketch transformation and population generation. In Book: Chmielewski L, Datta A, Kozera R, Wojciechowski K, eds. *Computer vision and graphics. ICCVG 2016: Computer vision and graphics* 2016: 329-340. DOI: 10.1007/978-3-319-46418-3_29.
- [26] Borga M. *Learning multidimensional signal processing*. Linköping, Sweden: Linköpings Universitet; 1998. ISBN: 91-7219-202-X.
- [27] Reiter M, Donner R, Langs G, Bischof H. 3D and infrared face reconstruction from rgb data using canonical correlation analysis. *Proc 18th Int Conf Patt Recogn (ICPR 2006)* 2006; 1: 425-428. DOI: 10.1109/ICPR.2006.24.
- [28] Reiter M, Donner R. Estimation of face depth maps from color textures using canonical correlation analysis. In Book: Chum O, Franc V, eds. *Proceedings of the Computer Vision Winter Workshop 2006 (CWW' 06)*. Telč: Czech Society for Cybernetics and Informatics; 2006. ISBN: 80-239-6530-1.
- [29] Donner R, Reiter M, Langs G, Peloschek P, Bischof H. Fast active appearance model search using canonical correlation analysis. *IEEE Trans Pattern Anal Mach Intell* 2006; 28(10): 1960-1964. DOI: 10.1109/TPAMI.2006.206.
- [30] Alonso J, Zepeda Y, Davoine F, Charbit M. Face tracking using canonical correlation analysis. *Proc VISAPP 2007* 2007; 2: 396-402.
- [31] Sharma A, Jacobs DW. Bypassing synthesis: PLS for face recognition with pose, low-resolution and sketch. *CVPR 2011*: 593-600. DOI: 10.1109/CVPR.2011.5995350.
- [32] Lee SH, Choi S. Two-dimensional canonical correlation analysis. *IEEE Signal Processing Letters* 2007; 14(10): 735-738. DOI: 10.1109/LSP.2007.896438.
- [33] Zou C-R, Sun N, Ji Z-H, Li Z. 2DCCA: a novel method for small sample size face recognition. *WACV'07* 2007. DOI: 10.1109/WACV.2007.1.
- [34] Shao M, Wang Y. Joint features for face recognition under variable illuminations. *Fifth International Conference on Image and Graphics* 2009: 922-927. DOI: 10.1109/ICIG.2009.128.
- [35] Gong X. Application to three-dimensional canonical correlation analysis for feature fusion in image recognition. *Journal of Computers* 2011; 6(11): 2427-2433.
- [36] Kamencay P, Hudec R, Benco M, Zachariasova M. 2D-3D face recognition method based on a modified CCA-PCA algorithm. *International Journal of Advanced Robotic Systems* 2014; 11(3). DOI: 10.5772/58251.
- [37] Kukharev G, Kamenskaya E. Two-dimensional canonical correlation analysis for face image processing and recognition. *Metody Informatyki Stosowanej* 2009; 3(21): 103-112.
- [38] Kukharev G, Tujaka A, Forczmański P. Face recognition using two-dimensional CCA and PLS. *International Journal of Biometrics* 2011; 3(4): 300-321. DOI: 10.1504/IJBM.2011.042814.
- [39] Kukharev G, Kamenskaya E, Matveev Y, Shchegoleva N. *Methods of facial images processing and recognition in biometrics [In Russian]*. Saint-Peterburg: "Politechnika" Publisher; 2013. ISBN: 978-5-73251-028-7.
- [40] Why dogs look like hosts [In Russian]. Source: http://www.house-dog.ru/about_391.html.
- [41] Gupta S, Markey MK, Castleman KR, Bovik AC. Texas 3D face recognition database. Source: <http://live.ece.utexas.edu/research/texas3dfr/index.htm>.
- [42] Tu C-T, Ho M-C, Lin M-Y. A new approach for face hallucination based on a two-dimensional direct combined model. *Patt Recogn* 2017; 62: 1-20. DOI: 10.1016/j.patcog.2016.07.020.
- [43] An L, Bhanu B. Face image super-resolution using 2D CCA. *Signal Processing* 2014; 103: 184-194. DOI: 10.1016/j.sigpro.2013.10.004.
- [44] Hou S, Sun Q, Xia D. A two-dimensional partial least squares with application to biological image recognition. *2010 Sixth International Conference on Natural Computation* 2010: 57-61. DOI: 10.1109/ICNC.2010.5583135.
- [45] Velkov VV. *Multidimensional biology and multidimensional medicine [In Russian]*. Chemistry and Life 2007; 3: 10-15.
- [46] Meng C, Zeleznik OA, Thallinger GG, Kuster B, Gholami AM, Culhane AC. Dimension reduction techniques for the integrative analysis of multi-omics data. *Briefings in Bioinformatics* 2016; 17(4): 628-641. DOI: 10.1093/bib/bbv108.
- [47] Qiu J, Wang H, Lu J, Zhang B, Du K-L. Neural network implementations for PCA and its extensions. *ISRN Artificial Intelligence* 2012; 2012: 847305. DOI: 10.5402/2012/847305.
- [48] Chan T-H, Jia K, Gao S, Lu J, Zeng Z, Ma Y. PCANet: A Simple Deep Learning Baseline for Image Classification? *IEEE Transactions on Image Processing* 2015; 24(12): 5017-5032. DOI: 10.1109/TIP.2015.2475625.
- [49] Tian L, Fan C, Ming Y. Multiple scales combined principle component analysis deep learning network for face recognition. *J Electron Imaging* 2016; 25(2): 023025. DOI: 10.1115/1.JEI.25.2.023025.
- [50] Hasegawa R, Hotta K. PLSNet: A simple network using Partial Least Squares regression for image classification.

- Proc 23rd ICPR 2016: 1601-1606. DOI: 10.1109/ICPR.2016.7899865.
- [51] Andrew G, Arora R, Bilmes J, Livescu K. Deep canonical correlation analysis. Proc 30th ICML: PMLR 2013; 28(3): 1247-1255.
- [52] Benton A, Khayrallah H, Gujral B, Reisinger D, Zhang Sh, Arora R. Deep generalized canonical correlation analysis. Source: <arXiv:1502.02519v2>.
- [53] Kukharev GA, Shchegoleva NL. Algorithms of two-dimensional projection of digital images in eigensubspace: History of development, implementation and application // Pattern Recognition and Image Analysis 2018; 28(2): 185-206. DOI: 10.1134/S1054661818020116.

Author's information

Georgy A. Kukharev Received Doctor of Technical Science degree (1986) from the Institute of Automatics and Computer Facilities (ABT, Riga, Latvia), Full Professor (2006). From 1993 he works at Szczecin University of Technology, Faculty of Computer Science & Information Systems (Poland). From 2003 he works at Saint Petersburg State Electrotechnical University LETI, Computer Software Environment department. In 2001–2003 visiting professor of Ecole Centrale de Lyon, Mathematics & Computer Science Department. Since 2005 visiting professor of Hanoi University of Technology, International Training Programmer department. Author of more than 10 monographs, over 100 scientific papers, and over 44 patents in the areas: computer architecture of signal processing, image processing, and pattern recognition. Current interests: biometrics, including face detection and face recognition, face/sketch retrieval, facial bar-code.

Nadegda L. Shchegoleva Doctor of Science. From 2001-2006 she was Senior Researcher in Federal State Unitary Enterprise Central Research Institute "Morphyspribor" (JSC "Concern" OKEANPRIBOR"). From 2007 she is Associate Professor in the Software Engineering and Compute Application department (Saint Petersburg Electrotechnical University "LETI", St.-Petersburg, Russia). Co-author of 2 monographs, 4 inventions, more than 80 scientific articles. Her research interests include biometric identification systems and access control systems, face recognition, synthesis and modeling of the recognition systems, face/sketch retrieval, facial bar-code.

Received November 30, 2017. The final version – July 17, 2018.