

*Рассмотрены несколько наиболее эффективных на сегодняшний день алгоритмов распознавания лиц. Показано, что производительность алгоритмов сильно зависит от природы поставленной задачи и архитектуры самих алгоритмов распознавания. Показано, что алгоритм ICA является наиболее эффективным и универсальным – архитектура II для задач глобального распознавания лиц, архитектура I – для задач локального распознавания (мимика, выражение лица, и т.д.).*

© В.В. Горин, 2011

УДК 004-931

В.В. ГОРИН

## СРАВНЕНИЕ АЛГОРИТМОВ АНАЛИЗА ИЗОБРАЖЕНИЙ В ЗАДАЧЕ РАСПОЗНАВАНИЯ ЛИЦ

Лицо человека является первостепенным объектом внимания при взаимоотношениях людей и играет ключевую роль при идентификации личности. Разработка вычислительной модели распознавания лиц – задача довольно трудная, так как лица являются естественно-природными, живыми объектами в отличие от «мира моноблоков» и других искусственно-созданных человеком объектов, изучаемых во многих работах по распознаванию образов. К тому же они сложные, многозначные трехмерные объекты. Ранее выполненные исследования показали, что индивидуальные черты и их взаимные отношения не достаточны для идентификации лиц взрослого человека [1].

В 1966 г. Бледсо первым предложил полуавтоматическую систему распознавания лиц с гибридной системой «человек-компьютер» [2, 3]. В 1973 г. Фишлер и Эльсшлагер предприняли попытку измерить похожие черты лица в автоматическом режиме [4]. В 1986 г. Стонхем предложил устройство «WISARD», предназначенное для распознавания образов и основанное на принципах нейронных сетей [5]. В 1991 г. Мэтью Турк и Пентлэнд предложили подход к проблеме распознавания лиц, основанный на методе главных компонент [6]. Сложные коммерческие системы распознавания лиц осуществляют поиск и идентификацию лица в реальном времени. Многие из них выполняют сопоставление изображений в два шага: проектирование изучаемого изображения в некоторое подпространство изученных изображений и

последующая классификация проекции в этом подпространстве. В простом каноничном случае сопоставление лица может быть реализовано как проектирование в подпространство и интерполяция методом ближайшего соседа [7].

**Метод главных компонент (РСА)** является самым распространенным методом проектирования изображения в подпространство для распознавания лиц. Основная идея РСА – построить для заданного  $t$ -мерного векторного пространства изображений  $x_1, x_2, \dots, x_n$   $f$ -мерное подпространство ( $f \ll t$ ), базисные векторы которого соответствуют направлениям максимальных вариаций в исходном пространстве. Базисные векторы вычисляются из набора обучающих изображений  $X = x_1, x_2, \dots, x_n$ . Сначала вычисляется усредненное изображение  $\mu$  из изображений обучающей выборки, которое затем вычитается из  $X$ , образуя тем самым набор образцов  $i_1, i_2, \dots, i_n \in X - \mu$ .

Этот набор выстраивается затем в виде матрицы  $I$ , столбцы которой составлены из образцов набора:

$$I = \left[ \begin{array}{c|c|c} \left[ \begin{array}{c} \vdots \\ i_1 \\ \vdots \end{array} \right] & \cdots & \left[ \begin{array}{c} \vdots \\ i_n \\ \vdots \end{array} \right] \end{array} \right]$$

Тогда  $I \cdot I^T$  – это ковариационная матрица для обучающего набора изображений, а главные компоненты ковариантной матрицы вычисляются путем решения системы  $R^T (I \cdot I^T) R = \Lambda$ , где  $\Lambda$  – диагональная матрица собственных значений, а  $R$  – это ортогональная матрица собственных векторов, матрица отображения исходной системы координат в систему, построенную на собственных векторах.  $f$  собственных векторов, которые соответствуют наибольшим собственным числам, выбираются в качестве базисных для нового подпространства. Число  $f$  – желаемая размерность этого подпространства [6].

Есть три аргумента в пользу проектирования изображений на подпространство из  $f$  собственных векторов. Первый – это компрессия. С точки зрения оптимизации вычислений, более эффективно выполнять сравнение изображений в подпространстве со значительно пониженной размерностью по сравнению с исходным. Второй аргумент подразумевает, что набор  $i_1, i_2, \dots, i_n$  подчиняется нормальному распределению. В этом случае оси, соответствующие наибольшим собственным значениям, соответствуют сигналу, а оси, соответствующие наименьшим собственным значениям, – шуму. Таким образом, удаление «зашумленных» осей улучшает качество сравнения изображений. Третий аргумент связан с первым шагом, в котором среднее значение вычитается из изображений обучающей выборки и полученный набор значений формирует исходное векторное пространство изображений, которое затем отображается в подпространство, где евклидово расстояние обратно пропорционально корреляции между исходными изображениями. В результате сравнение методом ближайшего соседа в новом подпространстве становится весьма эффективным.

Метод **линейного дискриминанта Фишера (LDA)** находит векторы в исходном векторном пространстве, которые наилучшим образом распределяются по разным классам, а не те, которые лучшим образом описывают данные. Для набора векторов, взятых из всех классов, составляются матрицы разброса – внутриклассовая  $S_W$  и межклассовая  $S_B$ :

$$S_B = \sum_{i=1}^c M_i \cdot (x_i - \mu) \cdot (x_i - \mu)^T, \quad S_W = \sum_{i=1}^c \sum_{x_k \in X_i} M_i \cdot (x_k - \mu_i) \cdot (x_k - \mu_i)^T,$$

где  $M_i$  – размер обучающей выборки в классе  $i$ ;  $c$  – количество различных классов;  $\mu_i$  – усредненное значение изображений в классе  $i$ ;  $X_i$  – набор изображений, принадлежащих классу  $i$ ;  $x_k$  –  $k$ -е изображение класса. Матрица  $S_W$  описывает разброс свойств вокруг среднего каждого из классов лиц, а матрица  $S_B$  описывает разброс свойств вокруг усредненного лица по всем классам.

Цель состоит в том, чтобы максимизировать  $S_B$ , минимизируя  $S_W$ , другими словами, максимизировать отношение  $\det|S_B|/\det|S_W|$ . Это отношение максимально, если векторы-столбцы матрицы-проекции являются собственными векторами матрицы  $S_W^{-1} \times S_B$ .

**Метод независимых компонент (ICA).** В то время как PCA генерирует некоррелированные (декоррелирует) данные используя статистику второго порядка и, таким образом, генерирует сжатые данные с минимальной квадратичной ошибкой при проектировании, ICA минимизирует как зависимости второго порядка, так и более высоких порядков во входных данных [8]. ICA близко связан с задачей *слепого разделения источников (BSS)*, где цель состоит в разложении наблюдаемого сигнала на линейную комбинацию неизвестных независимых сигналов. Пусть  $s$  – вектор неизвестных составляющих исходного сигнала, а  $x$  – вектор наблюдаемого суммарного сигнала. Если  $A$  – матрица смешивания, то модель смешивания будет выглядеть как  $x = As$ . При этом предполагается, что составляющие независимы, а матрица смешивания  $A$  обратима. Основываясь на этих предположениях, алгоритмы ICA пытаются найти матрицу смешивания  $A$  или разделяющую матрицу  $W$ , такие, что  $u = Wx = WA s$  является оценкой независимых исходных сигналов [9].

ICA может рассматриваться как обобщение PCA. PCA декоррелирует данные обучающей выборки так чтобы ковариация образцов обучающей выборки  $(i_1, i_2, \dots, i_n)$  была равна нулю. ICA делает еще один шаг вперед: он преобразует декоррелированные данные в набор статистически независимых сигналов [10].

Сигналы являются статистически независимыми, когда  $f_u(u) = \prod_i f_u(u_i)$ , где  $f_u$  – плотность вероятности  $u$ . К сожалению, матрицы  $W$ , которая полностью удовлетворяла бы требованию независимости исходных сигналов, может и не существовать, нет замкнутого выражения для нахождения  $W$ . Вместо этого

существует несколько алгоритмов, которые итерационно аппроксимируют  $\mathbf{W}$  так, чтобы косвенно минимизировать независимость исходных сигналов: InfoMax, JADE и FastICA [9]. Теоретически, все три алгоритма сходятся к одному и тому же решению для любой заданной обучающей выборки. На практике возможны расхождения, вызванные разными формулировками требования независимости исходных сигналов, которые закладываются в эти алгоритмы.

Независимо от того, какой алгоритм используется для вычисления независимых компонент, существует два фундаментально разных способа применения ИСА для распознавания лиц. В **архитектуре I** (статистически независимые базисные изображения) входные изображения лиц  $\mathbf{X}$  рассматриваются как комбинация статистически независимых базисных изображений  $\mathbf{S}$ , которые суммируются с помощью неизвестной матрицы смешивания  $\mathbf{A}$ . Алгоритм вычисляет весовую матрицу  $\mathbf{W}$ , которая используется для восстановления набора независимых базисных изображений. В этой архитектуре изображения лиц являются переменными, а значения пикселей предоставляют наблюдения этих переменных. Разделение исходного сигнала, таким образом, производится в пространстве лиц. Проектирование входных изображений на полученные весовые векторы порождает базисные независимые изображения. Сжатое представление изображения лица – это вектор коэффициентов, используемых для построения линейной комбинации базисных изображений, которая дает нужное изображение.

Базисные изображения, полученные в **архитектуре I**, статистически независимы. Коэффициенты, представляющие входные изображения в подпространстве базисных изображений, – не являются статистически независимыми. В **архитектуре II** (статистически независимые коэффициенты) пиксели – переменные, изображения – наблюдения этих переменных. Разделение источников производится на пикселях. Архитектура I порождает базис изображений пространственно локализованных объектов, совсем не похожих на базис, порождаемый PCA и эффективна для задач распознавания частей лица. Архитектура II наоборот – порождает базис, внешне напоминающий базис PCA, и эффективна для распознавания лица в целом.

**Сравнение.** Рассмотренные выше методы относятся к так называемым «appearance-based» методам, т.е. основанным на классификации внешности. Они не подразумевают существование и использование какой-то определенной модели или шаблона построения лица. Вместо этого используется большой набор изображений обучающей выборки и производится классификация по ней. В контексте «appearance-based» методов считается, что LDA-алгоритмы работают лучше чем PCA. Однако в [11] показано, что в случаях, когда объем обучающей выборки невелик, PCA может работать эффективнее чем LDA. PCA также менее чувствителен к замене одной обучающей выборки на другую. При этом построенные в [12] алгоритмы, базирующиеся на PCA и LDA, действительно показывают лучшую производительность LDA по сравнению с PCA, однако разница незначительна [12].

Одной из общих черт PCA и LDA является то, что они порождают пространство глобальных векторов-признаков. Применив ИСА для создания стати-

чески независимых базисных векторов, можно получить пространство локальных векторов-признаков, которые могут быть полезными для распознавания частей лица. ICA также может быть использован для создания векторов-признаков, которые равномерно распределяют изображения в подпространстве [8]. Это концептуально иное использования ICA создает векторы-признаки, которые не являются пространственно локализованными. Вместо этого ICA порождает векторы-признаки, которые хорошо различают похожие изображения.

Производительность ICA сильно зависит от задачи распознавания, архитектуры и алгоритма; производительность PCA зависит от метрики подпространства базисных векторов. Сравнение ICA архитектур I, II и PCA с различными метриками (L1, L2, косинусоидальная метрика, метрика Махаланобиса – L2, в которой каждое направление нормировано по соответствующему собственному значению) показывает, что для ICA архитектуры II косинусоидальная метрика работает лучше чем все остальные (82.3 % успешных распознаваний против 80.4 % для PCA L1, 74.9 % ICA II L2, 75.2 % PCA–Махаланобис, 72.8 % PCA L2, 70.7 % PCA– косинусоидальная) [13]. В работе [13] также утверждается, что для ICA архитектуры I косинусоидальная метрика обходит PCA L1 и что существенной разницы между косинусоидальной и Махаланобиса метриками нет. Однако приведенные результаты успешности применения алгоритмов с соответствующими метриками этого не подтверждают (ICA косинусоидальная 50.6 %, PCA L1 57.3 % и PCA–Махаланобис 75.2 %). Полученные в [13] выводы не зависят от размерности подпространства практически для всех метрик и алгоритмов. Исключение – PCA с метрикой Махаланобиса, производительность которого снижается и становится хуже ICA II и PCA L1 при увеличении размерности базисного подпространства.

Тестирование рассмотренных алгоритмов применительно к задаче распознавания выражения лица показывает эффективность алгоритма ICA (архитектура I) – 88.1 %. На втором месте PCA L2 (85.0 %), затем идет PCA L1 (79.0 %), затем ICA архитектура I (75.9 %) и PCA–Махаланобис (58.5 %) [13]. Этот результат находится в соответствии с гипотезой о том, что пространственно-локализованные базисные векторы работают лучше, чем пространственно-перекрывающиеся для распознавания выражения лиц. Это также свидетельствует о том, что выбор алгоритма распознавания зависит от поставленной задачи. Этот вывод подтверждается также и в [14].

Выбор алгоритма распознавания лиц представляет довольно трудную задачу, поскольку качество работы того или иного алгоритма сильно зависит как от природы поставленной задачи, так и от выбора архитектуры самого алгоритма. Тем не менее, можно сделать вывод, что среди рассмотренных алгоритмов ICA обеспечивает наилучший результат при правильном выборе архитектуры. При этом ICA отлично подходит как для задач глобального распознавания лиц (архитектура II), так и для задач локального распознавания – выражения лица (архитектура I). PCA (метрика L2) показывает хороший и стабильный результат по отношению к типу решаемой задачи.

*V.V. Gorin*

ПОРІВНЯННЯ АЛГОРИТМІВ АНАЛІЗУ ЗОБРАЖЕНЬ У ЗАДАЧІ РОЗПІЗНАВАННЯ ОБЛИЧ

Розглянуто декілька найбільш ефективних на сьогоднішній день алгоритмів розпізнавання облич. Показано, що продуктивність алгоритму сильно залежить від природи поставленої задачі та від архітектури самих алгоритмів розпізнавання. Показано, що алгоритм ІСА є найбільш ефективним та універсальним – архітектура II для задач глобального розпізнавання облич, архітектура I – для задач локального розпізнавання (міміка, вираз обличчя).

*V.V. Gorin*

IMAGE-BASED FACE RECOGNITION ALGORITHMS AND THEIR COMPARISON

This paper describes few of the most popular and effective image-based face recognition algorithms. It is shown that algorithms performance depends scientifically on the task statement and algorithm architecture as well. It is shown that ICA is the most universal and effective algorithm for both facial identity recognition tasks (ICA architecture II) and facial action (mimics/emotion) recognition (ICA architecture I)

1. *Carey S., Diamond R.* From piecemeal to configurational representation of faces // *Science* – 1977. – **195** (4275). – P. 312 – 314.
2. *Bledsoe W.* The model method in facial recognition // *Technical Report PRI-15.* – 1964. – 142 p.
3. *Bledsoe W., Chan H.* A Man-Machine Facial Recognition System // *Technical Report PRI-19A.* – 1965. – 103 p.
4. *Fischler M., Elschlager R.*, The Representation and Matching of Pictorial Structures // *IEEE Transactions on Computers* – 1973. – **22.** – P. 67 – 92.
5. *Stonham T. J.* Practical face recognition and verification with WISARD. // *Aspects of Face Processing* – 1984. – P. 426 – 441.
6. *Turk M., Pentland A.* Eigenfaces for recognition // *JoCN* – 1991. – **3** (1). – P. 71 – 86.
7. *Moon H., Phillips P.J.* Computational and Performance aspects of PCA-based Face Recognition Algorithms. // *Perception* – 2001. – **30.** – P. 303 – 321.
8. *Bartlett M., Movellan J., Sejnowski T.* Face Recognition by Independent Component Analysis // *IEEE Transactions on Neural Networks* – 2002. – **13** (6). – P. 1450 – 1464.
9. *Cardoso J.-F.* Infomax and maximum likelihood for blind source separation // *IEEE Letters on Signal Processing.* – 1997. – **4** (4). – P. 112 – 114.
10. *Hyvärinen A., Karhunen J., Oja E.* Independent Component Analysis. – New York: Wiley, 2001. – 504 p.
11. *Martinez A., Kak A.* PCA versus LDA // *IEEE TPAMI* – 2001. – **23** (2). – P. 228 – 233.
12. *Belhumeur P., Hespanha J., Kriegman D.* Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection // *IEEE TPAMI* – 1997. – **19** – P. 711 – 720.
13. *Draper B., Baek K., Bartlett M., Beveridge J.* Recognizing Faces with PCA and ICA // *Computer Vision and Image Understanding* – 2003. – **91** (1-2). – P. 115 – 137.
14. *Donato G., Bartlett M., Hager J., Ekman P., Sejnowski T.* Classifying Facial Actions // *IEEE TPAMI* – 1999. – **21.** – P. 974 – 989.

Получено 14.04.2011