

Методы автоматической идентификации лиц

Ю.И.Якименко, д. т. н., проф., В.Г.Дзюба

*Национальный технический университет Украины «КПИ», г. Киев
dzyuba@ee.ntu-kpi.kiev.ua*

В работе рассмотрены наиболее распространенные методы идентификации лиц. Основное внимание уделялось временным характеристикам работы алгоритмов, а также их качеству идентификации. Цель данной работы состоит в определении наиболее перспективного алгоритма для последующей его реализации в виде аппаратно-программного комплекса.

Введение

Процесс автоматической идентификации лиц предполагает создание аппаратно-программного средства, способного при минимальном вмешательстве человека проводить качественную идентификацию, используя только изображение лица человека. Актуальность этой задачи, а также ее предпочтительность по сравнению с другими средствами идентификации личности (например идентификация по отпечаткам пальцев или по сетчатке глаза) заключается в том, что нет необходимости непосредственного контакта системы и человека. Лицо человека само по себе содержит уникальную информацию для проведения безошибочной идентификации. Основная трудность, с которой пытаются справиться предложенные алгоритмы состоит в том, что задача идентификации человека по лицу принадлежит к классу трудноформализуемых задач, для решения которых и используются современные методы искусственного интеллекта.

Методы идентификации лиц

В работе [1] рассмотрен подход, основанный на геометрических характеристиках лица. В этой работе использовалась база данных лиц, состоящая из 188 фотографий (по четыре фотографии для каждого из 47 человек), изображения (фотографии) которых получались при помощи специальной ССD-камеры. Расстояние от объекта до камеры во время съемки не было четко фиксировано, поэтому изменение в масштабе

составляло до 30%. Авторы этого метода ставили следующие требования к алгоритму:

- максимальная простота оценочного алгоритма ;
- влияние освещения должно быть минимально ;
- минимальная чувствительность к мимике .

Первым ключевым моментом, который определяет качество работы алгоритма в целом, является предобработка или нормирование изображения с целью улучшения яркостно-контрастных характеристик. Извлеченные характеристические точки должны быть нормированы для обеспечения инвариантности по отношению к позиции (расположению) лица, масштабу и углу поворота. Другим ключевым моментом является процедура определения набора характеристических точек. Рассматриваемый геометрический метод опирается на свойство билатеральной симметрии лица. Используется множество отношений расстояний между такими характеристическими точками лица как глаза, кончик носа, центр рта. Алгоритм определения характеристических точек базируется на интегральной проекции изображения. Проекционный анализ проводится над бинарным изображением, полученным путем применения *оператора Лапласа* . Применяя вышеуказанный алгоритм были получены характеристические точки расположения глаз, носа, бровей, рта, а также был определен контур лица, на основании чего было выделено 35 геометрических признаков.

Применяя к описанным подобным образом лицам *классификатор Байеса* авторы этого метода добились 90%-ного распознавания на базе данных 47 человек.

В работе [2] также предложен геометрический метод идентификации лиц, основанный на геометрическом расстоянии между ключевыми точками. Схожесть лиц, согласно этому методу, оценивается в несколько этапов, действуя по принципу “от грубого к точному”. Сначала определяется множество ключевых точек для каждого лица, затем отбирается приблизительно k_1 изображений, близких по геометрическим характеристикам к классифицируемому лицу; k_1 определяется так:

$$k_1 = (\ln N)^2 \quad (1)$$

где N – количество портретов, содержащихся в базе данных.

После этого выполняется этап нормализации изображения для последующего вычисления корреляции между центральными лицевыми частями k_1 изображений и лицевой частью идентифицируемого изображения. Результатом будет получение k_2 изображений из k_1 отобранных:

$$\ln k_1 \leq k_2 \leq (\ln k_1)^2 \quad (2)$$

Ключевые точки определяются автоматически, но для проведения более точной идентификации возможен также ручной отбор точек. В экспериментах, проводимых авторами данного геометрического метода использовалось множество из 37 ключевых точек. Пример полученных геометрических точек представлен на рис. 1

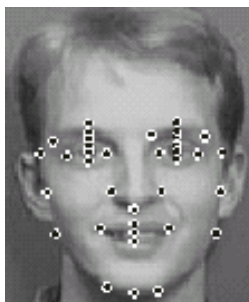


Рис. 1 Пример множества геометрических ключевых точек.

Для задачи идентификации необходимо выполнить этап объединения точек в сегменты, периметры и площади различных фигур – определение репрезентативного набора характеристик. Далее этот набор может быть оптимизирован для достижения более высокого качества идентификации. Сам процесс идентификации заключается в определении “самого близкого” изображения по отношению к идентифицируемому изображению с использованием Евклидовой метрики. Гипотеза, выдвинутая авторами этой работы, следующая – если два изображения содержат практически одинаковую информацию, то они должны содержать множество пикселей с близкими значениями интенсивности.

Изображение A можно представить как дискретную поверхность:

$$A = \{(i,j,a_{ij})\}, \quad 0 \leq i,j \leq N_0 \quad (3)$$

Расстояние от каждого пикселя одного изображения к ближайшим пикселям другого изображения отражает, так называемое, локальное отличие. Вычисляя множество локальных отличий и аккумулируя их в глобальное значение отличия, формируется окончательная оценка схожести изображений. Глобальное значение отличия двух изображений A и C может быть определено по следующей формуле:

$$D(A,C) = \sqrt{\frac{1}{2ZN} \left\{ \sum_{i,j} [d(a_{ij},C) + d(c_{ij},A)]^2 \right\}^{1/2}} \quad (4)$$

где

$$d(a_{ij}, C) = d(a_{ij}, C_W) = \min_{(l,m) \in W} \{d(A_{ij}, C_{lm})\} \quad (5)$$

и W – квадратное окно наблюдения размера $(2w+1) \times (2w+1)$ центрированное в точке (i,j) изображения C .

Эксперименты по определению качества распознавания проводились на базе данных лиц, составленной Olivetti Research Lab. Качество распознавания составило 98,5%.

Использование *скрытых марковских моделей* для идентификации лиц было выполнено в работе [3]. Скрытые марковские модели (СММ) – это множество статистических моделей, используемых для описания статистических свойств сигналов.

Можно выделить следующие элементы СММ :

- N - количество состояний модели . Если S – множество состояний, следовательно $S = \{S_1, S_2, \dots, S_N\}$. Состояние модели в момент времени t обозначается q_t ($q_t \in S$), $1 \leq t \leq T$, где T – длина последовательности наблюдений (количество фреймов).
- M - количество различных наблюдаемых символов. Если V – множество всех возможных наблюдаемых символов , тогда $V = \{v_1, v_2, \dots, v_M\}$.
- A - вероятностная матрица перехода состояний , т.к. $A = \{a_{ij}\}$, где $a_{ij} = P[q_t = S_j | q_{t-1} = S_i] \quad 1 \leq i, j \leq N, 0 \leq a_{ij} \leq 1$

$$\sum_{j=1}^N a_{ij} = 1, \quad 1 \leq i \leq N \quad (6)$$

- B - матрица вероятностей наблюдаемых символов, $B = \{b_j(k)\}$, где

$$b_j(k) = P[O_t = v_k | q_t = S_j] \quad (7)$$

O_t – наблюдаемый символ в момент времени t

- Π - распределение начального состояния, $\Pi = \{p_i\}$, где

$$p_i = P[q_1 = S_i], \quad 1 \leq i \leq N \quad (8)$$

Используя сокращенную нотацию, СММ обозначается как триплет:

$$I = (A, B, p) \quad (9)$$

Для фронтального изображения лица наиболее важные регионы (волосы, лоб, глаза, нос, рот) следуют сверху вниз, даже если изображение подверглось небольшому вращению. Каждый из этих лицевых регионов соответствует состоянию СММ.

Любое изображение шириной W и высотой H разделяется на перекрывающиеся блоки высотой L и шириной W . Величина перекрытия последовательных блоков равна P (рис.2).

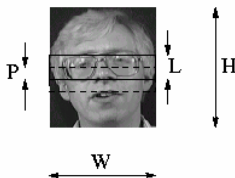


Рис.2 Извлечение перекрывающихся блоков для распознавания лиц

Количество извлекаемых блоков равно количеству векторов наблюдения T и определяется:

$$T = \frac{H - L}{L - P} + 1 \quad (10)$$

Выбор параметров P и L существенно влияет на процент правильного распознавания. Векторы наблюдения состоят из значений пикселей каждого извлекаемого блока. Использование значений пикселей несет в себе два недостатка: значение пикселей не дает помехоустойчивые особенности, т.к. они очень чувствительны к шуму, к вращению изображения, к изменениям в освещенности; большая размерность векторов влечет за собой высокую вычислительную стоимость. Для решения этой проблемы используется дискретное косинус-преобразование (ДКП) с целью извлечения характеристических точек из каждого блока.

Набор из пяти блоков используется для обучения СММ. Характеристические коэффициенты, извлекаемые из каждого блока, задействованы для формирования векторов наблюдений. Векторы наблюдений, ассоциируемые с каждым состоянием СММ, используются для получения первоначальных оценок матрицы вероятностей символов наблюдения. Далее параметры модели переоцениваются для максимума $P(O/I)$. Итерационный процесс продолжается до тех пор, пока не будет достигнута сходимость, т.е. разность между текущим значением и значением

на предыдущей итерации не будет меньше некоторого порога. После формирования векторов наблюдений вычисляются вероятности векторов наблюдений для каждого состояния СММ.

Изображение t классифицируется как лицо k если :

$$P(O^{(t)} | I_k) = \max_n P(O^{(t)} | I_k) \quad (11)$$

Данный метод распознавание лиц тестировался на базе данных лиц Olivetti Research Lab. Система достигла 84% уровня верного распознавания лиц.

Довольно перспективная технология распознавания лиц с использованием сверточной нейронной сети (СНС) рассмотрена в работе [4] . СНС использует три основных идеи: локальные рецепторные поля, разделяемые веса, пространственная подвыборка. Эта сеть эффективно может использоваться не только для распознавания лиц, но также и для распознавания символов и речи.

Типичная СНС представлена на рис.3. Эта сеть состоит из набора слоев, каждый из которых содержит одну или более плоскостей.

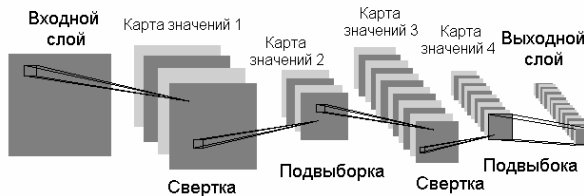


Рис.3 Сверточная нейронная сеть

Центрированное и нормированное изображение поступает на входной слой. Каждый нейрон в плоскости получает входной сигнал из области, которая находится “по соседству” на плоскости предыдущего слоя. Веса нейронов устанавливаются одинаковыми. Каждая плоскость может рассматриваться как карта особенностей или карта значений. Множество плоскостей всегда присутствует в каждом слое, поэтому слои называются сверточными. Следовательно, сверточные слои – это такие слои, которые следуют один за другим, при этом используются операторы усреднения и подвыборки (если фактор подвыборки равняется 2 : $y_{ij} = [x_{2i,2j} + x_{2i+1,2j} + x_{2i,2j+1} + x_{2i+1,2j+1}]/4$, где y_{ij} - выход плоскости подвыборки в точке с координатами i, j и x_{ij} – выход в предыдущей плоскости). Обучается сеть стандартным градиентным методом обучения. В

зависимости от выбора стратегии соединений возможна вариация количества весов СНС. Стратегия соединения определяется человеком вручную.

Этапы работы системы следующие:

1. Для каждого поступающего изображения из обучающей выборки окно фиксированного размера перемещается по всей площади этого изображения, извлекая локальные образцы. Перемещение происходит дискретными шагами по 4 пикселя.
2. Самоорганизующаяся карта обучается на векторах, полученных на предыдущем этапе. Таким образом она квантует 25-мерный вектор в одно из 125 топологических значений.
3. Сканирующее окно, описанное в первом этапе, перемещается по всем изображениям обучающей и тестирующей выборок. Локальные образцы подаются на карту Кохонена, на выходе которой формируются новые обучающая и тестирующая выборки. Теперь каждое входное изображение представляется посредством 3 карт. Размер этих карт равен размеру входного изображения, деленного на размер шага.
4. СНС обучается на новой обучающей выборке.

Эксперимент по исследованию характеристик работы системы проводился с фотографиями лиц (5 обучающих фотографий на 5 тестирующих фотографий одного и того же человека). Всего использовалось 400 фотографий 40 человек. Ошибка распознавания в зависимости от количества классов составила от 1,33% до 5.75%.

Заключение

Для оценки наиболее оптимального метода необходимо определить параметры, по которым проведем сравнительный анализ. Первый параметр – процент правильного распознавания (ППР) : диапазон от 0 до 100%; второй – относительная временная стоимость (ОВС), который делится на категории : высокая, средняя, низкая и отражает временные затраты работы системы; третий – сложность системы (высокая, средняя, низкая) и четвертый параметр – возможность обучения системы (да или нет). В таблице 1 приведены и оценены все вышеописанные методы.

Таблица 1. Сравнительный анализ методов

№	Название	ППР	ОВС	Сложность системы	Обучаемость
1	Геометрический метод №1	90%	-*	средняя	нет
2	Геометрический метод №2	98%	высокая	средняя	нет
3	Скрытые марковские модели	84%	средняя	средняя	да
4	Сверточная нейронная сеть	98-99%	средняя	средняя	да

Таблица показывает, что из всех приведенных систем только некоторые способны обучаться и переобучаться, т.е. улучшать качество своей работы. Это является одним из самых важных свойств при создании аппаратно-программного средства идентификации лиц. Однако вторым важным параметром является процент правильного распознавания системы. Руководствуясь этими двумя параметрами как критериями выбора наиболее оптимального метода, можно заключить, что оптимальным вариантом является система построенная с использованием сверточной нейронной сети.

Литература

1. Brunelli R. and Poggio T. Face recognition: features versus templates// IEEE Trans. on Pattern Analysis and Machine Intel.- 1993. - Vol.15. - No 10. - P. 1042-1052.
2. Starovoitov V., Samal D., G. Votsis, and S. Kollias "Face recognition by geometric features", *Proceedings of 5-th Pattern Recognition and Information Analysis Conference*, Minsk, May 1999.

* Символ "-" означает, что информация по данному параметру отсутствует

3. Ara V.Nefian , Monson H. Hayes. Hidden Markov models for face recognition // IEEE Trans. on Pattern Analysis and Machine Intel.- 1993. - Vol.27. - No 15. - P. 985-997.
4. Lawrence S., Giles C. L., Tsoi A. C. and Back A. D. Face recognition: a convolutional neural network approach// IEEE Trans. on Neural Networks.- Special Issue on Neural Networks and Pattern Recognition.- 1997.- P.97-113.