

УДК 519.7

А.А. Петрашко, В.Т. Калайда

Методы выделения фрагментов изображения лица человека в системах распознавания образов

В работе рассмотрены методы выделения фрагментов изображения лица человека. Делается вывод о целесообразности использования приведенных методов совместно с нейросетевыми технологиями.

При анализе и идентификации изображений имеется целый ряд практических задач, решение которых оказывается достаточно проблематичным. К числу подобных задач относится распознавание образов. Обусловлено это во многом сложностью процесса восприятия видимых образов. Несмотря на очевидную легкость, с которой человек решает задачу распознавания окружающих его предметов, все еще нет «универсального» математического или технологического подхода, позволяющего конструктивно разрабатывать методы, алгоритмы и автоматические устройства, эффективно осуществляющие процесс распознавания.

Весь спектр задач, которые приходится решать при распознавании образов на изображениях, можно подразделить на две группы:

- распознавание или классификация изображений;
- поиск и распознавание объектов (специфических локальных областей, сюжетной части) на изображениях.

Это разделение связано с особенностями реализации процесса распознавания. В первой группе задач распознавание или классификация производится для всего изображения целиком. То есть все изображение целиком в процессе распознавания относят к одному из нескольких классов. Таким образом, решением задачи распознавания в этой группе является реализация отображения: *изображение — номер класса*. Указанное отображение реализуется в виде следующих двух этапов: *изображение — признаки* и *признаки — класс*. Это позволяет представить процесс решения задачи классификации изображения в виде схемы, которая является традиционной и стандартной для задач распознавания образов (рис. 1).



Рис. 1 — Модель распознавания образов

Примером задач *первой группы* являются задачи распознавания лиц по фотографиям.

В задачах *второй группы* процесс распознавания оказывается включенным в более общую технологию обработки изображения, связанную с поиском распознаваемых геометрических объектов на всей области наблюдения. Объекты — в данном случае лица, представляют собой относительно небольшие локальные области, появление которых может произойти в любой точке изображения. Задача усложняется еще и тем, что информация о том, имеются ли на изображении лица, каково их количество, ориентация, размеры и т.д., чаще всего отсутствует.

Результатом решения задачи распознавания в этой ситуации является не только класс найденного объекта, но также и его характеристики: положение, размер, возможно, ориентация лица в области изображения и т.д.

Неопределенность в целом ряде характеристик объектов делает задачу их поиска и распознавания на изображениях в математическом и вычислительном плане более сложной по сравнению с задачами первой группы. Это приводит к тому, что процесс ее решения не укладывается в приведенную схему, а производится в соответствии со схемой, в упрощенном виде представленной на рис. 2, то есть включает в себя трудно формализуемую задачу выделения фрагментов [1].

В соответствии с данной схемой анализу подвергается каждый фрагмент на изображении. По текущему фрагменту, выделенному окном обработки, производится формирование

признаков и классификация. В зависимости от результатов классификации происходит расчет дополнительных параметров объекта.

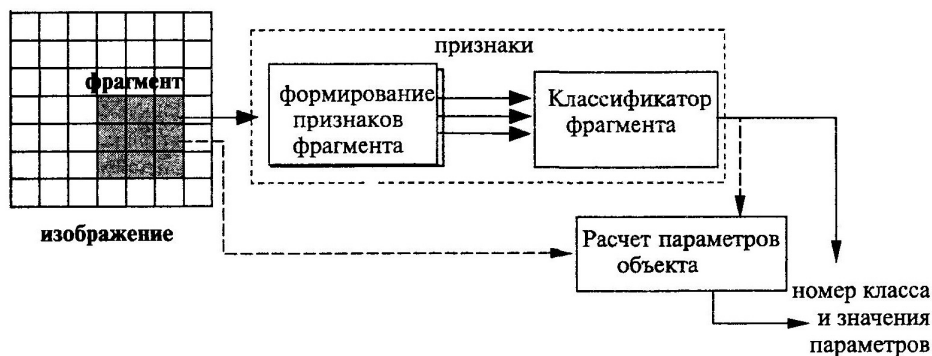


Рис. 2 — Схема поиска и распознавания объектов на изображении

Легко заметить, что схема решения классификации изображения входит как составной элемент в схему задачи поиска и распознавания. Действительно, в более широком понимании задачи второй группы относятся к задачам высокого уровня — задачам анализа наблюдаемого изображения или сцены [2]. Под анализом сцены в данном случае понимается составление полного описания изображенных на снимке предметов с указанием их местоположения и взаимного расположения.

Существует множество важных задач распознавания объектов, которые включают поиск на изображении окон, имеющих простую форму и стилизованное содержимое. Например, лицо анфас имеет вид овального окна, причем (при грубом масштабе) все лица выглядят приблизительно одинаково: темные горизонтальные полосы на уровне глаз и рта, светлая вертикальная полоса вдоль носа и слаботекстурные образования в районе щек и лба.

Исходя из приведенных соображений, напрашивается следующий подход к распознаванию лиц: определить все окна изображений, имеющие овальную форму, и проверить их на предмет наличия лица. Если о размерах ничего не известно, поиск можно проводить при разных масштабах; если не известна ориентация лица, можно дополнительно проводить поиск в пространстве ориентаций и т.д. Чтобы ответить на главный вопрос — представляет ли данный овал лицо — необходимо разработать тесты для проверки. В идеале для получения теста используется большой набор образцов, так называемый классификатор.

Известны классификаторы для нахождения кожи человека [3]. Кожа имеет довольно характерный диапазон цветов, так что детектор обнаружения кожи можно построить, классифицируя пиксели по их цвету.

Запишем через x вектор, содержащий цветовые значения пикселя. Разделим данное цветовое пространство на квадратики и подсчитаем процентную часть пикселей телесного цвета, попадающих в каждую ячейку. Величину $P(x | \text{пиксели кожи})$ можно оценить, определив квадратик, соответствующий x , а затем указать процентное содержание пикселей в этой ячейке. Подобным образом, определяя процентное содержание в каждом квадратике пикселей нетелесного цвета, получим величину $P(x | \text{иные пиксели})$.

Для оценки априорной вероятности можно смоделировать $P(\text{кожа})$ как долю пикселей телесного цвета в некотором (идеально большом) настроечном наборе. Полученный таким образом классификатор сравнивает

$$\frac{P(x | \text{кожа})P(\text{кожа})}{P(x)} L(\text{кожа} \rightarrow \text{иное}) \quad \text{с} \quad \frac{P(x | \text{иное})P(\text{иное})}{P(x)} L(\text{иное} \rightarrow \text{кожа}).$$

Теперь отметим, что $P(\text{пиксели кожи} | x) = 1 - P(\text{иные пиксели} | x)$, и получим такой классификатор:

- если $P(\text{кожа} | x) > \theta$, отнести к пикселям кожи;
- если $P(\text{кожа} | x) < \theta$, отнести к иным пикселям;
- если $P(\text{кожа} | x) = \theta$, случайно и равномерно выбрать из классов.

Здесь θ — выражение, которое не зависит от x и учитывает относительные потери. В результате получаем семейство классификаторов, по одному для каждого выбора θ . При подходящем выборе θ классификатор может быть удачным.

Все лица содержат небольшое число компонентов, достаточно строго упорядоченных. Как правило, на лице выделяют глаза, нос и рот. Расстояние между глазами примерно равно расстоянию от переносицы до рта. Линия, соединяющая глаза, перпендикулярна линии от переносицы до лба и т.д. Можно попытаться согласовать данные компоненты, а затем определить, какие объекты присутствуют на изображении, изучив предложенные связи между ними. Например, вместо поиска целого лица можно искать глаза рот и нос с приемлемым взаимным расположением.

Поставленная задача остается сложной. Требуется построить модели, которые представят, что является значимым, и позволят эффективно сформулировать задачу логического вывода. На данный момент не существует общепринятого подхода к решению этого вопроса. Можно описать только основные принципы решения — использование вероятностной модели для сопоставления фрагментов изображения.

При практической реализации сформулированной задачи возникает множество проблем. Основная из них — не известно, какая информация, полученная с изображения, порождена объектами, а какая — шумом. В теории данная проблема обходится, и задача сводится к исследованию связей между компонентами изображения. Например, если требуется найти лица, на изображение действуют рядом различных детекторов (детекторами глаз, носа и рта), а затем ищутся допустимые конфигурации.

Из этого вытекает вторая важная проблема — проблема соответствия. Плотность апостериорной вероятности оценить невозможно, если не знать значения всех переменных. По сути, это означает, что в процессе требуется положить, что один отклик — это левый глаз, второй — правый глаз, третий — нос, четвертый — рот, а только после этого оценить апостериорную вероятность. Очевидно, этим поиском важно управлять аккуратно, чтобы не требовалось перебирать все возможные соответствия.

Вообще, для вероятностной модели оценивается некоторый набор переменных. Множество значений этих переменных называется совокупностью [3]. Совокупность состоит из набора выходов детекторов, каждый из которых может выдавать положение, ориентацию и метки для каждого выхода. Необходимо отметить, что данная модель предполагает, что на изображении присутствует только одно лицо либо лица нет вообще.

Все совокупности сформировать и проверить невозможно, поскольку обычно их очень много. Например, предположим, что есть детекторы глаз, которые не различают левый и правый глаза, детекторы носа и детекторы рта, а лицо состоит из двух глаз, носа и рта. Если всего имеется N_e откликов с детекторов глаз, N_n откликов с детекторов носа и N_m откликов с детекторов рта, поиск надлежит проводить в пространстве $O(N_e^2 N_n N_m)$ совокупностей. Это нецелесообразное усложнение задачи. Основная идея состоит в том, что находится довольно малое число соответствий, а по нему предсказываются другие соответствия.

Предположим, детектор левого глаза дает ненулевой отклик в точке x_1 , детектор правого глаза — в точке x_2 , детектор рта — в x_3 и детектор носа — в x_4 . Предполагается, что лицо расположено в F и все остальные отклики детекторов вызваны шумом. Более того, предполагается, что на изображении либо присутствует единственное лицо, либо лица нет вообще. Итак, требуется сравнивать величины

$$P(\text{одно лицо в } F \mid X_{lg} = x_1, X_{ng} = x_2, X_p = x_3, X_n = x_4, \text{ все другие отклики})$$

и

$$P(\text{нет лица} \mid X_{lg} = x_1, X_{ng} = x_2, X_p = x_3, X_n = x_4, \text{ все другие отклики}).$$

Предположим, что отклики, порожденные шумом, не зависят от присутствия лица (что вполне правдоподобно). Получаем

$$P(\text{одно лицо в } F \mid X_{lg} = x_1, X_{ng} = x_2, X_p = x_3, X_n = x_4, \text{ все другие отклики})$$

равно

$$P(\text{одно лицо в } F \mid x_1, x_2, x_3, x_4)P(\text{все другие отклики}),$$

что в свою очередь пропорционально

$$P(x_1, x_2, x_3, x_4 \mid \text{одно лицо в } F)P(\text{все другие отклики})P(\text{одно лицо в } F).$$

Определенные группы откликов детекторов можно классифицировать как порожденные лицом или шумом, сравнивая апостериорную вероятность того, что данная конфигурация порождена лицом, с апостериорной вероятностью того, что она порождена шумом. В частности, сравниваются величины

$$P(x_1, x_2, x_3, x_4 \mid \text{одно лицо в } F)$$

и

$$\frac{P(\text{отклики порожденные шумом})P(\text{нет лица})}{P(\text{одно лицо в } F)}$$
 коэффициент допустимой погрешности.

Когда группа признаков удовлетворяет критерию классификации — апостериорная вероятность присутствия лица превышает апостериорную вероятность отсутствия лица, поиск можно прекратить. Вероятно, проверять все возможные признаки для определения наличия лица будет не нужно. Если конфигурация явно указывает на наличие лица, а не на последствия шума, то можно утверждать, что лицо присутствует, и прекращать поиск.

В заключение нужно отметить, что задачи распознавания образов не имеют точного аналитического решения в связи с тем, что не существует формализованного описания изображения (Старавойтов В.В., Брилюк Д.В. Нейросетевые методы распознавания изображений. http://rusnauka.narod.ru/lib/author/briluk_d_b/1/index.html). Поэтому наиболее эффективных результатов работы приведенных детекторов лиц следует ожидать при их совместном использовании с неформализованными методами, например многослойными нейронными сетями.

Литература

1. Сойфер В.А., Храмов А.Г. и др. Методы компьютерной обработки изображений. — М.: Физматлит, 2003. — 780 с.
2. Грузман И.С., Киричук В.С. и др. Цифровая обработка изображений в информационных системах. — Новосибирск: Изд-во НГТУ, 2000. — 168 с.
3. Форсайт Д., Понс Ж. Компьютерное зрение. Современный подход. — М.: Издательский дом «Вильямс», 2004. — 928 с.

Петрашко Андрей Анатольевич

Программист Института дополнительного образования Томского политехнического университета

Телефон: (3822) 49 22 42

Эл. почта: andrey@ido.tpu.edu.ru

Калайда Владимир Тимофеевич

Канд. техн. наук, ст. науч. сотр. кафедры автоматизированных систем управления ТУСУРа

Телефон: (3822) 49 22 42

Эл. почта: kvk@iao.ru

A.A. Petrashko, V.T. Kalaida

Image of a face extraction method| in systems of figures recognition

The image of a human face extraction method were described. Conclusion of expediency of using described methods in common with neuronetworking technologies was made.

УДК 681.3.067

П.В. Потапов

Двухпроходный режим компрессии видеоданных

В статье предлагается двухпроходный алгоритм распределения ресурсов при компрессии видеоданных. Проводится анализ преимуществ и недостатков использования двухпроходных алгоритмов.

Представление видеоизображений в цифровой форме дает много преимуществ при съемке, обработке и воспроизведении. Однако объем данных, описывающих видеоматериал, может достигать гигантских размеров. Для компактного хранения видеоданных и передачи их по каналам связи с ограниченной пропускной способностью применяются специальные алгоритмы сжатия. Для обеспечения совместимости аппаратуры и программных систем были