

УДК 004.93'1:004.932

Р. И. Ш у в а л о в

РАСПОЗНАВАНИЕ ОБРАЗОВ НА ЦИФРОВЫХ ИЗОБРАЖЕНИЯХ НА ОСНОВЕ ТЕОРИИ ИНВАРИАНТОВ

Рассмотрена задача распознавания образов на цифровых изображениях и определения их положения, ориентации и масштаба относительно заданной системы координат. Описан метод решения этой задачи на основе вычисления инвариантов. Приведен алгоритм решения задачи.

E-mail: shuvalov.r.bmstu@mail.ru

Ключевые слова: распознавание образов, цифровые изображения, инварианты, параметры геометрических преобразований.

В прикладных областях, связанных с цифровой обработкой изображений (дистанционное зондирование Земли, медицина, астрономия, системы компьютерного зрения), актуальна задача распознавания образов на изображениях и определения их положения, ориентации и масштаба относительно заданной системы координат [1]. Формализуем эту задачу следующим образом.

Даны два изображения, относительно которых предполагается, что:

- 1) изображения информативны (т. е. неоднородны и неизотропны);
- 2) одно изображение может быть получено из другого путем сложного преобразования, являющегося суперпозицией трансляции, поворота, масштабирования и наложения случайного шума:

$$f_2(x_t, y_t) = f_1((x \cos \theta - y \sin \theta)s + \Delta x, (x \sin \theta + y \cos \theta)s + \Delta y) + \eta(x_t, y_t); \quad (1)$$

$$\begin{pmatrix} x_T \\ y_T \\ 1 \end{pmatrix} = \begin{pmatrix} s \cos \theta & -s \sin \theta & \Delta x \\ s \sin \theta & s \cos \theta & \Delta y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix},$$

где $f_1(x, y)$, $f_2(x_t, y_t)$ — пара исходных изображений; θ — угол поворота второго изображения относительно первого; s — коэффициент растяжения (сжатия) второго изображения относительно первого; $(\Delta x, \Delta y)^T$ — вектор параллельного переноса второго изображения относительно первого; η — реализация случайного шума.

Требуется проверить при заданном уровне значимости гипотезу о том, что два изображения значительно перекрываются (изображают большей частью одну и ту же сцену) и связаны функциональным преобразованием (1). В случае установления факта наличия функциональной связи оценить ее параметры (т. е. Δx , Δy , θ , s).

Цифровое изображение. Изображение — функция двух аргументов $f(x, y)$, определенная в некоторой замкнутой ограниченной двумерной области $\Pi \subset \mathbb{R}^2$. Область Π называется полем зрения и обычно представляет собой прямоугольник $\Pi = \{(x, y) \in \mathbb{R}^2 : 0 \leq x \leq X, 0 \leq y \leq Y\}$. Если функция f является скалярной, то значение этой функции в любой точке (x, y) называют яркостью. Если величины x , y и f принимают лишь конечное число дискретных значений, то говорят о цифровом изображении. Цифровой обработкой изображений называется обработка цифровых изображений с помощью цифровых вычислительных машин (компьютеров). Цифровое изображение состоит из конечного числа элементов, каждый из которых расположен на конкретном месте и принимает определенное значение. Эти элементы называются элементами изображения, или пикселями — от англ. pixel (picture element).

Распознавание образов. Снятому физическому объекту на цифровом изображении соответствует некоторая группа изображающих его пикселей. Эта группа пикселей называется образом объекта. В более широком смысле образом объекта можно назвать любой хорошо определенный набор измерений характеристик этого объекта. Под распознаванием образов понимается отнесение данного образа к одному из ограниченного числа дискретных классов. Каждый класс обычно соответствует некоторому физическому объекту. Поэтому распознать заданный объект на изображении значит определить, может ли этот объект при некоторых условиях съемки дать заданный образ на изображении. Распознавание предполагает наличие априорной базы моделей образов распознаваемых объектов. В качестве модели образа может использоваться вектор числовых характеристик. Обнаружение образа объекта на изображении представляет собой статистическую задачу проверки гипотез, решение которой сводится к построению статистического критерия, позволяющего по выборочным данным отдавать предпочтение одной из конкурирующих гипотез. Гипотеза о наличии образа принимается, если различие между модельным и фактическим описанием (с учетом преобразования определенного вида, параметры которого предстоит оценить в случае принятия гипотезы) мало (ниже предвычисленного порога) и объяснимо погрешностями и шумом.

Основные причины сложности автоматического распознавания образов на изображениях:

- наличие в предъявленном образе случайного шума и искажений;
- заслонение одних объектов другими (нарушение однозначности и непрерывности отображения при формировании изображения);
- различие условий съемки (масштабирование, поворот, сдвиг, искажение формы образов или яркостного портрета);
- вычислительная ресурсоемкость.

Инварианты. Один из подходов к распознаванию образов основан на теории инвариантов. Он состоит в том, чтобы сравнивать с моделью не предъявленный образ, а некоторый набор характеристик, инвариантных по отношению к возможным преобразованиям, которые отличают наблюдаемый образ от его модели. Инвариантный подход предполагает три этапа: извлечение числовых характеристик, вычисление инвариантов и корреляция вычисленного инвариантного образа с модельным образом. Одно из преимуществ такого подхода заключается в легкости построения модели образа. Построение модели представляет собой вычисление набора инвариантов по одному предъявленному образу. При этом не требуется проводить измерения над распознаваемым объектом. Другое преимущество состоит в том, что инварианты могут быть использованы как индексы. Измеренное значение инварианта, являясь индексом, позволяет обращаться к модели по этому индексу за одну операцию без необходимости перебора всех имеющихся моделей. Если инвариантов много, индекс многомерный. Время распознавания не зависит от числа моделей. Инвариантное описание объектов является компактным. Основная трудность заключается в том, чтобы для заданного объекта найти инварианты возможных условий его съемки.

Инвариантом называется функция вектора характеристик, которая имеет одно и то же значение для всех преобразований этого вектора характеристик, составляющих группу [2]:

$$I(\vec{f}) = I(\vec{T}(\vec{f})),$$

где I — инвариант; \vec{f} — вектор характеристик; $\vec{T}(\cdot) \in G$ — преобразование вектора характеристик, входящее в группу преобразований G .

Архитектура системы распознавания образов, использующей теорию инвариантов, имеет следующий вид:

1. Выделение характеристик, т. е. контуров, границ, характерных точек.
2. Группирование характеристик, относящихся к одному объекту (сегментация).

3. Вычисление вектора инвариантов по каждой группе характеристик.

4. Формулирование статистической гипотезы о принадлежности наблюдаемого образа соответствующему классу (о близости вычисленного вектора инвариантов к вектору инвариантов, задающему соответствующий класс).

5. Проверка статистической гипотезы против альтернативы.

Спектр Фурье—Меллина. Согласно свойствам преобразования Фурье, спектры F_1 , F_2 изображений f_1 , f_2 связаны уравнением

$$F_2(\omega_x, \omega_y) = \frac{\exp(-j2\pi(\Delta x \omega_x + \Delta y \omega_y))}{s^2} F_1\left(\frac{\omega_x \cos \theta - \omega_y \sin \theta}{s}, \frac{\omega_x \sin \theta + \omega_y \cos \theta}{s}\right). \quad (2)$$

Переход к логарифмически-полярным координатам

$$\begin{cases} \omega_x = \exp(\rho) \cos \varphi; \\ \omega_y = \exp(\rho) \sin \varphi \end{cases}$$

и взятие амплитуды от правой и левой частей уравнения (2) дает

$$\begin{aligned} \left| F_2(e^\rho \cos \varphi, e^\rho \sin \varphi) \right| &= \frac{1}{s^2} \left| F_1(e^{\rho - \ln s} \cos(\theta + \varphi), e^{\rho - \ln s} \sin(\theta + \varphi)) \right|; \\ F_{2-LP}(\rho, \varphi) &= \frac{1}{s^2} F_{1-LP}(\rho - \ln s, \theta + \varphi), \end{aligned} \quad (3)$$

где F_{2-LP} и F_{1-LP} — спектры F_2 и F_1 , представленные в логарифмически-полярных координатах.

Соотношение (3) показывает, что поворот на угол θ и масштабирование с коэффициентом s являются сдвигами вдоль соответствующих осей в логарифмически-полярной системе координат. Взятие преобразования Фурье от функции F_{LP} дает так называемый спектр Фурье—Меллина, инвариантный относительно циклического сдвига, поворота и растяжения (сжатия). Некоторые свойства преобразования Фурье—Меллина [3] указаны в табл. 1.

Таблица 1

Свойства преобразования Фурье—Меллина

| Пространственная область | Область спектра Фурье—Меллина |
|---------------------------|--|
| Поворот на угол φ | Циклический сдвиг на величину $\Delta\theta = \varphi$ |
| Масштабирование в s раз | Сдвиг на величину $\Delta \log \rho = -s$ |

Итак, спектр Фурье—Меллина является инвариантным по отношению к преобразованию (1), если параметры Δx , Δy и s не очень велики, и его можно использовать для распознавания образов на цифровых изображениях [1, 3—8].

Алгоритм решения задачи. Приведем алгоритм решения задачи.

1. Взвешивание отсчетов внутри каждого изображения в соответствии с их удаленностью от центра (аподизация):

$$\hat{a}(x, y) = a(x, y)w(x - x_0, y - y_0), \quad \hat{b}(x, y) = b(x, y)w(x - x_0, y - y_0),$$

$$(x, y) \in \Pi, \quad \Pi = \{(x, y) \in \mathbb{R}^2 : 0 \leq x \leq X, 0 \leq y \leq Y\},$$

$$x_0 = \left\lfloor \frac{X}{2} \right\rfloor, \quad y_0 = \left\lfloor \frac{Y}{2} \right\rfloor,$$

где $\hat{a}(x, y)$, $\hat{b}(x, y)$ — пара взвешенных изображений; $a(x, y)$, $b(x, y)$ — пара исходных изображений; $w(x, y)$ — весовая функция, придающая центральным пикселям изображения больший вес, чем пикселям, расположенным на периферии; (x_0, y_0) — файловые координаты центра каждого изображения; Π — поле зрения; X , Y — ширина и высота исходных изображений, выраженные в пикселях; $\lfloor \cdot \rfloor$ — операция округления вниз до ближайшего целого числа.

В качестве весовой функции можно использовать одну из оконных функций, приведенных в табл. 2:

$$w(n) = c_0 - c_1 \cos \frac{2\pi n}{N-1} + c_2 \cos \frac{4\pi n}{N-1}, \quad n = 0, \dots, N-1.$$

Таблица 2

Некоторые оконные функции

| Оконная функция | c_0 | c_1 | c_2 |
|------------------------------|------------------|---------|------------|
| Прямоугольная | 1 | 0 | 0 |
| Ханна | 0,5 | 0,5 | 0 |
| Хэмминга | 0,53836 | 0,46164 | 0 |
| Блэкмана ($\alpha = 0,16$) | $(1 - \alpha)/2$ | 0,5 | $\alpha/2$ |

2. Умножение каждого изображения на функцию $(-1)^{x+y}$, для того чтобы нулевая частота вычисляемого на следующем этапе спектра Фурье оказалась в центре:

$$\hat{a}^0(x, y) = (-1)^{x+y} \hat{a}(x, y), \quad \hat{b}^0(x, y) = (-1)^{x+y} \hat{b}(x, y).$$

3. Дополнение изображений \hat{a}^0 , \hat{b}^0 нулями во избежание ошибок перекрытия спектров [9] и для приведения размеров изображений к виду 2^n (n — натуральное число).

4. Двумерное быстрое преобразование Фурье (БПФ) изображений пары:

$$A(\omega_x, \omega_y) = F[\hat{a}^0(x, y)], \quad B(\omega_x, \omega_y) = F[\hat{b}^0(x, y)],$$

$$(\omega_x, \omega_y) \in \Xi, \quad \Xi = \{(\omega_x, \omega_y) \in \mathbb{R}^2: -\Omega_x \leq \omega_x \leq \Omega_x, -\Omega_y \leq \omega_y \leq \Omega_y\},$$

где $A(\omega_x, \omega_y)$, $B(\omega_x, \omega_y)$ — спектры изображений; $F[\cdot]$ — оператор БПФ.

5. Фильтрация верхних частот:

$$\bar{A}(\omega_x, \omega_y) = A(\omega_x, \omega_y)H(\omega_x, \omega_y), \quad \bar{B}(\omega_x, \omega_y) = B(\omega_x, \omega_y)H(\omega_x, \omega_y).$$

Например, может быть использован фильтр со следующей передаточной функцией:

$$H(\omega_x, \omega_y) = (1 - \cos \pi \omega_x \cos \pi \omega_y)(2 - \cos \pi \omega_x \cos \pi \omega_y),$$

$$-0,5 \leq \omega_x, \quad \omega_y \leq 0,5.$$

6. Пересчет амплитуды спектра Фурье каждого из изображений в логарифмически-полярную систему координат:

$$A(\rho, \varphi) = |A(\omega_x(\rho, \varphi), \omega_y(\rho, \varphi))|,$$

$$B(\rho, \varphi) = |B(\omega_x(\rho, \varphi), \omega_y(\rho, \varphi))|,$$

$$(\rho, \varphi) \in P, \quad P = \{(\rho, \varphi) \in \mathbb{R}^2: 0 \leq \rho \leq \log_q \sqrt{\Omega_x^2 + \Omega_y^2}, -\pi \leq \varphi < \pi\},$$

$$\begin{cases} \omega_x = q^\rho \cos \varphi; \\ \omega_y = q^\rho \sin \varphi, \end{cases}$$

где q — основание логарифма (выбирается в зависимости от требуемой точности).

Так, при фиксированном исходном размере $R = \sqrt{\Omega_x^2 + \Omega_y^2}$ и заданном коэффициенте изменения числа отсчетов λ для определения q получаем уравнение $\log_q R = \lambda R$.

Алгоритм перехода к логарифмически-полярной системе координат таков. Задаем равномерную прямоугольную сетку на P ; последовательно заполняем ее узлы, вычисляя для каждого узла (ρ, φ) точку $(\omega_x, \omega_y) \in \Xi$ и определяя значения $|A(\omega_x, \omega_y)|$ и $|B(\omega_x, \omega_y)|$ линейной интерполяцией.

7. Повторное двумерное БПФ изображений пары (вычисление спектров Фурье—Меллина):

$$A(m_x, m_y) = F[A(\rho, \varphi)], \quad B(m_x, m_y) = F[B(\rho, \varphi)], \quad (m_x, m_y) \in M,$$

$$M = \{(m_x, m_y) \in \mathbb{R}^2: -M_x \leq m_x \leq M_x, -M_y \leq m_y \leq M_y\}.$$

8. Формулирование и проверка статистической гипотезы о равенстве спектров Фурье—Меллина $A(m_x, m_y)$, $B(m_x, m_y)$ изображений $a(x, y)$, $b(x, y)$. Если различие спектров Фурье—Меллина оказывается статистически значимым, то алгоритм прекращает свою работу — изображения не перекрываются.

9. Перемножение спектров Фурье—Меллина и нормализация произведения:

$$C(m_x, m_y) = \frac{A(m_x, m_y)B^*(m_x, m_y)}{|A(m_x, m_y)B(m_x, m_y)|},$$

где $*$ — знак комплексного сопряжения.

10. Обратное преобразование Фурье от результата:

$$D(\rho, \varphi) = F^{-1}[C(m_x, m_y)].$$

11. Поиск абсолютного максимума изображения $D(\rho, \varphi)$ путем перебора его элементов. Получение оценок угла поворота и коэффициента растяжения (сжатия) второго изображения относительно первого:

$$\{\hat{\theta}, \hat{s}\} = \arg \max_{\rho, \varphi} D(\rho, \varphi).$$

12. Компенсация масштаба и поворота второго изображения.

13. Оценивание сдвига $(\Delta \hat{x}, \Delta \hat{y})^T$ второго изображения относительно первого путем двумерной корреляции (с вычислением свертки

на основе свойств БПФ) компенсированного (на шаге 12) второго изображения с первым.

14. Компенсация сдвига второго изображения относительно первого, если необходимо.

Заключение. Описано решение задачи распознавания образов на цифровых изображениях методами теории инвариантов. В качестве набора инвариантов рассмотрен спектр Фурье—Меллина, слабо изменяющийся при повороте, небольших изменениях масштаба и смещениях образа относительно фона на цифровом изображении. Приведен алгоритм распознавания образа объекта и определения его положения, ориентации и масштаба относительно заданной системы координат на изображении.

СПИСОК ЛИТЕРАТУРЫ

1. Мясников Е. В. Определение параметров геометрических трансформаций для совмещения портретных изображений // Компьютерная оптика. – 2007. – Т. 31, № 3. – С. 77–82.
2. Abdallah S. M. Object recognition via invariance: PhD thesis. Sydney: The University of Sydney, 2000.
3. Агуаласомайяжула Р., Грасси С., Фарине Р. - A. Low complexity RST invariant image recognition using Fourier–Mellin Transform // 19th European Signal Processing Conf. (EUSIPCO 2011), Barcelona, Spain, Aug. 29 – Sept. 2, 2011.
4. Wisetphanichkij S., Dehghan K. Fast fourier transform technique and affine transform estimation-based high precision image registration method // GESTS Int'l Trans. Computer Science and Engr. – 2005. – Vol. 20. – N 1. – P. 179–191.
5. Sheng Y., Arsenaault H. Experiments on pattern recognition using invariant Fourier–Mellin descriptors // J. Opt. Soc. Am. – 1986. – A/Vol. 3. – N 6. – P. 771–776.
6. Reddy B., Chatterji B. An FFT-based technique for translation, rotation, and scale-invariant image registration // IEEE Transactions on Image Processing. Aug. 1996. – Vol. 5. – N 8. – P. 1266–1271.
7. McGuire M. An image registration technique for recovering rotation, scale and translation parameters // NEC Research Institute, Princeton NJ, 1998.
8. Abdelfattah R., Nicolas J., Tupin F. Interferometric SAR image coregistration based on the Fourier–Mellin invariant descriptor // IEEE. – 2002. – P. 1334–1336.
9. Гонсалес Р., Вудс Р. Цифровая обработка изображений. – М.: Техносфера, 2005.

Статья поступила в редакцию 03.07.2012.