

МОДЕЛИ И АЛГОРИТМЫ ИСКУССТВЕННОГО РАЗМНОЖЕНИЯ ДАННЫХ ДЛЯ ОБУЧЕНИЯ АЛГОРИТМОВ РАСПОЗНАВАНИЯ ЛИЦ МЕТОДОМ ВИОЛЫ–ДЖОНСА

А.В. Акимов¹, А.А. Сирота¹

¹Воронежский государственный университет, Воронеж, Россия

Аннотация

Описаны математические модели и алгоритмы внесения деформирующих искажений в решетчатые функции дискретных аргументов. Для изображений рассматриваются три модели внесения деформирующих искажений при размножении обучающей выборки в интересах построения алгоритмов распознавания лиц: на основе применения при деформации гармонических функций, на основе смещения контрольных точек и на основе поиска оптического потока при наложении фильтра энтропии. Для исходных и искусственно размноженных в соответствии с описанными моделями обучающих выборок проведены эксперименты по обучению алгоритмов распознавания изображений лиц методом Виолы–Джонса и сравнению точности работы полученных детекторов. Показана применимость данного подхода для размножения данных при обучении алгоритмов распознавания изображений объектов, характеризующихся определенной структурой.

Ключевые слова: распознавание изображений, деформирующие искажения, интерполяция, оптический поток, энтропия, метод Виолы–Джонса.

Цитирование: Акимов, А.В. Модели и алгоритмы искусственного размножения данных для обучения алгоритмов распознавания лиц методом Виолы–Джонса / А.В. Акимов, А.А. Сирота // Компьютерная оптика. – 2016. – Т. 40, № 6. – С. 911-918. – DOI: 10.18287/2412-6179-2016-40-6-911-918.

Введение

Одной из важных задач, которые приходится решать при построении алгоритмов машинного обучения, является формирование обучающей выборки, содержащей множество образов анализируемых объектов. Их поиск и подготовка к использованию часто требуют значительных затрат времени и ресурсов. В ряде случаев сбор необходимого числа таких образов оказывается сложной или даже невозможной из-за специфического характера предметной области.

Решение данной задачи особенно актуально для систем распознавания объектов на изображениях и, в частности, распознавания лиц, базирующихся на реализации современных методов построения алгоритмов машинного обучения. С учетом необходимости обеспечения представительности обучающих данных одним из возможных подходов является искусственное размножение некоторых «опорных» образов на основе стохастических или детерминистских моделей преобразования.

В ряде исследований, например, [1, 2], метод искусственного размножения применяется при работе с несбалансированными обучающими выборками с малым числом примеров. В данных выборках образы одного класса (необязательно речь идет об изображениях) значительно преобладают над образами другого. При этом классы уравниваются за счет искусственной генерации новых значений признаков входящих в них элементов на основе результата оценки их статистических параметров, проведенной для исходной выборки.

Также известным подходом в задачах распознавания изображений, связанных с обнаружением объектов в естественной среде, является размножение обучающей выборки за счет искажения перспективы или

изменения освещенности, осуществляемых для моделирования различных условий съемки и положений детектируемых объектов в кадре [3].

Целью данной работы является обоснование и исследование различных моделей искусственного размножения обучающих данных на основе внесения деформирующих искажений (ДИ) в цифровые изображения подлежащих детектированию и имеющих явно выраженные структурные характеристики объектов, а также анализ возможностей применения подобного подхода на примере алгоритма распознавания лиц по методу Виолы–Джонса [4, 5].

Данный алгоритм был выбран в качестве инструмента исследования как один из самых надежных и популярных на данный момент детекторов фронтальных лиц [6]. Метод основан на применении комплекса взаимосвязанных подходов и технологий, в центре которых лежит идея построения точного или, как еще говорят, сильного классификатора путем комбинирования других менее точных или слабых классификаторов [7]. Как и большинство других методов машинного обучения, метод Виолы–Джонса требует подготовки большого набора обучающей выборки. В качестве инструмента при проведении исследований авторами был использован разработанный ими в среде Matlab с применением технологии CUDA алгоритм распознавания, содержащий в себе параллельные реализации как процесса обучения, так и процесса принятия решений [8].

1. Математическая модель внесения деформирующих искажений

Математическая модель процесса внесения ДИ для любой исходной функции $f(x_1, \dots, x_n)$ от n переменных определяется следующим соотношением:

$$f(x_1 + \Gamma_1(x_1, \dots, x_n), \dots, x_n + \Gamma_n(x_1, \dots, x_n)) = f(u_1(x_1, \dots, x_n), \dots, u_n(x_1, \dots, x_n)) = g(x_1, \dots, x_n), \tag{1}$$

где $g(x_1, \dots, x_n)$ – результирующая деформированная функция; $\Gamma_i(x_1, \dots, x_n)$, $i = \overline{1, n}$ – функции вносимых деформирующих искажений по каждой координате, которые могут носить детерминированный или стохастический характер. Используя ранее введенные обозначения, а также введя векторы $\mathbf{x} = (x_1, \dots, x_n)^T$, $\mathbf{r}(\mathbf{x}) = (\Gamma_1(\mathbf{x}), \dots, \Gamma_n(\mathbf{x}))^T$ и $\mathbf{u}(\mathbf{x}) = (u_1(\mathbf{x}), \dots, u_n(\mathbf{x}))^T$, можно записать

$$f(\mathbf{x} + \mathbf{r}(\mathbf{x})) = f(\mathbf{u}(\mathbf{x})) = g(\mathbf{x}). \tag{2}$$

При реализации процесса внесения ДИ на основе (1) и (2) возникает ряд особенностей [9]. Первая из них: для того, чтобы значение деформированной функции $g(\mathbf{x}) = f(\mathbf{u}(\mathbf{x}))$ существовало, область ее определения не должна выходить за пределы области определения исходной функции $f(\mathbf{x})$, т.е. если для $f(\mathbf{x})$ вектор $\mathbf{x} \in \Omega_x$, то и вектор $\mathbf{u}(\mathbf{x}) \in \Omega_x$.

Для обеспечения выполнения условия $\mathbf{u}(\mathbf{x}) \in \Omega_x$ применимы различные подходы. Один из возможных вариантов состоит в том, чтобы при выходе какой-либо компоненты координат в $\mathbf{x} + \mathbf{r}(\mathbf{x}) = \mathbf{u}(\mathbf{x})$ из (2) за границы исходной области определения Ω_x в качестве значения данной компоненты $\mathbf{u}(\mathbf{x})$ при расчете $f(\mathbf{u}(\mathbf{x}))$ подставлять граничное значение соответствующей координаты исходной функции $f(\mathbf{x})$. Преимущество данного подхода состоит в том, что при его применении не ограничиваются значения функции деформации $\mathbf{r}(\mathbf{x})$ в случае, когда $\mathbf{x} + \mathbf{r}(\mathbf{x}) = \mathbf{u}(\mathbf{x})$ находится в пределах Ω_x . Именно этот подход был использован в нашей работе.

Вторая особенность процесса внесения ДИ на основе (1) и (2) состоит в следующем. Если исходная функция $\hat{f}(\mathbf{x})$ является решетчатой функцией дискретных аргументов, т.е. задана на многомерной дискретной сетке (например, в случае внесения ДИ в цифровые изображения), прямое применение (1) и (2) невозможно; и для обеспечения возможности модификации аргументов $\hat{f}(\mathbf{x})$ при внесении в них ДИ произвольного характера необходимо предварительно выполнить ее интерполяцию и представление в виде функции $\tilde{f}(\mathbf{x})$ непрерывнозначных аргументов. При этом общая процедура будет соответствовать схеме, представленной на рис. 1.

Утверждение. Пусть: $f(\mathbf{x})$ – функция, заданная на множестве значений n непрерывных аргументов Ω_x ; $\hat{f}(\mathbf{x})$ – решетчатая функция, сформированная на многомерной дискретной сетке, покрываемой Ω_x ; $\tilde{f}(\mathbf{x})$ – функция, интерполированная на основе значений $\hat{f}(\mathbf{x})$ и определенная на $\tilde{\Omega}_x \subset \Omega_x$; $g(\mathbf{x})$ – деформированная функция на основе внесения ДИ $\mathbf{r}(\mathbf{x})$ в $f(\mathbf{x})$ по формуле (1); $\tilde{g}(\mathbf{x})$ – деформированная аналогичным образом функция на основе внесения ДИ $\mathbf{r}(\mathbf{x})$ в $\tilde{f}(\mathbf{x})$;

X – некое связное подмножество значений непрерывных аргументов $\mathbf{x} \in X \subset \tilde{\Omega}_x$. Пусть также для любого $\mathbf{x} \in X$ выполняется $\mathbf{u}(\mathbf{x}) \in X$.

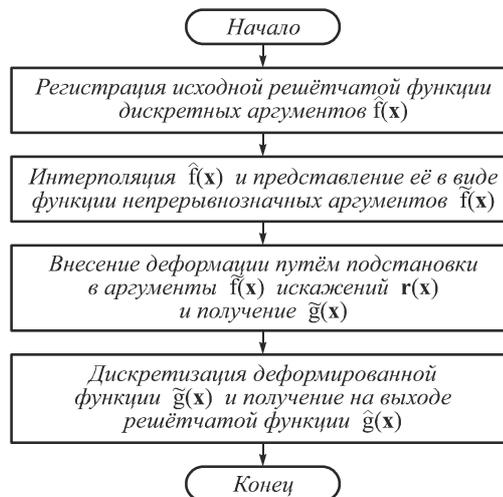


Рис. 1. Схема процесса внесения ДИ в решетчатую функцию

Тогда ошибка $|g(\mathbf{x}) - \tilde{g}(\mathbf{x})|$, возникающая при внесении деформации в интерполированную функцию $\tilde{f}(\mathbf{x})$, для $\mathbf{x} \in X \subset \tilde{\Omega}_x$ не превысит максимальную ошибку интерполяции $|f(\mathbf{x}) - \tilde{f}(\mathbf{x})|$.

Доказательство. Исходя из сформулированных условий, имеем, что $g(\mathbf{x}) = f(\mathbf{u}(\mathbf{x}))$ и $\tilde{g}(\mathbf{x}) = \tilde{f}(\mathbf{u}(\mathbf{x}))$. Вычтем эти два выражения друг из друга и возьмем абсолютные значения разностей в левой и правой части:

$$\Delta(\mathbf{x}) = |g(\mathbf{x}) - \tilde{g}(\mathbf{x})| = |f(\mathbf{u}(\mathbf{x})) - \tilde{f}(\mathbf{u}(\mathbf{x}))|.$$

Пусть U – множество всех значений, принимаемых $\mathbf{u}(\mathbf{x})$ внутри X , т.е. $\mathbf{u}(\mathbf{x}) \in U \subset X$. Таким образом, функции $g(\mathbf{x}) = f(\mathbf{u}(\mathbf{x}))$ и $\Delta(\mathbf{x})$ определены на множестве U . При этом функция $|f(\mathbf{x}) - \tilde{f}(\mathbf{x})|$ определена на множестве X . Возможны две ситуации.

Первая из них состоит в том, что функция искажения $\mathbf{r}(\mathbf{x})$ может оказаться такой, что для некоторого аргумента $\mathbf{x}^{(j)} \in X$ не найдется такого значения из U , что $\mathbf{u}(\mathbf{x}) = \mathbf{x} + \mathbf{r}(\mathbf{x}) = \mathbf{x}^{(j)}$ (данный элемент не будет использован при деформации). Действительно, достаточно, например, взять $\mathbf{r}(\mathbf{x}) = -\mathbf{x} + \mathbf{x}^{(k)}$, для некоторого $\mathbf{x}^{(k)} \in X$ такого, что $\mathbf{x}^{(k)} \neq \mathbf{x}^{(j)}$. Тогда:

$$\mathbf{u}(\mathbf{x}) = \mathbf{x} + \mathbf{r}(\mathbf{x}) = \mathbf{x} - \mathbf{x} + \mathbf{x}^{(k)} = \mathbf{x}^{(k)} \neq \mathbf{x}^{(j)}.$$

В этом случае множество X/U не пустое. Среди его элементов могут оказаться такие, что $|f(\mathbf{x}) - \tilde{f}(\mathbf{x})|$ будет принимать свое максимальное в X значение. Тогда $\max |g(\mathbf{x}) - \tilde{g}(\mathbf{x})| < \max |f(\mathbf{x}) - \tilde{f}(\mathbf{x})|$, иначе $\max |g(\mathbf{x}) - \tilde{g}(\mathbf{x})| = \max |f(\mathbf{x}) - \tilde{f}(\mathbf{x})|$.

Во втором случае, если для данного искажения $\mathbf{r}(\mathbf{x})$ для любого $\mathbf{x}^{(j)} \in X$ найдется $\mathbf{x} \in X$ такое, что

$\mathbf{x}^{(j)} = \mathbf{u}(\mathbf{x})$, то $U = X$, так как каждому $\mathbf{x} \in X$ соответствует свое уникальное значение $\mathbf{u}(\mathbf{x}) \in U$, при том, что по определению $U \subset X$, и в этом случае $\max |g(\mathbf{x}) - \tilde{g}(\mathbf{x})| = \max |f(\mathbf{x}) - \tilde{f}(\mathbf{x})|$.

Аналогичным образом можно доказать, что

$$\min |g(\mathbf{x}) - \tilde{g}(\mathbf{x})| \geq \min |f(\mathbf{x}) - \tilde{f}(\mathbf{x})|.$$

Следствие. Ошибка $|g(\mathbf{x}) - \tilde{g}(\mathbf{x})|$, возникающая при деформации интерполированной функции $\tilde{f}(\mathbf{x})$ с областью определения $\tilde{\Omega}_x$, не выйдет за пределы максимального и минимального значения ошибки интерполяции $|f(\mathbf{x}) - \tilde{f}(\mathbf{x})|$ при условии, что $\mathbf{u}(\mathbf{x}) \in \tilde{\Omega}_x$. Для подтверждения этого следствия достаточно повторить предыдущее доказательство в случае $X = \tilde{\Omega}_x$.

Таким образом, при внесении ДИ достаточно выполнить $\mathbf{u}(\mathbf{x}) \in \tilde{\Omega}_x$, что производится при помощи того же описанного выше подхода, что и при обеспечении условия $\mathbf{u}(\mathbf{x}) \in \Omega_x$. При этом $\mathbf{u}(\mathbf{x}) \in \tilde{\Omega}_x$ автоматически означает, что $\mathbf{u}(\mathbf{x}) \in \Omega_x$, так как $\tilde{\Omega}_x \subset \Omega_x$.

Внесение ДИ в рассмотренной схеме может также производиться и обратным образом:

$$f(\mathbf{x}_1) = g(\mathbf{u}^{-1}(\mathbf{x}_1)) = g(\mathbf{x}_1 + \mathbf{r}^{-1}(\mathbf{x}_1)),$$

где $\mathbf{r}^{-1}(\mathbf{x}_1)$ – функция деформации, обратная к $\mathbf{r}(\mathbf{x}_2)$, и \mathbf{x}_1 с \mathbf{x}_2 такие, что

$$f(\mathbf{x}_2 + \mathbf{r}(\mathbf{x}_2)) = f(\mathbf{u}(\mathbf{x}_2)) = g(\mathbf{x}_2),$$

$$\mathbf{x}_1 = \mathbf{x}_2 + \mathbf{r}(\mathbf{x}_2), \quad \mathbf{x}_2 = \mathbf{x}_1 + \mathbf{r}^{-1}(\mathbf{x}_1).$$

В некоторых публикациях данную технологию можно обнаружить под названием «варпинг» (warping) [10]. Он используется при «морфинге» (morphing) – процессе постепенного замещения одного изображения другим путем плавных изменений положений, схожих по некоему признаку частей этих двух изображений и интенсивностей их пикселей. Основной областью применения подобных технологий является компьютерная графика в кино и анимации.

2. Модели и алгоритмы внесения деформирующих искажений с целью размножения обучающей выборки

При проведении исследования нами были рассмотрены три модели внесения ДИ с целью искусственного размножения обучающих данных.

Первая модель внесения ДИ основана на применении в качестве функции деформации $\mathbf{r}(\mathbf{x}) = \mathbf{r}(x, y) = \{r_i(x, y), i = \overline{1, 2}\}$, гармонических функций (ГФ) вида

$$r_i(x, y) = A_i \sin(\omega_{xi}x + \omega_{yi}y + \psi_i),$$

где A_i – амплитуда смещения при внесении ДИ, ω_{xi} и ω_{yi} – частоты функции деформации по каждой из осей координат изображения и ψ_i – фаза.

Значение амплитуды A_i ограничивает максимальную величину смещения координат пикселей при внесении ДИ и определяется тем, насколько можно независимо друг от друга сместить черты лица относительно их исходных позиций так, чтобы не разрушить его общую структуру.

Частоты ω_{xi} и ω_{yi} определяют размеры деталей изображения, которые будут деформироваться. При малых значениях осуществляется смещение и плавная деформация изображения объекта в целом с сохранением его структурных свойств, а при их увеличении – деформируются все более мелкие его части. Исходя из этих соображений используемые в экспериментах значения ω_{xi} и ω_{yi} подбирались так, чтобы период функции деформации $r_i(x, y)$ примерно соответствовал размерам основных черт лица на изображении.

Фаза функции деформации ψ_i определяет ее смещение, и для обеспечения возможности получения нескольких разных деформированных изображений на основе одного экземпляра исходной обучающей выборки ее значение генерируется случайным образом по равномерному закону на всей области определения $\tilde{\Omega}_x$.

В итоге реализуемый в соответствии с данной моделью алгоритм состоял в том, что для каждого из изображений исходной обучающей выборки заданное число раз с учетом описанных выше соображений генерировались параметры функции деформации, и на их основе производилось внесение ДИ с целью получить новые изображения.

На рис. 2 показаны примеры результатов внесения ДИ такого типа в изображения лиц, которые могут использоваться в качестве обучающей выборки в методе Виолы–Джонса.

Вторая модель внесения ДИ основана на часто применяемом при реализации морфинга [10, 11] подходе, заключающемся в ручной расстановке соответствий на двух совмещаемых изображениях в виде контрольных точек или контуров. Далее, на основе известных для отмеченных элементов попарных смещений или значений функции деформации в данных точках, при помощи интерполяции или некоторого другого метода находят все остальные значения функции деформации. Одним из используемых при таком подходе способов интерполяции является применение радиально-базисных функций (РБФ) [12].



Рис. 2. Примеры внесения ДИ в изображения лиц с применением гармонической функции деформации (первый столбец слева является исходным)

Реализуемый в соответствии с данной моделью алгоритм состоит в следующем. Для каждого из изображений подлежащей размножению обучающей выборки сначала производится расстановка контрольных точек (x_k, y_k) , $k = \overline{1, m}$, где число контрольных точек $m = 5$: по одной на глаз, кончик носа и уголки рта (рис. 3).



Рис. 3. Примеры расстановки контрольных точек на изображениях лиц

Далее рассчитываются средние значения (x_k, y_k) и их матрица ковариации, которые используются при генерации случайных положений контрольных точек для новых, искусственно получаемых данных. После этого внесение ДИ в одно из изображений исходной обучающей выборки производится следующим образом.

Генерируется случайный набор координат контрольных точек (x_k, y_k) в соответствии с полученными ранее статистическими параметрами их распределения (при генерации вместо средних значений также могут подаваться координаты самих контрольных точек размножаемого изображения). Для каждой из этих точек на основе (1) и (2) рассчитываются значения функции деформации $r_i(x_k, y_k)$, $i = \overline{1, 2}$, в данной точке так, чтобы

$$f(x_k + r_1(x_k, y_k), y_k + r_2(x_k, y_k)) = f(\mathbf{u}(x_k, y_k)) = g(x_k, y_k)$$

где $f(x, y)$ – деформируемое изображение, а $g(x, y)$ – результат деформации, соответствующий набору случайно сгенерированных положений контрольных точек. При этом $(x_k + r_1(x_k, y_k))$, $y_k + r_2(x_k, y_k)$ – координаты контрольных точек деформируемого изображения обучающей выборки.

В данной модели конечные функции деформации $r_i(x_k, y_k)$, $i = \overline{1, 2}$, представляются в виде сумм РБФ-функций:

$$r_i(x, y) = \sum_{k=1}^m a_{ki} \exp\left(-\frac{(x-x_k)^2 + (y-y_k)^2}{2\sigma^2}\right), \quad (3)$$

где σ – ширина окна RBF-функций, a_{ki} – неизвестные коэффициенты отображения. Элементам суммы соответствуют RBF-функции с центрами по координатам (x_k, y_k) .

Если для x и y в координатах контрольных точек (x_k, y_k) , $k = \overline{1, m}$, вместо $r_i(x, y)$, поочередно подставлять в (3) известные значения $r_i(x_k, y_k)$, то для каждой из функций деформации, соответствующей смещению по координатам x или y , получится по системе из m линейных уравнений (число неизвестных также равно m):

$$\sum_{l=1}^m \exp\left(-\frac{(x_k-x_l)^2 + (y_k-y_l)^2}{2\sigma^2}\right) a_{li} = r_i(x_k, y_k),$$

при $k = \overline{1, m}$.

В матричном виде это эквивалентно следующему уравнению:

$$\mathbf{E}\mathbf{A}_i = \mathbf{R}_i,$$

где матрица \mathbf{E} состоит из элементов

$$\sum_{l=1}^m \exp\left(-\frac{(x_k-x_l)^2 + (y_k-y_l)^2}{2\sigma^2}\right),$$

со столбцами, пронумерованными значениями индексов $l = \overline{1, m}$, и строками, пронумерованными значениями $k = \overline{1, m}$, вектор $\mathbf{A}_i = (a_{i1}, \dots, a_{im})^T$ содержит неизвестные коэффициенты РБФ-функций, и вектор $\mathbf{R}_i = (r_i(x_1, y_1), \dots, r_i(x_m, y_m))^T$ – известные значения функции деформации соответствующей компоненты изображения в контрольных точках. При этом матрица \mathbf{E} одинакова для всех $r_i(x_k, y_k)$, $i = \overline{1, 2}$, и не зависит от компоненты изображения.

В простейших вариантах значение σ определяется экспериментально путем перебора в заданном диапазоне значений или для каждой контрольной точки и соответствующей ей РБФ-функции выставляется равным расстоянию от нее до ближайшей к ней другой контрольной точки.

Получаемые таким образом системы уравнений не всегда могут иметь стандартное решение $\mathbf{A}_i = \mathbf{E}^{-1}\mathbf{R}_i$ из-за плохой обусловленности матрицы уравнения или в ситуации, когда число контрольных точек меньше числа используемых для интерполяции функций. В этих случаях применяют методы регуляризации. В большинстве случаев используют расчет псевдообратной матрицы, тогда $\mathbf{A}_i = \mathbf{E}^+\mathbf{R}_i$, где \mathbf{E}^+ – псевдообратная матрица для \mathbf{E} , или получают решение на основе регуляризации по А.Н. Тихонову

$$\mathbf{A}_i = (\lambda\mathbf{I} + \mathbf{E}^T\mathbf{E})^{-1}\mathbf{E}^T(\mathbf{R}_i - \mathbf{E}\mathbf{A}_{i0}),$$

где \mathbf{I} – единичная матрица размером $m \times m$, λ – параметр регуляризации, \mathbf{A}_{i0} – априорная оценка решения) [13].

На рис. 4 показаны примеры результатов внесения ДИ на основе описанного подхода.



Рис. 4. Примеры внесения ДИ в изображения лиц на основе расстановки контрольных точек и интерполяции функции деформации при помощи РБФ (первый столбец слева является исходным)

Третья модель основана на идеях определения оптического потока (ОП) в процессе изменения положений и ориентации объектов сцены в разные моменты времени или при разных условиях съемки [14,

15]. Вычисление оптического потока состоит в оценке по координатам x и y смещений содержимого сцены, которые и предлагается напрямую использовать в качестве функций деформации $r_i(x, y), i=1,2$. Существует множество механизмов определения оптического потока [15 – 17]. Каждый из них ограничен своей областью применимости из-за заложенных в него допущений, на основе которых и производится его расчет. Выбранный для данной модели метод Фарнебака [17] работает следующим образом. Оба изображения (предварительно сглаженные) представляются в виде квадратичных полиномов вида:

$$f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{b} \mathbf{x} + c,$$

где $\mathbf{x}=(x, y)^T$, и $\mathbf{A}, \mathbf{b}, c$ – параметры полинома, значения которых во время его построения вычисляются в соответствии с критерием наименьших квадратов: \mathbf{A} – симметричная матрица, \mathbf{b} – вектор, c – скаляр. При расчете оптического потока между кадрами $f_1(\mathbf{x})$ и $f_2(\mathbf{x})$ ищется смещение \mathbf{d} , удовлетворяющее уравнениям:

$$f_2(\mathbf{x}) = f_1(\mathbf{x} - \mathbf{d}) = (\mathbf{x} - \mathbf{d})^T \mathbf{A}_1 (\mathbf{x} - \mathbf{d}) + \mathbf{b}_1 (\mathbf{x} - \mathbf{d}) + c_1 = \mathbf{x}^T \mathbf{A}_2 \mathbf{x} + \mathbf{b}_2 \mathbf{x} + c_2.$$

Для повышения точности используемого приближения построение полинома при этом производится не для всего изображения, а для некоей его области, и, таким образом, все коэффициенты $\mathbf{A}, \mathbf{b}, c$ и смещение \mathbf{d} становятся зависимыми от координат $\mathbf{x}=(x, y)^T$: $\mathbf{A}(\mathbf{x}), \mathbf{b}(\mathbf{x}), c(\mathbf{x})$ и $\mathbf{d}(\mathbf{x})$. Особенностью метода Фарнебака является возможность вычисления потоков большой величины всего для двух соседних по времени кадров за счет итерационного характера расчета $\mathbf{d}(\mathbf{x})$ в разных масштабах от меньшего к большему с постепенным уточнением получающихся значений [17].

Из-за специфики предметной области – обработка изображений лиц, разницу между которыми нельзя увязать к смещению положений их пикселей – перед процедурой поиска оптического потока требуется провести дополнительное преобразование. Им является операция наложения фильтра энтропии [18], который для каждой координаты изображения и ее окрестности из n пикселей вычисляется следующим образом. Рассчитываются вероятности $p_i, i=1, n$, появления каждого значения интенсивности пикселя в данной области, и на их основе – энтропия ее центрального пикселя:

$$e(x, y) = -\sum_{i=1}^n p_i \log p_i,$$

где $e(x, y)$ – результат наложения фильтра, который и используется для поиска оптического потока.

Таким образом, изображения лиц заменяются образами, отражающими их основную структуру. В них мелкие черты лица соответствуют высоким значениям энтропии, а фон и кожа лба и щек – более низким (рис. 5).

В итоге размножение изображений обучающей выборки осуществлялось следующим образом. На каждое из них производилось наложение фильтра энтропии и вычисление $e(x, y)$. После чего рассчитыва-

лись выборочные параметры энтропии: среднее и матрица ковариации.



Рис. 5. Примеры наложения фильтра энтропии на изображения лиц

Далее, при размножении каждого изображения исходной обучающей выборки производилась генерация случайных значений энтропии $e_g(x, y)$ и расчет функции деформации $r_i(x, y), i=1,2$, путем вычисления оптического потока между сгенерированным значением энтропии и энтропией размножаемого изображения. Полученный результат $r_i(x, y), i=1,2$, и использовался при внесении ДИ в качестве функции деформации. На рис. 6 показаны примеры работы алгоритма, основанного на данной модели.



Рис. 6. Примеры внесения ДИ в изображения лиц на основе наложения фильтра энтропии и расчета ОП (первый столбец слева является исходным)

3. Результаты экспериментальных исследований

Для оценки влияния искусственного размножения элементов обучающей выборки на качество детектирования были проведены эксперименты по обучению детекторов лиц по методу Виолы–Джонса с последующими замерами точности их работы. Обучение было осуществлено с использованием пяти вариантов формирования обучающих выборок:

- оригинальная, содержащая 4916 изображений лиц, собранных и подготовленных вручную авторами метода Виолы–Джонса [4, 5];
- усеченная, содержащая 492 изображения лиц из числа изображений оригинальной выборки;
- три искусственно размноженные выборки, содержащие по 4916 изображений лиц, из которых 492 изображения принадлежали усеченному набору изображений, а остальные (по 9 на каждое из исходных) сгенерированы искусственно, путем внесения ДИ на основе каждой из описанных выше моделей.

Оценка качества детектирования производилось следующим образом. Для каждого детектора модифицировался порог его реагирования, и в соответ-

ствии с полученными откликами были построены зависимости относительной частоты правильно найденных лиц от числа ложных срабатываний (рис. 7). В качестве изображений для тестирования был использован специальный набор изображений из выборки CMU+MIT [19], часто применяемый для сравнения работы детекторов фронтальных лиц и содержащий 42 изображения со 170 лицами.

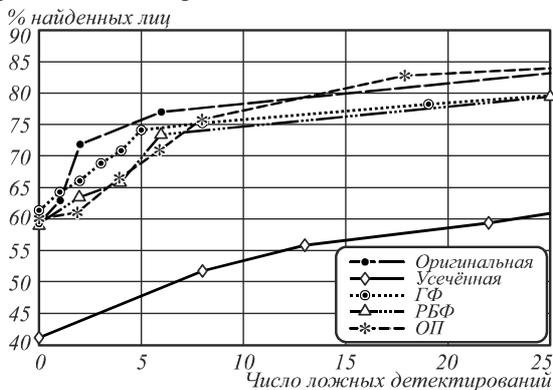


Рис. 7. Зависимости для относительной частоты правильно найденных лиц от числа ложных срабатываний, полученные при обучении по оригинальной и усеченной выборкам, а также с использованием трех рассмотренных моделей размножения обучающих данных

В табл. 1 представлены типичные результаты, полученные для каждого из детекторов при схожем числе ложных срабатываний.

Табл. 1. Относительные частоты правильно найденных лиц при схожем числе ложных срабатываний, полученные при обучении по оригинальной и усеченной выборкам, а также с использованием трех рассмотренных моделей размножения обучающих данных

	Оригинальная	Усеченная	ГФ	РБФ	ОП
Процент найденных лиц	83,53	59,41	78,23	79,41	82,94
Число ложных срабатываний	26	22	19	25	18

Анализ данных табл. 1 и полученных зависимостей (рис. 7) показывает, что при схожем числе ложных срабатываний детектор, обученный с использованием усеченной выборки, стабильно выдает результаты хуже оригинального примерно на 20%. В то же время детекторы, основанные на трех рассмотренных выше моделях размножения элементов обучающих данных путем внесения в них ДИ, дают сравнимый с оригиналом результат с отклонением от него по большей части в пределах 5%. Таким образом, показано значительное увеличение точности их работы при сравнении с классификатором, основанным на усеченной обучающей выборке (которая и была использована для размножения).

Заключение

Полученные результаты позволяют говорить о применимости данного подхода к задачам построения

детекторов обнаружения объектов, имеющих ярко выраженные структурные характеристики, для сокращения затрат на поиск и подготовку обучающих данных, а также в условиях невозможности сбора достаточного для начала процесса обучения количества элементов тренировочной выборки.

Литература

1. **Guo, H.** Learning from Imbalanced data sets with boosting and data generation: the DataBoost IM approach / H. Guo, H.L. Viktor // ACM SIGKDD Explorations Newsletter. – 2004. – Vol. 6(1). – P. 30-39. – DOI: 10.1145/1007730.1007736.
2. **Chawla, N.** SMOTE: synthetic minority over-sampling technique / N. Chawla, K. Bowyer, L. Hall, W. Kegelmeyer // Journal of Artificial Intelligence Research. – 2002. – Vol. 16(1). – P. 321-357. – DOI: 10.1613/jair.953.
3. **Жуковский, А.Е.** Синтез обучающей выборки на основе реальных данных в задачах распознавания изображений / А.Е. Жуковский, С.А. Усилин, Н.А. Тарасова, Д.П. Николаев // Информационные технологии и системы (ИТиС'12): сборник трудов конференции. – М., 2012. – С. 377-382.
4. **Viola, P.** Rapid object detection using a boosted cascade of simple features / P. Viola, M. Jones // Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2001. – 2001. – Vol. 1. – P. 511-518. – DOI: 10.1109/CVPR.2001.990517.
5. **Viola, P.** Robust real-time face detection / P. Viola, M. Jones // International Journal of Computer Vision. – 2004. – Vol. 57(2). – P. 137-154. – DOI: 10.1023/B:VISI.0000013087.49260.fb.
6. **Калиновский, И.А.** Обзор и тестирование детекторов фронтальных лиц / И.А. Калиновский, В.Г. Спицын // Компьютерная оптика. – 2016. – Т. 40, № 1. – С. 99-111. – DOI: 10.18287/2412-6179-2016-40-1-99-111.
7. **Freund, Y.** A short introduction to boosting / Y. Freund, R. Schapire // Journal of Japanese Society for Artificial Intelligence. – 1999. – Vol. 14(5). – P. 771-780.
8. **Акимов, А.В.** Разработка и исследование алгоритмов распознавания изображений на основе метода Виолы-Джонса с использованием технологии вычислений на графических процессорах CUDA / А.В. Акимов, А.А. Сирота // Вестник ВГУ, Серия: Системный анализ и информационные технологии. – 2014. – № 3. – С. 100-108.
9. **Акимов, А.В.** Модели и алгоритмы внесения деформирующих искажений на изображениях с использованием радиально-базисных функций / А.В. Акимов, М.А. Дрюченко, А.А. Сирота // Вестник ВГУ, Серия: Системный анализ и информационные технологии. – 2014. – № 1. – С. 130-137.
10. **Wolberg, G.** Image morphing: a survey / G. Wolberg // The Visual Computer. – 1998. – Vol. 14(8). – P. 360-372. – DOI: 10.1007/s003710050148.
11. **Steyvers, M.** Morphing techniques for manipulating face images / M. Steyvers // Behavior Research Methods, Instruments, & Computers. – 1999. – Vol. 31(2). – P. 359-369. – DOI: 10.3758/BF03207733.
12. **Arad, N.** Image warping by radial basis functions: applications to facial expressions / N. Arad, N. Dyn, D. Reissfeld, Y. Yeshurun // CVGIP: Graph Models Image Processing. – 1994. – Vol. 56, Issue 2. – P. 161-172. – DOI: 10.1006/cgip.1994.1015.
13. **Сизиков, В.С.** Устойчивые методы обработки результатов измерений: Учебное пособие / В.С. Сизиков. – СПб.: СпецЛит, 1999. – 240 с.

14. **Brown, L.G.** A survey of image registration techniques / L.G. Brown // ACM Computing Surveys. – 1992. – Vol. 24(4). – P. 325-376. – DOI: 10.1145/146370.146374.
15. **Barron, J.** Performance of optical flow techniques / J. Barron, D. Fleet, S. Beauchemin // International Journal of Computer Vision. – 1994. – Vol. 12(1). – P. 43-77. – DOI: 10.1007/BF01420984.
16. **Horn, B.** Determining optical flow / B. Horn, B. Schunk // Artificial Intelligence. – 1981. – Vol. 17(1-3). – P. 185-203. – DOI: 10.1016/0004-3702(81)90024-2.
17. **Farneback, G.** Two-frame motion estimation based on polynomial expansion / G. Farneback // Proceedings of the 13th Scandinavian Conference on Image Analysis, Halmstad, Sweden, June 29 – July 02, 2003. – 2003. – P. 363-370. – DOI: 10.1007/3-540-45103-X_50.
18. **Gonzalez, R.C.** Digital image processing using MATLAB / R.C. Gonzalez, R.E. Woods, S.L. Eddins. – 2nd ed. – New Jersey: Prentice Hall. – 2009. – 826 p. – ISBN: 978-0982085400.
19. CMU/VASC image database: Frontal face images [Electronic Resource]. – URL: http://vasc.ri.cmu.edu/idb/html/face/frontal_images/index.html (дата обращения 30.10.2016).

Сведения об авторах

Акимов Алексей Викторович, 1990 года рождения, в 2013 году окончил магистратуру Воронежского государственного университета по специальности «Информационные системы и технологии». Аспирант кафедры технологий обработки и защиты информации Воронежского государственного университета. Область научных интересов: распознавание изображений, машинное обучение. E-mail: akimov@rcnit.vsu.ru.

Сирота Александр Анатольевич, 1954 года рождения, в 1976 году окончил Воронежский государственный университет по специальности «Радиофизика и электроника». Доктор технических наук (1995 год), профессор, заведует кафедрой технологий обработки и защиты информации Воронежского государственного университета. Область научных интересов: синтез и анализ систем сбора и обработки информации, методы и технологии компьютерного моделирования информационных процессов и систем, системный анализ в сфере информационной безопасности, компьютерная обработка изображений, нейронные сети и нейросетевые технологии в системах принятия решений. E-mail: sir@cs.vsu.ru.

ГРПТИ: 28.23.15

Поступила в редакцию 15 августа 2016 г. Окончательный вариант – 3 ноября 2016 г.

SYNTHETIC DATA GENERATION MODELS AND ALGORITHMS FOR TRAINING IMAGE RECOGNITION ALGORITHMS USING THE VIOLA-JONES FRAMEWORK

A.V. Akimov¹, A.A. Sirota¹

¹ Voronezh State University, Voronezh, Russia

Abstract

The paper describes mathematical models and algorithms of warping grid functions with discrete parameters. For images, three warping models, applied to the generation of extra training data to build face recognition algorithms, are examined: the one based on harmonic functions, the one based on offsetting user-specified control point coordinates, and the one based on the computation of the optical flow between entropy-filtered images. For the training sets, both initial and those synthetically generated using the above three models, learning of face detection algorithms based on the Viola-Jones framework was performed and corresponding detection rates were compared. It is shown that this approach is applicable for synthetic data generation when training image recognition algorithms for recognition of objects characterized by inherent structure.

Keywords: image recognition, warping, interpolation, RBF, optical flow, entropy, Viola-Jones framework.

Citation: Akimov AV, Sirota AA. Synthetic data generation models and algorithms for training image recognition algorithms using the Viola-Jones framework. Computer Optics 2016; 40(6): 911-918. DOI: 10.18287/2412-6179-2016-40-6-911-918.

References

- [1] Guo H, Viktor HL. Learning from imbalanced data sets with boosting and data generation: the DataBoost IM approach. ACM SIGKDD Explorations Newsletter 2004; 6(1): 30-39. DOI: 10.1145/1007730.1007736.
- [2] Chawla N, Bowyer K, Hall L, Kegelmeyer W. SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research 2002; 16(1): 321-357. DOI: 10.1613/jair.953.
- [3] Zhukovsky A, Usilin S, Tarasova N, Nikolaev D. Synthetic training sets based on real data in the problems of image recognition [In Russian]. Proceedings of the Conference “Information Technology and Systems (ITaS’12)”, Moscow 2012; 377-382.
- [4] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2001; 1: 511-518. DOI: 10.1109/CVPR.2001.990517.
- [5] Viola P, Jones M. Robust real-time face detection. International Journal of Computer Vision 2004; 57(2): 137-154. DOI: 10.1023/B:VISI.0000013087.49260.fb.

- [6] Kalinovskii IA, Spitsyn VG. Review and testing of frontal face detectors. *Computer Optics* 2016; 40(1): 99-111. DOI: 10.18287/2412-6179-2016-40-1-99-111.
- [7] Freund Y, Schapire R. A Short Introduction to Boosting. *Journal of Japanese Society for Artificial Intelligence* 1999; 14(5): 771-780.
- [8] Akimov AV, Sirota AA. Design and analysis of algorithms for image recognition based on the method of Viola-Jones using computing technology on GPU CUDA [In Russian]. *Vestnik VSU, Series: System Analysis and Information Technology* 2014; 3: 100-108.
- [9] Akimov AV, Dryuchenko MA, Sirota AA. Models and algorithms for making distorting distortion in images using radial basis functions [In Russian]. *Vestnik VSU, Series: System Analysis and Information Technology* 2014; 1: 130-137.
- [10] Wolberg G. Image Morphing Survey. *The Visual Computer* 1998; 14(8): 360-372. DOI: 10.1007/s003710050148.
- [11] Steyvers M. Morphing Techniques for Manipulating Face Images. *Behavior Research Methods, Instruments, & Computers* 1999; 31(2): 359-369. DOI: 10.3758/BF03207733.
- [12] Arad N, Dyn N, Reisfeld D, Yeshurun Y. Image warping by radial basis functions: applications to facial expressions. *CVGIP: Graph Models Image Processing* 1994; 56(2): 161-172. DOI: 10.1006/cgip.1994.1015.
- [13] Sizikov VS. Robust methods of processing measurement results. Tutorial [In Russian]. Saint-Petersburg: "SpecLit" Publisher; 1999.
- [14] Brown LG. A survey of image registration techniques. *ACM Computing Surveys* 1992; 24(4): 325-376. DOI: 10.1145/146370.146374.
- [15] Barron J, Fleet D, Beauchemin S. Performance of Optical Flow Techniques. *International Journal of Computer Vision* 1994; 12(1): 43-77. DOI: 10.1007/BF01420984.
- [16] Horn B, Schunk B. Determining Optical Flow. *Artificial Intelligence* 1981; 17(1-3): 185-203. DOI: 10.1016/0004-3702(81)90024-2.
- [17] Farneback, G. Two-Frame Motion Estimation Based on Polynomial Expansion. *Proc SCIA'03* 2003; 363-370. DOI: 10.1007/3-540-45103-X_50.
- [18] Gonzalez RC, Woods RE, Eddins SL. Digital image processing using MATLAB. 2nd ed. New Jersey: Prentice Hall; 2009. ISBN: 978-0982085400.
- [19] CMU/VASC image database: Frontal face images. Source: (http://vasc.ri.cmu.edu/idb/html/face/frontal_images/index.html).

Authors' information

Alexey Viktorovich Akimov (b. 1990) graduated from Voronezh State University in 2013 with master's degree in Information Systems and Technology. Currently postgraduate student at Information Processing and Security Technologies chair at Voronezh State University. Research interests: image recognition, machine learning. E-mail: akimov@rcnit.vsu.ru.

Alexander Anatolievich Sirota (b. 1954) graduated from Voronezh State University in 1976 majoring in "Radiophysics and Electronics". Professor, Doctor of Technical Sciences (since 1995). Currently head of Information Processing and Security Technologies chair at Voronezh State University. Research interests: analysis and design of information collection and processing systems, methods and techniques of information processes and systems computer modeling, system analysis in information security, digital image processing, neural networks and neural network technologies in decision-making systems. E-mail: sir@cs.vsu.ru.

Received August 15, 2016. The final version – November 3, 2016.
