

М.А. Тищенко

Генеративная модель мимики человеческого лица

Предложена технология автоматического построения генеративной модели мимики человеческого лица, позволяющая синтезировать различные выражения лица на основе нескольких опорных его фотоснимков. Каждое выражение лица в предложенном подходе описывается как деформация некоторого базового его фотоснимка. Поиск соответствующей деформации выполняется путем решения (max, +)-задачи разметки.

A technology for the automatic construction of a facial expression generative model is suggested. The technology allows the synthesizing of different facial expressions on the basis of several support photographs. Each facial expression in the suggested approach is described as a deformation of some basis photograph of the individual. A search for an appropriate deformation is executed by solving the (max, +)-labeling problem.

Запропоновано технологію автоматичної побудови генеративної моделі мимики людського обличчя, що дозволяє синтезувати різні вирази обличчя на основі декількох опорних його фотознімків. Кожен вираз обличчя в запропонованому підході описується як деформація деякого базового його фотознімку. Пошук відповідної деформації виконується шляхом розв'язання (max, +)-задачі розмітки.

Введение. Область компьютерного зрения, связанная с моделированием и распознаванием человеческих лиц, находит широкое применение в различных сферах деятельности [1]. Технологии генерирования реалистических изображений человеческого лица используются в кино- и фотопроизводстве, создании удобных для пользователя *web*-приложений, они – составные части систем безопасного доступа, основанные на биометрической идентификации. Незирая на широкий выбор различных технологий генерирования мимики лица, сложно найти среди них полностью автоматическую систему, которая по нескольким фотографиям человека способна генерировать выражения его лица, ранее не наблюдаемые камерой [1–4]. Вместе с тем существующие системы такого типа, как правило, не основываются на решении формально поставленной оптимизационной задачи, что существенно усложняет оценку качества таких алгоритмов.

В данной статье предложена технология генерирования мимики человеческого лица, основывающаяся на решении точно сформулированной оптимизационной задачи, известной как (max, +)-задача. Эта технология базируется на предположении, что различные выражения человеческого лица можно получить в результате линейной комбинации нескольких опорных. В такой постановке задачи самый сложный момент – поиск соответствий между пикселями опорных изображений. Проблема по-

иска таких соответствий в статье формулируется в виде (max, +)-задачи.

Генеративная модель**Определения**

Поле зрения T будем называть двумерную прямоугольную решетку

$$\{(i, j): 0 \leq i < W, 0 \leq j < H\}.$$

Элементы поля зрения назовем *пикселями*.

Структурой соседства \mathfrak{S} будем называть множество пар соседних пикселей:

$$\mathfrak{S} = \left\{ ((i, j), (i', j')) : (i, j) \in T, (i', j') \in T, ((i' = i + 1) \wedge (j' = j)) \vee ((i' = i) \wedge (j' = j + 1)) \right\}.$$

Разметкой поля зрения T на $l \times l$ сегментов назовем функцию

$$k_T : T \rightarrow K, K = \{0, 1, \dots, l - 1\} \times \{0, 1, \dots, l - 1\}.$$

Элемент разметки обозначим $k(i, j)$, а первую и вторую его компоненты $k_x(i, j)$ и $k_y(i, j)$ соответственно. При этом K будем называть *множеством меток*.

Изображением I будем называть функцию

$$I: T \rightarrow RGB, RGB = \{0, 1, \dots, 255\} \times \{0, 1, \dots, 255\} \times \{0, 1, \dots, 255\}.$$

Поле движения изображения I лица относительно изображения I' того же лица над полем зрения T назовем разметку k_T , которая ставит в соответствие каждому пикселу (i, j) изображения I пиксел $(i + k_x(i, j), j + k_y(i, j))$ изображения I' такой, что пикселы (i, j) и

$(i + k_x(i, j), j + k_y(i, j))$ соответствуют одной и той же точке лица.

Формализация проблемы

Обозначим через I^T множество всех возможных изображений лица какого-либо человека на поле зрения T . Пусть задано некоторое выделенное изображение I_0 лица этого человека: фронтальный снимок с «нейтральным» выражением лица. Назовем это изображение *базовым*. Пусть задано некоторое множество изображений $\tilde{I}^T \subseteq I^T$, содержащее фронтальные снимки этого же человека с различными типичными выражениями лица. Пусть для каждого изображения $I \in \tilde{I}^T$ задано поле движения данного изображения относительно базового. С помощью этой информации можно генерировать выражения лица, представляющие собой различные комбинации выражений лица $I_0 \cap \tilde{I}^T$. Основная сложность состоит в поиске поля движения каждого изображения из \tilde{I}^T относительно базового изображения I_0 . Сформулируем эту проблему в терминах (max, +)-задачи.

Определим отличие между точкой (i, j) на изображении I_0 и точкой $(i + k_x(i, j), j + k_y(i, j))$ на изображении $I \in \tilde{I}^T$:

$$q_{i,j}(k_x(i, j), k_y(i, j)) = \sum_{c=1}^3 |I_1(i, j, c) - I_2(i + k_x(i, j), j + k_y(i, j), c)|, \quad (1)$$

где c – номер *RGB*-канала.

Наложим некоторые естественные ограничения на взаимное расположение соответствующих пикселей на паре изображений, а именно запретим им «перепутываться»:

$$g_{i,j,i+1,j}(k_x(i, j), k_y(i, j), k_x(i+1, j), k_y(i+1, j)) = \begin{cases} 0, & \text{если } k_x(i+1, j) \geq k_x(i, j) - 1, \\ & |k_y(i, j) - k_y(i+1, j)| < \Delta \\ \infty, & \text{иначе} \end{cases} \quad (2)$$

$$g_{i,j,i,j+1}(k_x(i, j), k_y(i, j), k_x(i, j+1), k_y(i, j+1)) =$$

$$= \begin{cases} 0, & \text{если } k_y(i, j+1) \geq k_y(i, j) - 1, \\ & |k_x(i, j) - k_x(i, j+1)| < \Delta \\ \infty, & \text{иначе} \end{cases} \quad (3)$$

Функции (2) и (3) равны нулю для допустимых пар соответствий и бесконечности для недопустимых. На рис. 1 приведена визуализация этих ограничений: если черный пиксел левого изображения соответствует черному пикселу правого, то серый пиксел левого изображения может соответствовать только одному из серых пикселей правого.

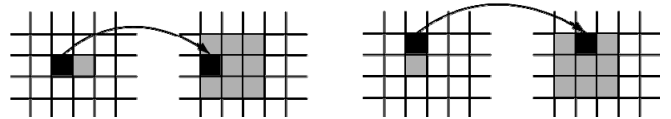


Рис. 1. Ограничения на взаимное расположение соответствующих пикселей на паре изображений

Определим *штраф* за разметку k :

$$Q(k) = \sum_{x=1}^X \sum_{y=1}^Y q_{x,y}(k_x(x, y), k_y(x, y)) + \sum_{x=1}^{X-1} \sum_{y=1}^Y g_{x,y,x+1,y}(k_x(x, y), k_y(x, y), k_x(x+1, y), k_y(x+1, y)) + \sum_{x=1}^X \sum_{y=1}^{Y-1} g_{x,y,x,y+1}(k_x(x, y), k_y(x, y), k_x(x, y+1), k_y(x, y+1)), \quad (4)$$

где W и H – вертикальный и горизонтальный размеры поля зрения.

Задача отыскания оптимального соответствия между двумя изображениями заключается в поиске разметки k с минимальным штрафом:

$$k^* = \operatorname{argmin}_k Q(k). \quad (5)$$

Множество задач (5), определяемых различными значениями параметров $g_{i,j,i+1,j}$ и $g_{i,j,i,j+1}$, составляет *NP*-полный класс задач. Для частного случая, когда функции $g_{i,j,i+1,j}$ и $g_{i,j,i,j+1}$ заданы выражениями (2) и (3) соответственно, также не известен полиномиальный алгоритм точного решения задачи (5). Для приближенного решения этой задачи мы использовали метод, предложенный в [5].

Усовершенствования

Непосредственное применение алгоритма [5] в задачах большой размерности неэффективно. Для решения этой проблемы используется метод, идея которого состоит в том, что

для поиска разметки k_T для изображений I_0 и $I \in \tilde{T}^T$ сначала выполняется поиск разметки k_T для изображений I'_0 и I' , где I'_0 и I' – уменьшенные в два раза изображения I_0 и I соответственно. После этого задается начальное приближение разметки k_T :

$$k(i, j) = 2k(i', j').$$

Затем выполняется поиск разметки k_T относительно этого начального приближения, но уже с меньшим количеством меток. Таким образом, строится «каскад» разметок.

Как показали эксперименты, время работы алгоритма с применением одного каскада уменьшается более чем в восемь раз. При этом точность нахождения поля движения даже улучшилась.

Функция штрафа (4) не учитывает многих физических факторов, поэтому если изображения I_0 и I сильно отличаются, поле движения может находиться неправильно, хотя штраф будет минимизирован. Решение этой проблемы заключается в том, что организационным образом получают промежуточные между I_0 и I изображения и находится последовательность полей движения между парами соседних изображений, на основании которой определяется результирующее поле движения между I_0 и I .

Работу алгоритма можно упростить, если предварительно найти аффинное преобразование одного изображения в другое, и на основе этого преобразования инициализировать начальное приближение поля движения. Например, можно найти координаты глаз на обоих изображениях и вычислить коэффициенты совмещающего их аффинного преобразования.

Синтез мимики на основе генеративной модели

Пусть задано множество опорных фотографий \tilde{T}^T с различными выражениями лица, и базовое изображение I_0 , представляющее «нейтральное» выражение лица. Пусть для каждого изображения $I \in \tilde{T}^T$ задано поле движения k^I между базовым изображением I_0 и изображением I , а также поле движения \tilde{k}_i между

I и I_0 . Пусть для каждого изображения $I \in \tilde{T}^T$ задан весовой коэффициент $\alpha_I \in [0, 1]$. Генерирование изображения I' , представляющего комбинацию выражений лица \tilde{T}^T , осуществляется следующим образом.

Для каждого пиксела (i', j') изображения I' вычислим соответствующий ему пиксел (i_0, j_0) изображения I_0 :

$$i_0 = i' + \sum_{I \in \tilde{T}^T} \alpha_I \cdot \tilde{k}_x^I(i, j), \quad j_0 = j' + \sum_{I \in \tilde{T}^T} \alpha_I \cdot \tilde{k}_y^I(i, j) \quad (6)$$

и соответствующие пикселы каждого из опорных изображений:

$$i_I = i_0 + k_x^I(i, j), \quad j_I = j_0 + k_y^I(i, j), \quad I \in \tilde{T}^T. \quad (7)$$

Вычислим нормированные весовые коэффициенты:

$$\tilde{\alpha}_I = \frac{\alpha_I}{\sum_{I \in \tilde{T}^T} \alpha_I}, \quad \alpha_0 = 1 - \sum_{I \in \tilde{T}^T} \tilde{\alpha}_I. \quad (8)$$

Цвет в каждом пикселе результирующего изображения определяется как линейная комбинация цветов соответствующих пикселов на изображениях из \tilde{T}^T :

$$I'(i', j') = \alpha_0 \cdot I_0(i_0, j_0) + \sum_{I \in \tilde{T}^T} \tilde{\alpha}_I \cdot I(i_I, j_I). \quad (9)$$

Пример. На рис. 2 проиллюстрировано применение предлагаемой технологии для генерирования мимики человеческого лица. В верхней части рисунков приведены опорные изображения, в нижней – другие выражения лица этого же человека, сгенерированные в соответствии с построенной генеративной моделью.



Рис. 2. Пример генерирования мимики человеческого лица на основе опорных изображений

Заключение. Предложенная технология построения генеративной модели мимики человеческого лица позволяет синтезировать различные выражения лица на основе нескольких опорных его фотоснимков. При этом она не требует вмешательства оператора в процесс построения генеративной модели. Каждое выражение лица человека в предложенном подходе описывается как деформация некоторого базового его фотоснимка. Мы предполагаем, что эту деформацию можно представить в виде линейной комбинации деформаций, соответствующих нескольким опорным выражениям лица. Это позволяет свести проблему построения генеративной модели мимики человеческого лица к задаче поиска соответствий между пикселями опорных изображений, которая, в свою очередь, формулируется в виде (max, +)-задачи.

1. *Cootes T.F., Taylor C.J.* Statistical Models of Appearance for Computer Vision, 2001.
2. *Ghent J., McDonald J.* 2005. Photo-realistic facial expression synthesis // *Image Vision Comput.* – 2005. – **23**, N 12. – P. 1041–1050.
3. *Geometry-Driven* Photorealistic Facial Expression Synthesis / Qingshan Zhang, Zicheng Liu, Baining Guo et al. // *IEEE Transactions on Visualization and Computer Graphics.* – Jan./Febr. 2006. – P. 48–60.
4. *Parameterized* facial expression synthesis based on MPEG-4.EURASIP / A. Raouzaoui, N. Tsapatsoulis, K. Karpouzis et al. // *J. Appl. Signal Process.* – Jan. 2002. – N 1. – P. 1021–1038.
5. *Kolmogorov V.* Convergent Tree-Reweighted Message Passing for Energy Minimization // *IEEE Transactions on Pattern Analysis and Machine Intelligence.* – Oct. 2006. – **28**. – P. 1568–1583.

Поступила 23.12.2010
Тел. для справок: (044) 502-6336 (Киев)
E-mail: maxim.tischenko@gmail.com
© М.А. Тищенко, 2011

