

Генерування зображень різних виразів людських облич

Недашківський О.В., Тищенко М.А.

Міжнародний науково-навчальний центр
інформаційних технологій та систем, Київ, Україна

Анотація

В даній роботі запропоновано технологію генерування різних виразів обличчя людини на основі кількох її опорних фотознімків. Ключовим елементом технології є пошук відповідностей між пікселями опорних зображень. В даній роботі проблема пошуку таких відповідностей сформульована у вигляді $(\min, +)$ задачі розмітки. Для розв'язку цієї задачі використано відомий TRW-S алгоритм. В роботі також запропоновано алгоритм оцінювання параметрів отриманої $(\min, +)$ задачі, який належить до класу EM-алгоритмів.

1. Вступ

Галузь комп'ютерного зору, пов'язана з моделюванням та розпізнаванням людських облич, знаходить все більш широке застосування в найрізноманітніших сферах людської діяльності. Технології генерування реалістичних зображень обличчя людини використовуються в кіно- та фотовиробництві, створенні зручних для користувача web-додатків, вони є складовими частинами систем безпечного доступу, заснованих на біометричній ідентифікації. Попри широкий вибір різних технологій генерування міміки людських облич, серед них важко знайти повністю автоматизовану систему, що за кількома фотознімками людини здатна генерувати вирази її обличчя, які раніше не спостерігалися камерою. Разом з тим, існуючі системи такого типу, здебільшого, не базуються на розв'язанні формально поставленої оптимізаційної задачі, що ускладнює оцінку якості таких алгоритмів.

В даній роботі запропоновано технологію генерування зображень різних виразів обличчя людини, що заснована на розв'язанні точно сформульованої оптимізаційної задачі, відомої як $(\min, +)$ задача розмітки. Ця технологія базується на припущенні, що різні вирази людського обличчя можна отримати як лінійну комбінацію кількох опорних. В такій постановці найскладнішим моментом технології є пошук відповідностей між пікселями на



Рис. 1: Приклад генерування різних виразів обличчя людини на основі опорних зображень.



Рис. 2: Приклад генерування різних виразів обличчя людини на основі опорних зображень.

опорних фотознімках. Проблема пошуку таких відповідностей в даній роботі формулюється у вигляді $(\min, +)$ задачі розмітки. Оскільки ця задача є NP-повною, для пошуку наближеного її розв'язку використано TRW-S алгоритм [1, 2]. Для оцінки параметрів цієї задачі в роботі запропоновано алгоритм самонавчання, що належить до класу EM-алгоритмів [3].

На рис. 1 та 2 наведено приклади використання запропонованої технології для генерування різних виразів обличчя людини. У верхній частині рисунків наведено наявні в базі даних зображення, в нижній — інші вирази обличчя цієї ж людини, згенеровані відповідно до побудованої генеративної моделі.

2. Технологія генерування зображень різних виразів людських облич

2.1. Базові означення

Полею зору T називатимемо двовимірну прямокутну ґратку $\{(i, j) : 0 \leq i < Width, 0 \leq j < Height\}$. Елементи поля зору називатимемо пікселами.

Структурою поля зору T називатимемо певну множину \mathfrak{S} підмножин $\tau \subsetneq T$ поля зору T , $\mathfrak{S} \subseteq \{\tau : \tau \subsetneq T\}$.

Розміткою поля зору T на $l \times l$ сегментів називатимемо функцію $k_T : T \rightarrow K; K = \{0, 1, \dots, l - 1\} \times \{0, 1, \dots, l - 1\}$. При цьому K називатимемо множиною позначок. У даній статті елементами множини K є вектори $(k_{t_x}, k_{t_y}) = k_{(t_x, t_y)} = k_t$.

Полею зсувів зображення $x : T \rightarrow X$ деякого об'єкта A відносно зображення $x' : T \rightarrow X$ того самого об'єкта A називатимемо функцію $s : T \rightarrow T$, що ставить у відповідність кожному пікселу t зображення x піксел t' зображення x' такий, що піксели t та t' відповідають одній і тій самій точці об'єкта A .

Нехай також задано зображення людського обличчя $x^L : T \rightarrow X$ та $x^R : T \rightarrow X$, які надалі називатимемо відповідно першим та другим зображеннями. Надалі пару x^L, x^R також позначатимемо x_{LR} . Розмітка $k_T : T \rightarrow K$ задає поле зсувів першого зображення відносно другого.

Під символом x розуміється в залежності від контексту, або зображення, або x -ова координата вектору.

2.2. Генеративна модель

Нехай задано поле зору T та множину кольорів X . Нехай X_H^T – множина всіх зображень обличчя деякої людини H . Нехай, далі, маємо деяке виділене зображення x_H^0 обличчя людини H , яке є фронтальним знімком з «нейтральним» виразом обличчя. Будемо називати таке зображення базовим. Нехай, також, маємо деяку множину зображень $\tilde{X}_H^T \subsetneq X_H^T$, що містить фронтальні знімки цієї ж людини з різними типовими виразами обличчя. Нехай, для кожного зображення x із \tilde{X}_H^T маємо поле зсувів даного зображення відносно базового. За допомогою цієї інформації можна генерувати вирази обличчя людини H , які являють собою різні комбінації виразів обличчя $x_H^0 \cup \tilde{X}_H^T$. Основна складність полягає у пошуку поля зсувів кожного зображення з \tilde{X}_H^T відносно базового зображення x_H^0 .

2.3. Технологія

Час роботи однієї ітерації TRW-S алгоритму, використаного в даній роботі для обчислення поля зсувів, залежить від $|K|$ та $|T|$, як $O(|K|^2 \cdot |T|)$, тому безпосереднє його використання стає недоцільним. Для вирішення цієї проблеми був розроблений метод, ідея якого полягає у тому, що для пошуку поля зсувів k_T для зображень x^L, x^R спершу потрібно знайти поле зсувів $k_{T'}$ для зображень x'^L, x'^R , де x'^L, x'^R зменшені удвоє зображення x^L, x^R . Після цього можна оцінити k_T^0 , як $k_t = 2 \cdot k_{t'}$, і шукати k_T відносно k_T^0 з меншою кількістю позначок. Таким чином будується „каскад“ полів зсувів.

Як показали експерименти, час роботи алгоритму при застосуванні одного каскаду зменшується більш ніж у 8 раз. При цьому результати роботи покращились на суб'єктивний людський погляд.

Описана вище функція штрафу не враховує багато фізичних чинників. Тому, якщо зображення x^L, x^R дуже відрізняються, поле зсувів може знаходитися не вірно, хоча штраф буде мінімізовано. Вирішення цієї проблеми полягає у тому, що організаційним чином отримується x^{LR} – зображення між x^L, x^R . Тоді можна знайти поле зсувів $k_T^1 : x^L \rightarrow x^{LR}, k_T^2 : x^{LR} \rightarrow x^R$, результуюче поле зсувів $k_t = k_{k_t^1+t}^2 + k_t^1$.

Роботу алгоритму можна спростити, якщо попередньо знайти афінне перетворення одного зображення в інше, та на основі цього перетворення ініціалізувати початкове поле зсувів. Наприклад, якщо необхідно знайти поле зсувів між двома зображеннями обличчя людини, то можна знайти координати очей на них і обрахувати коефіцієнти афінного перетворення, яке переводить дві точки на одному зображенні в дві точки на іншому.

3. Пошук поля зсувів як $(\min, +)$ задача розмітки

Нехай пікселу з координатами (x, y) на одному зображенні відповідає піксел з координатами $(x + k_x, y + k_y)$ на іншому (зсув піксела тоді рівний (k_x, k_y)). Таке співставлення штрафується довільною функцією яка показує, наскільки кольори в співставлених пікселях різні. Для пари сусідніх по горизонталі (або по вертикалі) пікселів, якщо зсув одного піксела фіксований, зсув іншого не може бути довільним. Вигляд цих обмежень може бути різним для різних прикладних задач. Обмеження, використані в даній роботі, схематично наводяться на рис. 3.

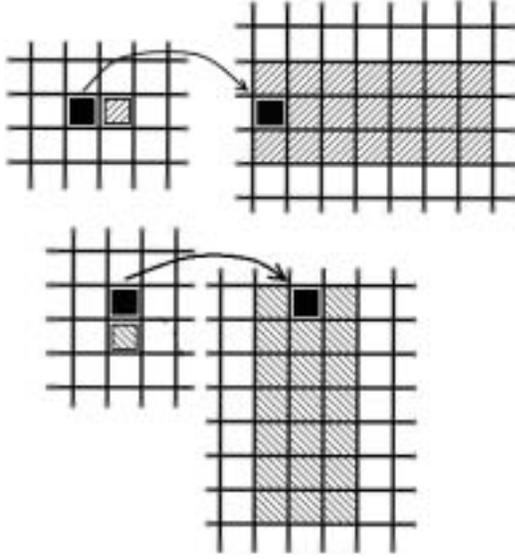


Рис. 3: Зліва - пара двох сусідніх по горизонталі/вертикалі пікселів, зсув темного пікселя зафіксований. Праворуч - заштрихованою областю позначено можливий зсув заштрихованого пікселя.

Нехай, x^L, x^R — два зображення, поле зсувів між якими необхідно знайти; $th = (t_x + 1, t_y)$ — сусідній піксел, що знаходиться праворуч пікселя t ; $tv = (t_x, t_y + 1)$ — сусідній піксел, що знаходиться нижче пікселя t ; $\rho(a, b) = |a - b|$ — відстань між значенням пікселю a та b ; C — коефіцієнт гладкості поля зсувів; $2f_t(k) = C \cdot \rho(x^L(t), x^R(t + k))$ — штраф за зсув k в пікселі t ; $f_1(v, v'), f_2(v, v')$ — штрафи за „паралельний“ та „перпендикулярний“ зсув відповідно, зображені на рис 3.

Тоді штраф за поле зсувів можна записати як

$$\begin{aligned}
 Q(k_T) &= \sum_{t \in T} [2f_t(k_{t_x}, k_{t_y}) + \\
 &\quad + f_1(k_{t_x}, k_{th_x}) + f_2(k_{t_y}, k_{th_y}) + \\
 &\quad + f_1(k_{t_y}, k_{tv_y}) + f_2(k_{t_x}, k_{tv_x})] = \\
 &= \sum_{t \in T} [2f_t(k_t) + f_h(k_t, k_{th}) + f_v(k_t, k_{tv})] \quad (1)
 \end{aligned}$$

Задача полягає в пошуку такої розмітки k_T , що мінімізує функцію $Q(k_T)$. Сформульована оптимізаційна задача відома під назвою (min, +) задачі розмітки. В даній роботі її розв'язок знаходиться за допомогою TRW-S алгоритму, описаного в роботі [1].

4. Алгоритм налаштування параметрів задачі пошуку поля зсувів

Нехай, елементами множини кольорів X будуть вектори (X_1, X_2, X_3) . Вважатимемо, що розмітка та зображення x^L, x^R пов'язані спільним розподілом ймовірностей $p(k, x_{LR})$. Його можна записати у вигляді:

$$p(k, x_{LR}) = p(k) \cdot p(x_{LR}|k).$$

Нехай, тепер, маємо вибірку $\{x_{LR}^i \mid i = 1..N\}$ пар зображень. Вважатимемо, що розподіли ймовірностей $p(k)$ та $p(x_{LR}|k)$ відомі з точністю до деяких параметрів. Оцінимо ці параметри за допомогою методу максимальної правдоподібності, тобто будемо шукати такі їх значення, при яких функція правдоподібності $L(\{x_{LR}^i \mid i = 1..N\})$ досягає максимуму:

$$L(\{x_{LR}^i \mid i = 1..N\}) = \prod_i \sum_k p(k, x_{LR}^i) \rightarrow \max. \quad (2)$$

Деталізуємо вигляд розподілів ймовірностей $p(k)$ та $p(x_{LR}|k)$.

Нехай $p(k)$ має вигляд:

$$p(k) = \frac{\prod_{tt' \in \mathfrak{S}} g_{tt'}(k_t, k_{t'})}{\sum_k \prod_{tt' \in \mathfrak{S}} g_{tt'}(k_t, k_{t'})},$$

де

$$\begin{aligned}
 g_{tt'}(k_t, k_{t'}) &= \\
 &= \exp\left(-\alpha_{tt'} \cdot \|(k_t - \beta_t) - (k_{t'} - \beta_{t'})\|^2\right), \quad (3)
 \end{aligned}$$

де β_t та $\beta_{t'}$ є векторами $(\beta_{t_x}, \beta_{t_y})$ та $(\beta_{t'_x}, \beta_{t'_y})$ відповідно.

Значення параметрів β_t в (3) можна інтерпретувати як „середню“ конфігурацію поля зсувів в пікселі t , а значення параметрів $\alpha_{tt'}$ — як ваговий коефіцієнт, що характеризує „міру важливості“ величини „стрибка“ поля зсувів в парі пікселів (t, t') .

Зауважимо, що у одновимірному випадку (коли k_t та β_t є скалярами) функції $g_{tt'}(k_t, k_{t'})$ стають субмодулярними, що дозволяє використовувати їх в алгоритмах, які базуються на розв'язанні субмодулярних (min,+) задач.

Таким чином, параметрами, що мають оцінюватись в розподілі ймовірностей $p(k)$, є: $\{\alpha_{tt'} \mid tt' \in \mathfrak{S}\}$ та $\{\beta_t \mid t \in T\}$.

Конкретизуємо вигляд $p(x_{LR}|k)$:

$$p(x_{LR}|k) = \prod_{t \in T} q_t(x_{LR_t}, k_t),$$

де

$$\begin{aligned}
 q_t(x_{LR_t}, k_t) &= \\
 &= p_{N(0, \sigma)}(Ax_t^L + \mu - x_{t+k_t}^R) = \\
 &= \frac{1}{(2\pi)^{3/2} \sqrt{|\det \sigma|}} \\
 &\exp \left(- \left(Ax_t^L + \mu - x_{t+k_t}^R \right)^T \right. \\
 &\quad \left. \sigma^{-1} \left(Ax_t^L + \mu - x_{t+k_t}^R \right) \right).
 \end{aligned}$$

Зупинимося детальніше на сенсі параметрів A та μ . Зазвичай баланс кольорів пари зображень не є ідентичним. Для зменшення впливу цього фактору на якість пошуку поля зсувів здебільшого користуються деякими емпіричними методами. Введення лінійного оператора A та зсуву μ дозволяє врахувати відмінності в балансі кольорів вже на етапі оцінки параметрів моделі.

Отже, в розподілі ймовірностей $p(x_{LR}|k)$ необхідно оцінювати значення параметрів A , μ та σ .

4.1. Явний вигляд алгоритму оцінки параметрів задачі пошуку поля зсувів

Наведемо безпосередньо алгоритм налаштування параметрів задачі пошуку двовимірного поля зсувів.

1. Ініціалізувати A , μ , $\left\{ \sum_i p(k_t = k^1, k_{t'} = k^2 | x_{LR}^i) \mid tt' \in \mathfrak{S} \right\}$, $\{p(k_t = k^1, k_{t'} = k^2) \mid tt' \in \mathfrak{S}\}$, $\{\alpha_{tt'} \mid tt' \in \mathfrak{S}\}$ та $\{\beta_t \mid t \in T\}$. Ініціалізувати γ . Задати критерій зупинки градієнтної оптимізації ϵ . Задати критерій завершення роботи алгоритму ϵ_1 .

2. Оцінити параметри A та μ .

3. Оцінити

$$\left\{ \sum_i p(k_t = k^1, k_{t'} = k^2 | x_{LR}^i) \mid tt' \in \mathfrak{S} \right\}$$

за допомогою генератора Гіббса [4] на підставі поточних значень A , μ , $\{\alpha_{tt'} \mid tt' \in \mathfrak{S}\}$ та $\{\beta_t \mid t \in T\}$.

4. Оцінити $\{p(k_t = k^1, k_{t'} = k^2) \mid tt' \in \mathfrak{S}\}$ за допомогою генератора Гіббса на підставі поточних значень $\{\alpha_{tt'} \mid tt' \in \mathfrak{S}\}$ та $\{\beta_t \mid t \in T\}$.

5. Змінити значення $\{\alpha_{tt'} \mid tt' \in \mathfrak{S}\}$ та $\{\beta_t \mid t \in T\}$ наступним чином:

$$\alpha_{tt'} + = \gamma \cdot \frac{\partial l}{\partial \alpha_{tt'}}, \quad \beta_t + = \gamma \cdot \frac{\partial l}{\partial \beta_t}.$$

Зменшити γ .

6. Підрахувати

$$\begin{aligned}
 \delta = \sum_{tt' \in \mathfrak{S}} \left| \sum_i p(k_t = k^1, k_{t'} = k^2 | x_{LR}^i) - \right. \\
 \left. - N \cdot p(k_t = k^1, k_{t'} = k^2) \right|
 \end{aligned}$$

7. Якщо $\delta > \epsilon$ — перейти до пункту 4, якщо $\delta \leq \epsilon_1$ — завершити виконання алгоритму, інакше — перейти до пункту 2.

5. Висновки

Запропонована в роботі технологія дозволяє генерувати різні вирази обличчя людини на основі кількох опорних її фотознімків. При цьому вона не вимагає втручання оператора в процес побудови генеративної моделі. Кожен вираз обличчя людини описується як деформація деякого базового її фотознімка. Ми припускаємо, що цю деформацію можна представити у вигляді лінійної комбінації деформацій, що відповідають кільком опорним виразам обличчя. Це дозволяє поставити задачу побудови такої моделі, як задачу пошуку відповідностей між пікселями опорних зображень, яка, в свою чергу, формулюється у вигляді $(\min, +)$ задачі розмітки. Запропонований в роботі алгоритм оцінки параметрів цієї $(\min, +)$ задачі може бути використано для підвищення точності пошуку міжпіксельних відповідностей і, як наслідок, покращення якості згенерованих зображень.

6. Література

- [1] Kolmogorov V. Convergent Tree-Reweighted Message Passing for Energy Minimization. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.28, pages 1568–1583, Washington, DC, USA, October, 2006.
- [2] Shekhovtsov A., Kovtun I., Hlaváč V. Efficient MRF Deformation Model for Image Matching. Center for Machine Perception, K13133 FEE Czech Technical University, CTU–CMP–2006–08, Prague, Czech Republic, October, 2006. <ftp://cmp.felk.cvut.cz/pub/cmp/articles/shekhovtsov/Shekhovtsov-TR-2006-08.pdf>
- [3] Шлезингер М., Главач В. Десять лекцій по статистическому и структурному распознаванию.— Киев: Наукова думка, 2004.— С. 545.
- [4] Gimel'farb G.L. Modeling image textures with Gibbs random fields. In *Pattern Recognition Letters*, Vol.20, no. 12, pages 1123 - 1132, 1999.