

人脸配准坐标变换解析

Jianwei @ Face Hacker

在很多人脸配准文章中，作者都提到对图像或者 landmarks 进行 similarity transformation 的操作。在我看来，该操作主要为了消除人脸面内旋转对基于级联回归的人脸配准方法造成的干扰。那么所谓的干扰是什么？坐标变换又是如何消除这一干扰的呢？

一般地，基于级联回归模型进行人脸配准的基本形式如下所示：

$$\Delta S^t = W\Phi(S^t) \quad (1)$$

其中 $\Delta S^t = S^{gt} - S^t$ 为第 t 个 stage 形状的残差； W 为线性回归矩阵； $\Phi(\bullet)$ 为特征提取算子。在模型训练过程中，我们需要给定特征和残差，使用最小二乘或其他优化方法得到回归矩阵。无论使用何种策略，为了减小回归误差，一个基本原则是：**相似的特征对应相似的残差**。尽量避免出现相似的输入特征对应不同的残差，或者不同的输入特征对应相似的残差。无论是在模型的训练还是测试阶段，都需要依据这条准则进行。

接下来，我们首先分析人脸的面内旋转如何对(1)式的求解带来干扰。如图(1)所示，假设训练集中存在左右两张图。其中右图由左图进行面内旋转得到；绿色的特征点为当前 stage 的形状，即 S_1^t 和 S_2^t ；红色的点为对应的 S_1^{gt} 和 S_2^{gt} ，两者均基于图像平面坐标系表示。

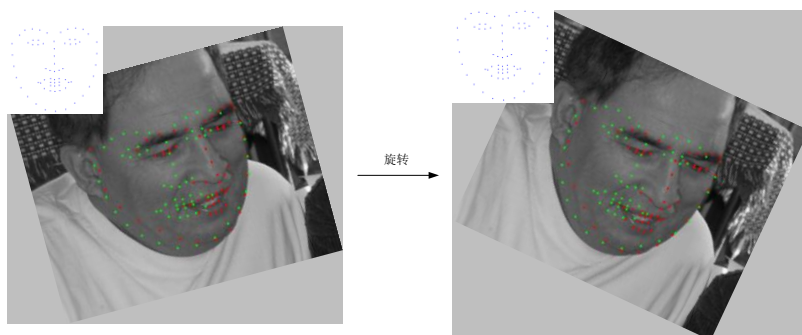


图 (1)

在给定当前 stage 的形状 S_1^t 和 S_2^t 之后，使用 $\Phi(\bullet)$ 提取相应的特征，如 SIFT，HOG 或者 LBP 等等。由于左右两图中 S_1^t 和 S_2^t 相对于人脸的位置是一样的，所以提取的特征在某种程度上也是相似的（相似程度因使用的特征不同而有所差异，如 SIFT 特征会有较高的相似性，而 HOG 和 LBP 相对较弱）。因此， $\Phi(S_1^t) \approx \Phi(S_2^t)$ 。然而， ΔS_1^t 与 ΔS_2^t 之间却有着明显的差异。由于图像的旋转，我们可以得到：

$$\begin{aligned} S_2^t &= cR(S_1^t) + \hat{t} \\ S_2^{gt} &= cR(S_1^{gt}) + \hat{t} \end{aligned} \quad (2)$$

显然，由 (2) 式进一步的得到 ΔS_1^t 与 ΔS_2^t 之间的变换关系：

$$\Delta S_2^t = cR(\Delta S_1^t) \quad (3)$$

即 ΔS_2^t 由 ΔS_1^t 旋转(R)，缩放(c)得到。

此时，我们再定义一个处在图像坐标中的平均形状 S^0 ，如图(1)中左右两图的左上角。

在对 S_1^t ， S_2^t 和 S^0 进行中心化之后表示为 \bar{S}_1^t ， \bar{S}_2^t 和 \bar{S}^0 ，求得如下变换：

$$\begin{aligned} \bar{S}^0 &= c_1 R_1(\bar{S}_1^t) \\ \bar{S}^0 &= c_2 R_2(\bar{S}_2^t) \end{aligned} \quad (4)$$

根据公式(2)，可以得到 $\bar{S}_2^t = cR(\bar{S}_1^t)$ 。根据公式(4)可得 $cR = \frac{c_1}{c_2} R_2^{-1} R_1$ ，代入式(3)可得：

$$\Delta S_2^t = \frac{c_1}{c_2} R_2^{-1} R_1(\Delta S_1^t) \Rightarrow c_2 R_2 \Delta S_2^t = c_1 R_1 \Delta S_1^t \quad (5)$$

根据(5)式可知，估计出 c_1 ， R_1 和 c_2 ， R_2 之后，我们可以分别对残差 ΔS_1^t 和 ΔS_2^t 进行坐标变换使得变换后的残差数值上基本相等，从而满足了如前所述的人脸配准中的基本原则。

为了方便和我的 Matlab 代码进行对照，上述变换主要以下两段代码：

```
129 - Data{i}.tf2meanshape{1} = fitgeotrans(bsxfun(@minus, Data{i}.intermediate_shapes{1}(1:end,:), 1), mean(Data{i}.intermediate_shapes{1}(1:end,:), 1)), ...
130 - (bsxfun(@minus, meanshape_resize(1:end,:), mean(meanshape_resize(1:end,:))), 'NonreflectiveSimilarity');
131 - Data{i}.meanshape2tf{1} = fitgeotrans((bsxfun(@minus, meanshape_resize(1:end,:), mean(meanshape_resize(1:end,:))), ...
132 - bsxfun(@minus, Data{i}.intermediate_shapes{1}(1:end,:), 1), mean(Data{i}.intermediate_shapes{1}(1:end,:), 1)), 'NonreflectiveSimilarity');
```

对应式(4)中的参数估计；

```
138 - [u, v] = transformPointsForward(Data{i}.tf2meanshape{1}, shape_residual(:, 1)', shape_residual(:, 2)');
```

对应(5)式中的坐标变换。

在级联回归训练过程中的每个 stage，都需要对变换参数进行估计，并应用于当前 stage 的残差分量上，对其进行坐标系变换，该操作见 globalregression 文件；而在测试过程中，我们同样需要对坐标进行变换。具体地，使用前一个 stage 所估计的 shape 与平均 shape 求得式(4)中的变换参数，然后对线性回归计算出的残差进行反变换并叠加到前一个 stage 所估计的人脸关键点位置上，得到当前 stage 新的人脸关键点位置，具体参见 globalprediction 中的代码。

以上仅仅介绍了代码中两个变换中的一个，即对残差坐标系的变换。该变换能够基本消除图像面内旋转对级联回归训练过程带来的负面影响，尽可能保证在输入特征一致或相似的情况下，待回归的目标数值（残差）也是一致的。有助于降低模型的训练误差，保证输入输出的一致性，从而也降低了模型的测试误差。除了要确保面内旋转对残差数值不产生影响之外，我们还需要保证面内旋转对特征提取不产生影响。关于该变换的介绍，下回分晓。