

Report for 1+N fusion Project

——2018 年春季学期算分课程项目

邓康乐、陈浩然、金庆哲、汪畅
北京大学信息科学技术学院

June 28, 2018

Abstract

本文提出了同一场景下的多照片融合技术用以实现“1+N fusion”，即将一张含有人像的自拍或近距摄影与 N 张背景照进行融合，得到的结果中既含有清晰的人像，又提供了完整开阔的背景。首先，采用白平衡调节、基于图像匹配的伽马校正等一系列算法对每组数据集进行亮度统一，以消除同一场景由于拍摄导致的色彩特征差异。然后通过 SIFT 特征计算以及根据 RANSAC 算法找出典型的特征匹配点，应用单应性变换使得场景照片与人像照片对齐。在尝试了几种方法后，我们提出了基于 distanceTransform-距离变换的渐变加权融合用于解决关键的照片融合问题。由于拍摄角度的差异，最后得到的结果并非完美的矩形，因此还进行了场景补全。最后进行全局光照调节以解决过曝光和次曝光的问题。几组实验结果表示我们的方法可以取得理想的结果。

1 研究背景

在当今时代，随着社交媒体的发展兴盛，分享照片也愈发流行起来。^[1]提出了“1+N fusion problem”，并且提出了一种解决办法。但在目前的手机照相技术中，存在着很多的技术问题，这些问题影响了人们对于自拍效果和场景重建的满意程度。

- 由于自拍大多使用前置摄像头进行手持拍摄，或者用后置摄像头进行近距离拍摄，人像在画面占比一般很高，难以给与人身临其境之感。
- 人的皮肤和背景的色差很明显，由于人的存在使得提取场景特征变得困难。
- 同一场景由于照射角度的不同，照片的色彩特征会出现偏差，使得强行统一的场景变得不协调。

为解决以上问题，我们提出了本文的来对于一组相同场景的照片进行融合的方法，通过将一组照片合成，我们得到的照片不仅可以包含清晰的自拍，并且提供了完整开阔的背景，我们称之为“同一场景下的多照片融合技术”。

在我们的方法中，一组照片必须基于以下一些假设：

- (a) 照片中包含一张含有人的照片，该照片必须含有完整的人脸
- (b) 其余照片是人像照片背景的延展，但是必须与人像照片有一定的背景重合

在我们的方法中，只要一组照片满足以上两个条件，可以对任意多张场景图像进行合成。在本文中，为了方便展示，我们统一使用 1 张人像照片和 5 张背景照片进行展示。但需要说明的是，只需要稍加改动，我们的方法可以完美的推广到任意多张照片。

我们的方法包含以下五个步骤：

1. 单照片照明统一
2. 多照片对齐
3. 照片融合
4. 场景补全
5. 全局光照调节

2 技术详解

2.1 单照片照明统一

由于不同照片的拍摄角度和光照条件的不同，即使是同一个场景下的照片，照明风格也有较大差异，如图 1 所示，六张照片分别同一场景的六个不同的角度，其中包含一个人像和五个场景图。可以看出，这五张照片的色调、亮度等属性均有较大差异，若强行融合，会导致明显的色差和违和。



Figure 1: 未处理之前的图片

在单照片照明统一中，我们首先进行白平衡调整。白平衡是描述红、绿、蓝三基色混合生成后白色精确度的一项指标。由于光照的不同，照片的白平衡并未达到一致，自动白平衡的目的就是消除不同光源造成的色彩偏差。

在这里，我们选取“白点检测算法”来实现白平衡操作 [2]。为了更加准确的找出图中的白点，我们首先将 RGB 颜色空间转换为 YcbCr 空间，该空间更加符合人类的视觉认知，转换公式如下：

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.2990 & 0.5870 & 0.1140 \\ -0.1687 & -0.3313 & -0.5000 \\ 0.5000 & -0.4187 & -0.0813 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

过滤并得到所有处于 A 区域的像素值，在我们的算法中， x 取所有像素 Y 值的 10% 分位数， a 取所有像素 Cb 绝对值的 10% 分位数。将该区

域像素值转换回 RGB 空间，计算得到每个颜色通道的均值，并由以下公式得到校正值。

$$\begin{cases} R_{ave} = \sum_A \frac{R}{|A|} \\ G_{ave} = \sum_A \frac{G}{|A|} \\ B_{ave} = \sum_A \frac{B}{|A|} \end{cases} \quad (2)$$

$$A_{ave} = (R_{ave} + 2 \cdot G_{ave} + B_{ave})/4 \quad (3)$$

$$\begin{cases} K_R = \frac{R_{ave}}{A_{ave}} \\ K_G = \frac{G_{ave}}{A_{ave}} \\ K_B = \frac{B_{ave}}{A_{ave}} \end{cases} \quad (4)$$

利用得到的三个校正值 K_R, K_G, K_B ，对于每张照片的 RGB 三通道进行调节，调节公式如下：

$$\begin{cases} R' = \frac{R}{K_R} \\ G' = \frac{G}{K_G} \\ B' = \frac{B}{K_B} \end{cases} \quad (5)$$

由此便得到了白平衡处理后的图片，如图 2 所示，左方未处理过的图片色调偏暖偏黄，右方处理过后色调发生明显变化。



Figure 2: 白平衡前后对比

照明统一的第二步，是进行亮度统一化。我们首先定义一个照片的亮度，公式如下：

$$Y = 0.2990 \cdot R_{ave} + 0.5870 \cdot G_{ave} + 0.1440 \cdot B_{ave} \quad (6)$$

其中 X_{ave} 表示 X 通道所有像素的平均值, $X \in \{R, G, B\}$

为了全方面利用照片的结构信息来进行照明统一, 我们创造性的将算法第二个步骤——多照片对齐, 运用到第一个步骤中。

我们将照片集合定义为 $S = \{S_0, S_1, \dots, S_n\}$, 其中 S_0 为人像照片, 其余皆为背景照片。取出 S 中元素, 进行白平衡处理, 而后重新放回集合。对于任意的 $i \in \{1, 2, \dots, n\}$, 我们利用特征点匹配算法 S_i 和 S_0 两张照片的图像中相同或相似的场景进行匹配, 并利用单应性变换算法将 S_i 匹配部分投影到 S_0 的相应区域, 由此所得的图片记为 S'_i (特征点匹配和单应性变换算法详见步骤 2)。下面构造蒙版图片 M_i :

- M_i 为灰度图, 其取值为 0 或 255
- M_{ijk} 表示图片第 j 行第 k 列的元素
- $M_{ijk} = 0 \Leftrightarrow S'_{ijk} = (0, 0, 0)$
- $M_{ijk} = 255 \Leftrightarrow S'_{ijk} \neq (0, 0, 0)$

如图 3 所示, 从左到右依次为 S_0, S_i, S'_i, M_i

定义蒙版照片中 $M_{ijk} = 0$ 的区域为 B_i , 找到 S_0, S_i 中的对应区域, 计算该区域的亮度为 Y_0 和 Y_i , 将他们视为 S_0 和 S_i 的亮度。

定义亮度调节因子 $rate = \frac{\ln \frac{Y_0}{256}}{\ln \frac{Y_i}{256}}$, 利用伽马矫正算法进行亮度调节, 伽马矫正公式如下, 其中 P_{old} 为原图像素值, P_{new} 为处理后的像素值。

$$P_{new} = \left(\frac{P_{old} + 0.5}{256} \right)^{\frac{1}{rate}} \times 256 - 0.5 \quad (7)$$

将 RGB 三通道分别处理完成后, 合成为新图即可。如图 4, 即为经过白平衡和伽马矫正的一组图像。

由此, 我们通过运行白点检测算法以及基于图片匹配的伽马矫正算法, 成功将多张图片的照明条件相统一, 为后面的照片匹配和融合奠定了基础。

2.2 多照片对齐

在调节了各张照片的光照之后, 还有一个重要的问题没有解决, 那就是各张照片的对齐问题。显然, 由于拍摄角度的不同, 背景图片与 S_0 中的公共

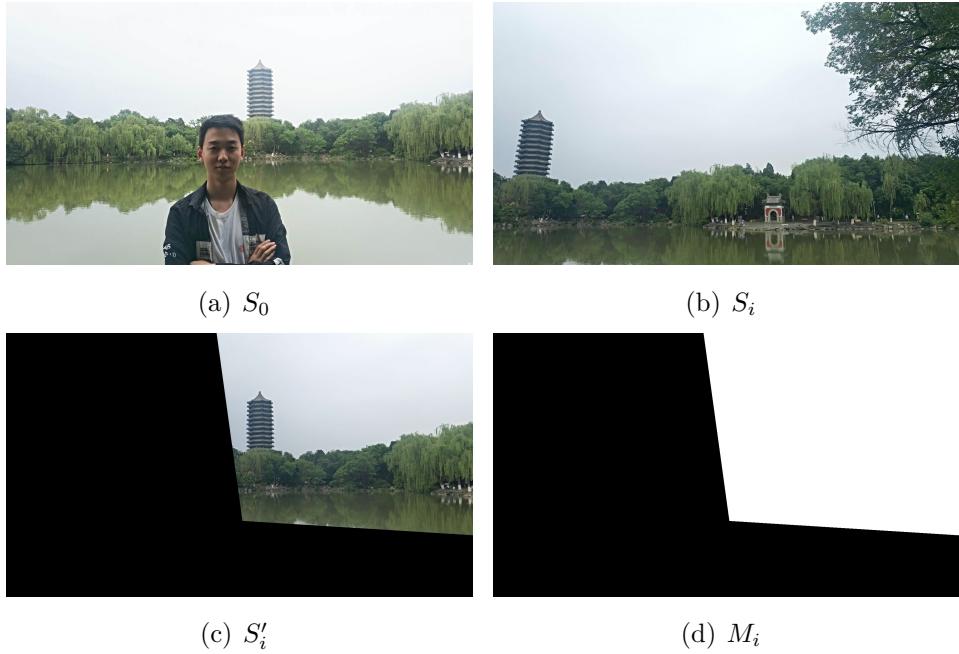


Figure 3: 特征点匹配和单应性变换算法演示



Figure 4: 完成了白平衡和亮度调节的一组照片

部分是无法直接对齐的。因此，我们必须将背景图片透视变换到 S_0 的视平面下。

2.2.1 特征点匹配

首先，想要将一个图片变换到另一个图片的视平面，就要寻找一个能够使得两个图片对齐效果最好的变换规则，这就需要寻找两张图片中那些能够反映图像本质特征的点并进行匹配，也就是所谓的“特征点”。主流的方法有 SIFT, SURF, ORB 等，在尝试了各种方法之后，我们所使用的方法是 SIFT。

SIFT 由 David Lowe 所发表 [3]，分为如下四步：

1. **尺度空间的极值检测：**尺度空间指一个变化尺度 () 的二维高斯函数 $G(x, y, \sigma)$ 与原图像 $I(x, y)$ 卷积（即高斯模糊）后形成的空间，尺度不变特征应该既是空间域上又是尺度域上的局部极值。极值检测的大致原理是根据不同尺度下的高斯模糊化图像差异（Difference of Gaussians, DoG）寻找局部极值，这些找到的极值所对应的点被称为关键点或特征点。
2. **关键点定位：**在不同尺寸空间下可能找出过多的关键点，有些关键点可能相对不易辨识或易受噪声干扰。该步借由关键点附近像素的信息、关键点的尺寸、关键点的主曲率来定位各个关键点，借此消除位于边上或是易受噪声干扰的关键点。
3. **方向定位：**为了使描述符具有旋转不变性，需要利用图像的局部特征为每一个关键点分配一个基准方向。通过计算关键点局部邻域的方向直方图，寻找直方图中最大值的方向作为关键点的主方向。
4. **关键点描述子：**找到关键点的位置、尺寸并赋予关键点方向后，将可确保其移动、缩放、旋转的不变性。此外还需要为关键点建立一个描述子向量，使其在不同光线与视角下皆能保持其不变性。Lowe 在原论文中建议描述子使用在关键点尺度空间内 4×4 的窗口中计算的 8 个方向的梯度信息，共 $4 \times 4 \times 8 = 128$ 维向量表征。

在进行特征点检测之前，我们需要将源图片转换为单通道灰度图以便计算。之后我们计算出源图片 S_1 和目标图片 S_0 中的关键点集合 kp_1, kp_0 以及描述符集合 $\text{des}_1, \text{des}_0$ 。在进行匹配时，我们使用 brute-force 方法，这种方法是在第一幅图像中选取一个关键点，然后依次与第二幅图像的每个关键点进行描述符距离测试（距离为欧式距离）。由于我们的图片规模不是特别大，应当优先选择准确性最强的算法，而显然暴力算法最符合我们的需求。结果如下图（绘制距离最近的 50 个匹配）



Figure 5: 裸的 bf 算法结果

可以看到如果不经过任何处理的话会产生许多误配，而这会对于我们 的变换产生严重的影响。针对这一问题，我们考虑使用 Lowe 提出的比值 测试法。即对于每个关键点 A 寻找两个距离最近的点 B(最近), C(次近)，而只有当 $B/C < d$ ($0 < d < 1$) 时，我们认为 A 与 B 为一个匹配。 d 越接近 0，获取的匹配越理想，但也越少（当 $d=0$ 时匹配就为一一对应），而 d 越大，就能获得更多的匹配，不过也更可能出现误配。下图是调整 d 为 0.6 的结果（绘制距离最近的 50 个匹配）。



Figure 6: 筛选后

下图绘制了全部的匹配



Figure 7: 全部的匹配

能够看出进行了这样的比例测试之后误配基本被消除了。下面我们来看看另外两种方法。在这里不赘述这另外两者的原则。

SURF 的速度在大规模图像的处理上远快于 SIFT，但在我们所要处理的日常照片中并没有本质区别。而 SURF 在计算的过程中运用了一些近似的思想，同时对于边缘的抑制没有 SIFT 做得好，导致了一些误配的出现，图 8 是 SURF 匹配的结果，可以看到房顶及地面处有一些误配。

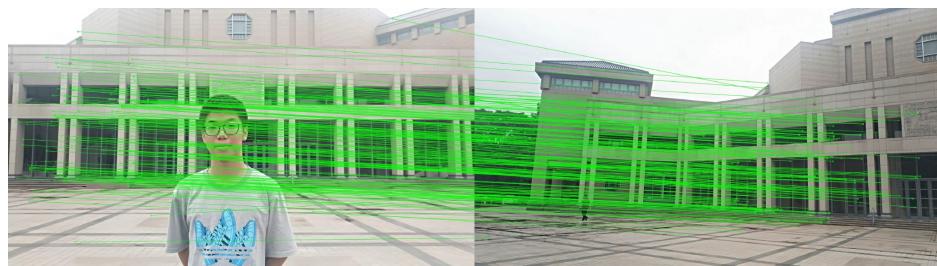


Figure 8: SURF 匹配的结果

ORB 的速度甚至要快于 SURF，它基于 BRIEF 特征描述子，解决了其不具备旋转不变性和对噪声敏感的缺点，但是并不具备尺度不变性，导致了检测出的特征点明显减少。而同样的，在我们日常照片之中它的速度优势并不足以让我们放弃更加精确的方法。图 9 是 ORB 匹配不经过比例测试的结果，可以看到效果是惨不忍睹的。而图 10 是经过了比例测试的结果，虽然消除了误配，但是匹配数量完全无法满足我们的需求。

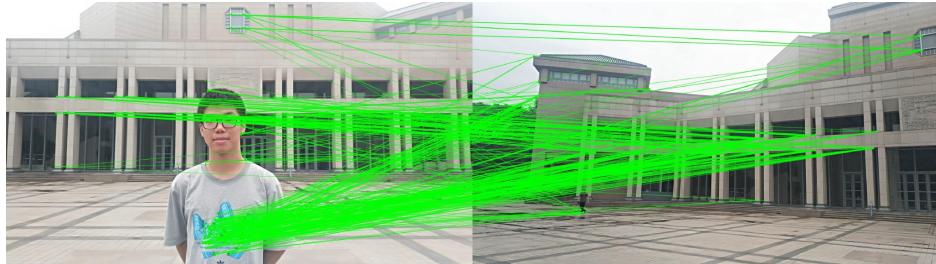


Figure 9: ORB 匹配的结果

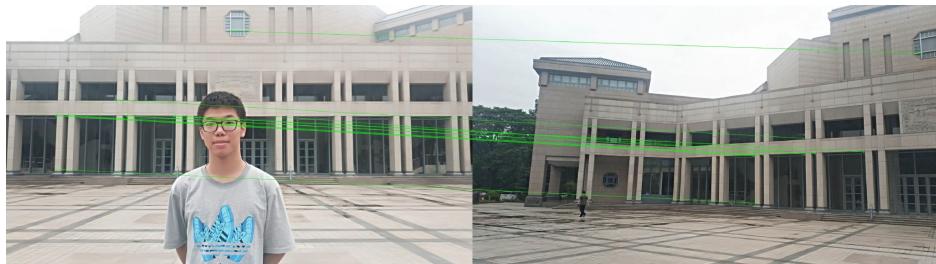


Figure 10: ORB 匹配经过比例测试的结果

综上，我们选择 SIFT 进行匹配。

2.2.2 计算单应性矩阵进行变换

单应性矩阵描述的是针对同一事物，在不同的视角下拍摄的两幅图像之间的关系。单应性矩阵，也就是透视变换矩阵 \mathbf{H} 定义如下

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \quad (8)$$

于是有

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (9)$$

于是我们只需要恢复出 \mathbf{H} 中的 8 个参数，这个过程至少需要四对匹配点。在计算之前，我们考虑使用 RANSAC 方法去除误匹配点对，这个方法使用尽可能少的点来估计模型参数，然后尽可能扩大得到的模型参数的影响范围。

RANSAC 算法的具体描述是：

给定 N 个数据点组成的集合 P ，假设集合中大多数的点都是可以通过一个模型来产生的，且最少通过 n 个点 ($n < N$) 可以拟合出模型的参数，则可以通过以下的迭代方式拟合该参数。对下面的操作执行 k 次：

- (1) 从 P 中随机选择 n 个数据点；
- (2) 用这 n 个数据点拟合出一个模型 M ；
- (3) 对 P 中剩余的数据点，计算每个点与模型 M 的距离，距离超过阈值的则认定为局外点，不超过阈值的认定为局内点，并记录该模型 M 所对应的局内点的值 m ；

迭代 k 次以后，选择 m 最大的模型 M 作为拟合的结果。

于是我们每次从所有的匹配点中选出四对来计算 \mathbf{H} ，选出内点个数最多的作为最终的结果。内点判断方法如下：

$$\left| \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} - \mathbf{H} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \right| \leq t \quad (10)$$

计算出 \mathbf{H} ，我们用透视变换可以将背景图片与 S_0 对齐，为了保证每次对齐的准确性，我们采用边拼接边对齐的方法。设 $fuse(img1, img2)$ 为将两张图片拼接的函数。那我们每次得到 $S'_i = fuse(S_{i-1}, S_i)$ 之后，将 S_{i+1} 与 S'_i 对齐。

2.3 照片融合

为了将各照片融合为一张，首先要确定最终成果照片的大小，这里我们与之前调光、对齐的部分相同，依然以包含人像的照片 S_0 作为基准。设融合完成的图像为 T ，将 T 的尺寸设置为 S_0 长宽的两倍。

然后将进行过光照统一过后的照片集合(依然记为 $S = \{S_0, S_1, \dots, S_n\}$)投影至 T , 即通过之前对齐部分对各照片所得出的单应性矩阵 H_i 计算得到 S_i 在 T 中所处的位置 T_i 。

我们规定人像照片 S_0 在 T 中所处的位置为水平方向居中, 坚直方向处于底部, 即图 11:



Figure 11: S_0 在 T 中所处的位置

除 S_0 以外的其余照片的投影结果如图 12 所示

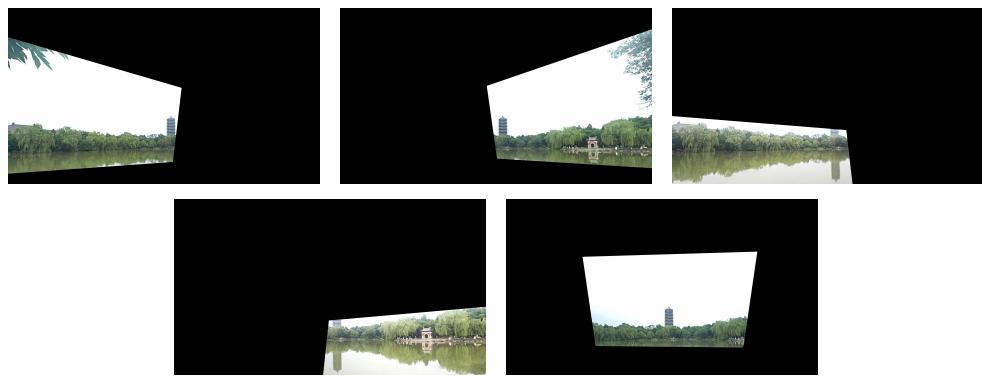


Figure 12: 其余照片的投影结果

确认了各照片在融合图像 T 中的位置后，就可以开始进行照片融合了。我们主要尝试了两种方法：

1. 直接覆盖拼接
2. 基于 distanceTransform- 距离变换的渐变加权融合 (alpha Blending)

2.3.1 直接覆盖拼接

直接覆盖拼接的思想很简单，为了防止人像被覆盖，首先将 T 初始化为 T_0 ， T 中还未被覆盖的区域依次由 T_1, T_2, \dots, T_n 覆盖。效果如图 13 所示：



Figure 13: 暴力拼接

可以看到仅仅只融合了四张，拼接的痕迹就很明显，由光照调节无法完美统一的些微色调差异也很突出。为了改善拼接处的突变，我们对原始的简单算法进行了改进。在几张照片的重叠部分（除包含人像部分），我们并不简单地取某张照片的 RGB 值作为 T 在该点的值，而是取所有照片的 RGB 值最大值作为结果，即：

$$T(i, j) = \max_k T_k(i, j) \quad (11)$$

最后的融合结果如图 14 所示

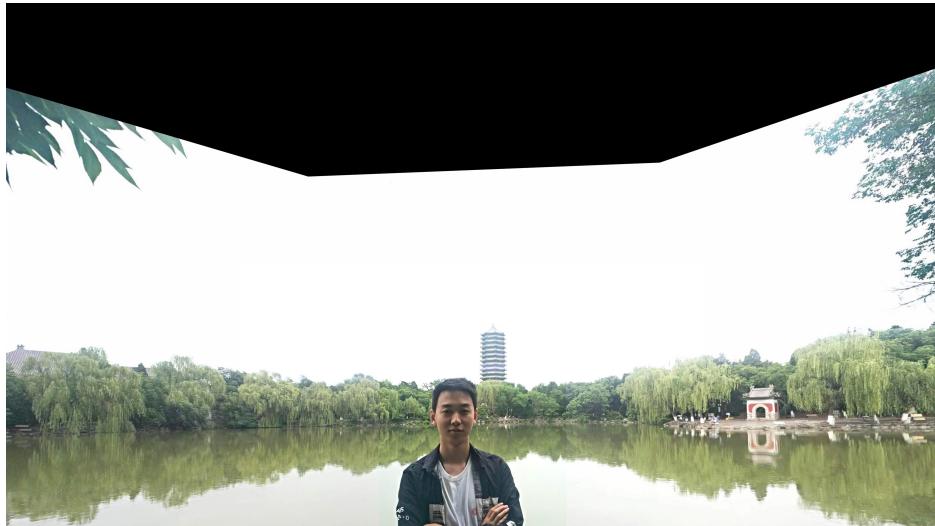


Figure 14: 改进的暴力拼接

相较于之前简单的方法，这次的结果拼接痕迹明显减少，仔细观察仍然能发现包含人像的照片周围有一个具有色差的边界，而且由于取最大值的局限性，使得照片的局部区域泛白；另外，由于各照片之间对齐的微小误差，求最大值使得景物有一些重影。

将该方法运用于另几组数据后，我们发现这个方法并不具有普适性，对于一般的数据依然有比较明显的拼接痕迹，如图 15

2.3.2 基于 distanceTransform-距离变换的渐变加权融合 (alpha Blending)

为了使得照片融合的交界部分过渡更加自然，我们采用了加权融合的技术 (alpha Blending)，即用各图像的 RGB 值的加权平均作为结果图像的 RGB 值。然而，如何确定各个图像的权值是值得商榷的问题，一个自然的想法是越靠近边界的地方权值越小，而远离的方向权值则越来越大。

首先我们将问题简化为两张图片间的融合(原问题可由 T 与 $\{T_1, T_2, \dots, T_n\}$ 依次两两融合实现)，要实现图片间的自然过渡只需要关注两个图像的交集



Figure 15: 失败的拼接

(重叠部分) 即可。假设要对 T 和 T_k 进行图像融合, I 为两者的交集, 矩阵 α 和 β 分别表示图像 T 和 T_k 各对应点的权值。

为了计算 α 和 β , 我们采用了 distanceTransform-距离变换算法 [?]. 距离变换的基本含义是计算一个图像中非零像素点到最近的零像素点的距离, 也就是到零像素点的最短距离。我们对图像 T 和 T_k 在区域 I 分别运用距离变换算法, 就得到了两幅图的 I 中各点到最近零像素点的距离, 也就是 I 中各点到边界的最近距离, 分别记为 $dis_\alpha(x, y)$ 和 $dis_\beta(x, y)$, 其中 $(x, y) \in I$ 。

两图权值的计算表达式应满足以下条件:

1. 归一化: $\alpha(x, y) + \beta(x, y) = 1, \forall (x, y) \in I$
2. $\alpha(x, y)$ 与 $dis_\alpha(x, y)$ 正相关, $\beta(x, y)$ 与 $dis_\beta(x, y)$ 正相关

由此我们给出 α 和 β 的计算公式如下:

$$\alpha(x, y) = \begin{cases} \frac{dis_\alpha^n(x, y)}{dis_\alpha^n(x, y) + dis_\beta^n(x, y)} & (x, y) \in I \\ 1 & (x, y) \notin I \end{cases} \quad (12)$$

$$\beta(x, y) = \begin{cases} \frac{dis_\beta^n(x, y)}{dis_\alpha^n(x, y) + dis_\beta^n(x, y)} & (x, y) \in I \\ 1 & (x, y) \notin I \end{cases} \quad (13)$$

若上式中 $n = 1$, 则表示权值与到边界的最近距离呈线性, 根据多次实验比较, 我们经验性地取 $n = 2$ 。

另一方面, 我们还需要保证人像不与景物重叠, 即把 T 中有人像的区域在 T_k 中设为零。在这里我们采用了 Haar 分类器进行人脸识别, 并按照比例确定整个人体的范围。效果如图 16 所示:

根据上面的讨论, 我们可以得到将 T 和 T_k 融合后的图像 \tilde{T} 的像素值计算公式:

$$\tilde{T}(x, y) = \alpha(x, y) \cdot T(x, y) + \beta(x, y) \cdot T_k(x, y), \forall (x, y) \quad (14)$$

另外, 如对齐部分所提到的那样, 为了取得更加精确的特征点匹配使得对齐带来的误差尽可能小, 我们采用了对齐和融合交替进行的方法, 每

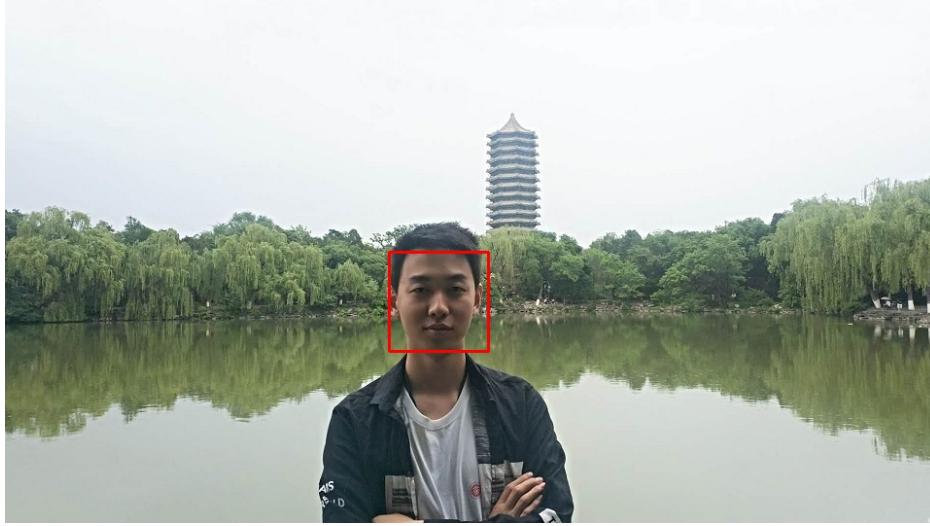


Figure 16: 人脸识别的结果

得到 T 和 T_k 新的融合图像 \tilde{T} 后，再将 S_{k+1} 以 \tilde{T} 为基准对齐得到 T_{k+1} 。最后的融合结果如图 17 所示：

可以看到这个方法不仅基本消除了图像拼接的痕迹，也能够根据参数 n 的大小一定程度上抑制景物的重影，同时还具有普适性。

2.4 场景补全

融合拼接后的照片存在两个问题：1. 照片轮廓为不规则的多边形而非长方形；2. 部分照片角落有失真情况。如图 18 所示：

在场景补全中，用透视的方法将照片轮廓变成长方形，同时也拉伸两个角落。对角落的处理让照片看起来更加真实。

透视补全的步骤依次为：1. 选取要变换的区域，即找不动点；2. 利用不动点对图片做透视变换；3. 将生成的图像与原图像融合。其中第一步最为关键。

以图 18 为例，确定变换最高点的方法是直接比较左上和右上的两个顶点坐标。在图 19 中以 A, B 标出。以较高的点的高度坐标为变换后照片的上边界坐标，则得到两个变换后的不动点 A^* , B^* 。图 18 中 B^* 即 B。

然后用二阈值图的边界判定得到 C, D。在最高边界上作出对应的 C^* ,



Figure 17: 融合结果



Figure 18: 未补全的照片

D^* 。最低变换边界的选择用到景物判定。在图 19 中的最低边界线的理想情况是一条不经过塔，也不经过树的水平直线。进而得到最后的不动点 E , E^* , F , F^* , G , G^* , H , H^* 。

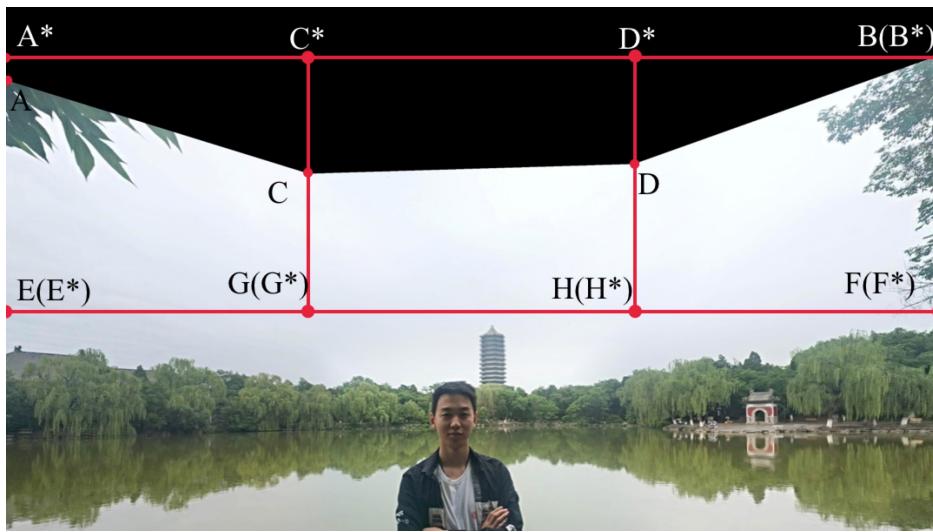


Figure 19: 选取不动点图示

在景区判定中选取恰当的阈值得到二值图后进行边界识别，然后选取一条不经过景物的下界，即选取的下界与识别出的边界不相交。若不存在这种下界，如有的照片每个水平线上都存在景物，则选取照片高度的中点坐标为下界。对图 18 的下界识别如图 20 所示。

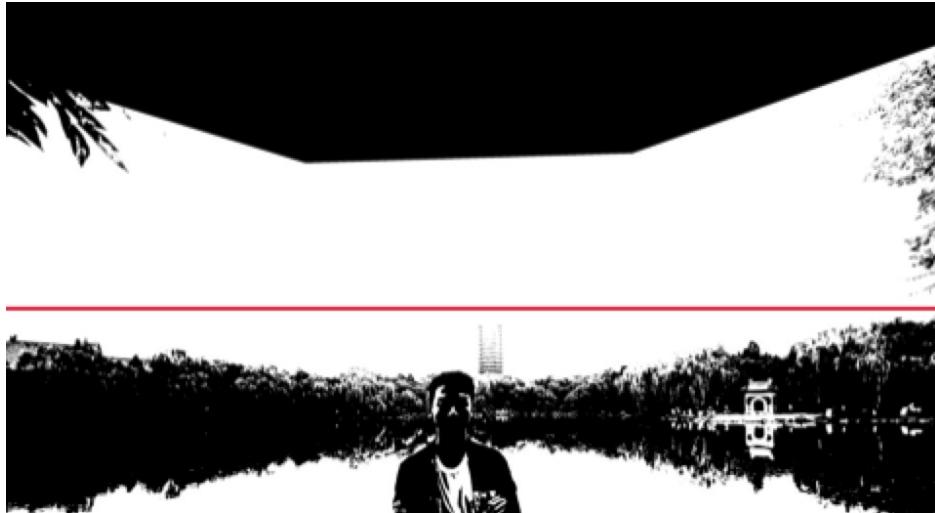


Figure 20: 变换下界的判定图示

得到所有不动点后，用透视变换的方法将 A, E, C, G 作为第一组变换前不动点， A^* , E^* , C^* , G^* 作为第一组变换后不动点处理。C, G, D, H 为第二组变换前不动点， C^* , G^* , D^* , H^* 为第二组变换后不动点。D, G, B, F 为第三组变换前不动点， D^* , G^* , B^* , F^* 为第三组变换后不动点。最后将变换后得到的图与未补全的图片融合即可。



Figure 21: 补全结果

2.5 全局光照调节

在经过以上的步骤之后，我们成功得到了拼接和补全出来的图片。然而，仅仅步骤一中的照明统一是不够的，原因如下：

1. 我们每次只考虑了 S_0 和 S_i 两个图片的照明条件，将其中相匹配的部分来代表总体进行统一化，但是不同照片在 S_0 中的匹配部分是不相同的，这些匹配部分的照明条件的差异我们并未考虑在内。
2. 即使我们完美的统一了他们的照明条件，然而自然界的光照本就不是完美的，很可能存在过量和过暗的地方，反映在照片上就是过曝光和欠曝光。

因此，进行全局光照调节是很有必要的操作，在上诉图片实例中，可以看出人脸部分明显发黑，曝光不够，而天空很亮，曝光充足，我们的算法就是要解决这些问题。首先，我们将照片由 RGB 空间转化为 HSV 空间，该空间可以比较准确的估算出照片的光照分量，转化关系如下：

令 \max 为 R, G, B 三个分量的最大值； \min 为三个分量的最小值，定义域为 $[0, 255]$

若 $\max = \min$, 则 $H = 0, S = 0, V = \frac{\max}{255}$

若 $\max \neq \min$:

当 $G \geq B$ 时, $H = \frac{\max - R' + G' - \min + B' - \min}{\max - \min} \times 60, S = 1 - \frac{\min}{\max}, V = \frac{\max}{255}$

当 $G < B$ 时, $H = 360 - \frac{\max - R' + G' - \min + B' - \min}{\max - \min} \times 60, S = 1 - \frac{\min}{\max}, V = \frac{\max}{255}$

我们定义 $F(x, y)$ 为照片在 HSV 空间中位于 x 行 y 列的像素的 V 值大小。紧接着, 我们利用多尺度高斯函数的算法来提取出该图像的光照分量 [6], 公式如下:

$$G(x, y) = \lambda e^{-\frac{x^2+y^2}{c}} \quad (15)$$

对于特定的 c , 我们求得相应的 λ 使得

$$\iint G(x, y) dx dy = 1 \quad (16)$$

我们将 $F(x, y)$ 与 $G(x, y)$ 进行卷积操作, 得到:

$$H(x, y) = F(x, y)G(x, y) \quad (17)$$

我们知道, c 取不同的值, 我们会得到不同范围大小的卷积核, 为了同时保持较好的全局特性和局部特性, 我们取不同的 c 进行采样并加权:

$$H(x, y) = \sum_{i=1}^N F(x, y)G(x, y) \quad (18)$$

在本文中, 我们取 $N = 3, c1 = 15, c2 = 80, c3 = 250$, 可以取得较好的效果。

因此我们的得到了二维高斯函数卷积后的图像, 用矩阵 H 来表示。由此我们可以构造光照的校正函数。其表达式如下:

$$O(x, y) = 255 \left(\frac{F(x, y)}{255} \right)^\alpha, \alpha = \left(\frac{1}{2} \right)^{\frac{m-H(x, y)}{m}}, m = \text{ave}(F) \quad (19)$$

由该公式, 我们可以对 F 进行校正, 得到新的光照分量 O , 我们由 O 和颜色空间转换时得到的 H 、 S 分量矩阵, 重新合成图像, 记为 F' , 该图像即为最终的答案。

如图??所示, 该算法可以显著的提高画面质量, 降低画面高亮区域亮度, 增加欠曝光区域的暗部细节, 使得整体画面协调性增强, 光照条件趋于合理, 便于进行场景重建和还原。



Figure 22: 处理结果

3 实验结果

我们将我们的方法运用于其他几组数据，并选取一些结果图如下：



Figure 23: 实验结果 1



Figure 24: 实验结果 2



Figure 25: 实验结果 3

4 总结

我们提出了一种解决”1+N fusion”问题的一系列步骤，我们称之为“同一场景下的多照片融合技术”。本文以“1+5”为例展示了这种方法的具体步骤。在此重申如下：

1. 单照片照明统一（白点检测算法以及基于图片匹配的伽马矫正算法）
2. 多照片对齐（基于 RANSAC 的 SIFT 特征点匹配以及单应性变换）
3. 照片融合（基于 distanceTransform-距离变换的渐变加权融合 (alpha Blending)）
4. 场景补全（透视变换）
5. 全局光照调节（基于 HSV 空间的全局光照矫正）

成功的几组数据表明，我们的方法具有一定的普适性，并且可以取得理想中的效果——不仅包含清晰的自拍，并且提供了完整开阔的背景。

References

- [1] Xia S, Yang S, Liu J, et al. Novel self-portrait enhancement via multi-photo fusing[C]//Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2016 Asia-Pacific. IEEE, 2016: 1-4.
- [2] 基于灰度世界和白点检测的自动白平衡算法, 金黃斌; 秦兴; 郭斌林, 电子器件, 2011-04-20
- [3] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International journal of computer vision, 2004, 60(2): 91-110.
- [4] Rosenfeld A, PfaltzJ.L, Sequential operations in digital picture processing. Journal of ACM,1966, 13(4):471-494.
- [5] Frank Y. Shih, YiTa Wu, Fast Euclidean distance transformation in two scans using a 3*3 neighborhood. Journal of Computer Vision and Image Understanding 2004, 195–205.
- [6] 基于二维伽马函数的光照不均匀图像自适应校正算法, 刘志成; 王殿伟; 刘颖; 刘学杰, 北京理工大学学报, 2016-02-15