概述

图像拼接技术是将一个包含多个图像且图像间相互具有重叠区域的图像序列拼接成一幅的无缝高清的全景图像[10]。图像拼接技术的一般步骤为[11]：

1. 图像预处理，包含有基本数字图像处理技术如滤波去噪、灰度图转换等，以及建立图像匹配模板、提取图像特征点等操作。
2. 图像配准，使用特定的配准方法，将拼接图像中的模板或特征点与对应图像中的目标像匹配。
3. 建立变换模型，根据图像间的特征点的匹配关系，采用相关技术方法，得到反映待拼接图像的几何变换的变换模型。
4. 统一坐标变换，利用前一步建立的反映相邻图像的几何关系的变换模型，将待拼接图像转换到相邻图像所在的平面中，即转换到同一坐标系下，为了避免待拼接图像经过处理后的坐标出现负数或非整数的状况，因此需要对拼接图像进行平移或插值。
5. 图像融合，为了消除拼接错位、亮度和光度差异过大等问题，促成拼接图像重叠区域的平滑过渡。

单应性矩阵

单应性矩阵通常描述位于同一几何空间位置上的点对应于两张图像位置之间的几何变换关系，单应性矩阵H把图像中表示相同位置的点一一对应起来[16]。单应性矩阵的定义与图像变换的旋转、平移及平面参数有关，它是一个 3x3 的矩阵，假设两张图像中的点的齐次坐标分别为(u1, v1,1)和(u2, v2,1)，则有

$$\begin{pmatrix} u2 \\ v2 \\ 1 \end{pmatrix} \simeq \begin{bmatrix} h1 & h2 & h3 \\ h4 & h5 & h6 \\ h7 & h8 & h9 \end{bmatrix} \begin{pmatrix} u1 \\ v1 \\ 1 \end{pmatrix} \qquad (2-1)$$

H为 8 自由度，令$h9$为 1 上式可以整理成两个等式，一组匹配点对可以构造出两项约束，因此至少 4 组匹配点对就可以计算出单应性矩阵H。但在实际情况中，由于噪声等因素影响会出现特征点误匹配的情况，因此一般会使用多对点对求解单应性矩阵。

特点:

1. 原点不再是原点；
2. 平行线不再是平行线；
3. 长度比例关系丢失；
4. 两个H相乘依然是H；

相机纯旋转，原点不变，可以准确计算单应性矩阵；当场景与相机很远时，场景就类似于平面，并且相机的位移也可以相对地认为只有旋转。(室外容易拼接，室内困难)

网格点提取

基于直接法的图像配准利用网格将图像中的网格点提取为特征点通过光流法计算其匹配点进行配准。文献[13]提到特征点选取有两步，首先将待拼接图像划分为一个网格，如图 2.1 所示，网格的尺寸根据实际情况进行调整，文献[13]给出图像分辨率为 1280x720pixels 的网格尺寸经验值为 70pixels，网格点是潜在的特征点。

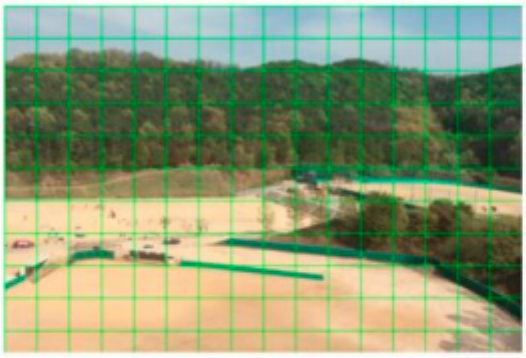


图 3.1 图像网格划分[13]

接着利用网格点在待拼接图片I_t、I_{t-1}之间的灰度值差异提取特征点：

$$|I_t(u,v) - I_{t-1}(u,v)| > \tau_i \qquad (3-1)$$

上式中的u、v是网格点的坐标，$I_t(u,v)$、$I_{t-1}(u,v)$是网格点在待拼接图像中的灰度值，τ_i是灰度值之差阈值。文献[12]指出基于网格的特征提取方法不仅快速而且稳定。

Lucas-Kande光流法

Lucas-Kande 光流法也称作 LK 光流法。LK 光流法存在三个假设前提分别是：

1. 亮度恒定，相邻视频帧间即待拼接图像之间的像素亮度是恒定的。
2. 时间连续，相邻视频帧间的目标物体位移非常缓慢。
3. 空间一致性，在图像中的一个邻域内所有的相邻像素具有相同的光流矢量。

考虑假设 1 和 2 可以得到基本约束方程，设视频帧中一个像素点光强度为$I(x,y,t)$，(x,y)为像素点坐标值，t代表时间维度。到下一视频帧位移矢量为(dx,dy)，时间变换量为dt，根据亮度恒定假设得到：

$$I(x,y,t) = I(x+dx, y+dy, t+dt) \tag{3-2}$$

将上式右端进行泰勒展开，并将泰勒展开式代入上式进行化简变换得到：

$$\frac{\partial I}{\partial x}\frac{dx}{dt} + \frac{\partial I}{\partial y}\frac{dy}{dt} + \frac{\partial I}{\partial t}\frac{dt}{dt} = 0 \tag{3-3}$$

令$u = \frac{dx}{dt}$、$v = \frac{dy}{dt}$表示光流沿 X 轴和 Y 轴的速度矢量，像素点的灰度值沿 X、Y、T 方向的偏导数分别以$I_x = \frac{\partial I}{\partial x}$、$I_y = \frac{\partial I}{\partial y}$、$I_t = \frac{\partial I}{\partial t}$来表示，综上，式子写为:

$$I_x u + I_y v + I_t = 0 \tag{3-4}$$

其中I_x、I_y、I_t可通过图像求得，u和v即为所求光流矢量。根据 LK 光流法的假设前提，可以得到 LK 光流约束方程

$$\sum_{(x,y\in\Omega)} W^2(x)\left(\mathrm{I}_x u+\mathrm{I}_y v+\mathrm{I}_t\right)^2=0 \qquad (3-5)$$

上式中$W^2(x)$是一个窗口权重函数，该函数使得邻域Ω中心的权重大于周围值，在领域上使用最小二乘法等计算得到(u, v)。

LK 算法的假设前提是针对小运动的假设，当目标运动较快时计算就会出现较大的偏差，因此为了解决次问题，提出了金字塔 LK 光流法。金字塔 LK 光流法是通过缩小图片尺寸相对缩小目标物体的位移距离，使其满足假设[19]。

图像金字塔本质上是不同尺寸的图像组合，其最底层为原始图像，对原始图像进行$\frac{1}{2^n}$次降采样得到第 n 层图像，根据目标物体位移距离是否满足光流假设前提确定降采样的层数，一般金字塔层数为 3~5 层。

通过金字塔 LK 光流法可以计算得到特征点的跟踪点，为了消除跟踪误差，文献[13]提到了一种舍弃误跟踪点方法

$$\frac{\mathrm{d}(\mathrm{p},p_{track})}{\mathrm{N}(\omega)}<\tau \qquad (3-6)$$

式子中的$\mathrm{d}(\mathrm{p},p_{track})$为网格点p和其跟踪点$p_{track}$的 L1 距离，N(ω)是搜索窗口内的像素点数，τ是决定跟踪点是否正确的阈值，根据上式可以删去错误的跟踪点。

最近邻比次近邻特征点匹配

根据待拼接图片P和C中的特征点P_a和C_b的描述子向量相似程度，确定特征点是否匹配。设P_a、C_b的描述子向量的第 i 个分量分别为$Descr_{ai}$、$Descr_{bi}$，则P_a、C_b之间的距离为：

$$\mathrm{D}(P_a,C_b)=\sqrt{\sum_{i=1}^{n}(Descr_{ai}-Descr_{bi})^2} \qquad (3-11)$$

如上式所示，特征点之间距离越近则表示特征点之间的描述子向量越相似，特征点匹配的可能性越高。

Lowe 在文献[24]中提到了一张图片中的特征点可能会没有任何匹配，因为这些特征点可能来自背景噪声或者在拼接图像中没有检测到，因此需要根据某种方法舍弃这些特征点。文献[14]提出了最近邻比次近邻的方法，最近邻距离比次近邻距离的数值Ratio存在一个阈值threshold，当Ratio ≤ threshold时，认为特征点与最近邻的特征点相匹配。

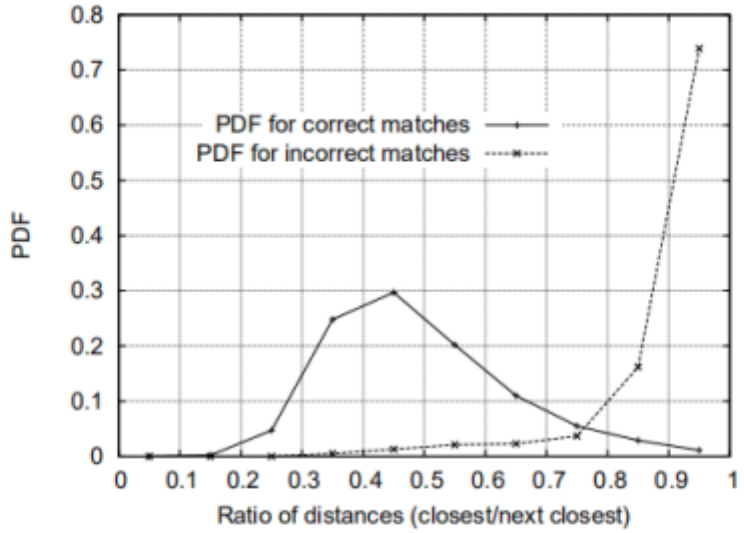


图 3.7 匹配概率与 Ratio 的关系[24]

图 3.7 是 Lowe 使用 40000 个特征点的数据库产生的，如图 2.4 所示当 threshold ∈ [0.4,0.6]这个范围时，匹配成功的概率比较理想。这个方法本质上体现了匹配成功的特征点，其最近邻距离显著短于次近邻距离。

RANSAC(随机采样一致算法)

RANSAC(Random Sample Consensus)，即随机采样一致算法。RANSAC 算法与单应性矩阵相结合消除误配点对。

设待拼接图像P和C，其特征点为P_i和对应的匹配点为C_i。RANSAC 算法对单应性矩阵的自动稳健估计如下[25]

1. 随机选择三对初始匹配成功的特征点对计算单应性矩阵，用单应性矩阵对P中各个点进行变换得到C'_i，求取各点与匹配点的误差$\delta_i = ||C'_i - C_i||$，当$\delta_i$小于某一阈值时，$\delta_i$对应的$P_i$记为内点。
2. 经过N次随机采样得到的取得内点数量最大的集合。
3. 重复步骤1、2，直到两次计算出的内点数量较为相似，则确定为内点集合。

经过上面的步骤，能够剔除明显存有误差的匹配点对，得到成功匹配点对的内点集。需注意 RANSAC 的阈值不能设置的太严苛，否则会删去图像上满足其他单应性矩阵的正确匹配的点对

图像质量评估方法

PSNR

PSNR 评价方法主要基于两个指标，一个是 MSE(Mean Squared Error)，均方误差；另一个是 PSNR(Peak Signal to Noise Ratio)，峰值信噪比，即峰值信号的能量与噪声的平均能量的比值，通常比值取对数。PSNR 评价方法是一种根据像素误差进行评估的图像质量评价方法，首先需要计算 MSE，假设尺寸为MxN的标准图像即 groundtruth 图为I(u,v)，待评价图像为I′(u,v)，则 MSE 为

$$\mathrm{MSE} = \frac{1}{MN}\sum_{u=0}^{M-1}\sum_{v=0}^{N-1}\left(\mathrm{I}(u,v) - \mathrm{I}'(u,v)\right)^2 \qquad (2-5)$$

MSE 对 groundtruth 图像与待评价图像之差进行能量计算并取均值，其值大小反映了两者之间的误差即待评价图像中的噪声，根据峰值信号能量与 MSE 之比可以求得 PSNR

$$\mathrm{PSNR} = 10 \times log_{10}\left(\frac{(2^n-1)^2}{MSE}\right) \qquad (2-6)$$

上式中n为图像位深，PSNR 值越大则说明待评价图像越接近于 groundtruth，图像质量越好。

由于 PSNR 评价方法计算复杂度小，已经广泛应用于图像评估工作中。但是这种方法仅考虑了两个图像中每个对应像素的灰度值之差反映的情况，缺少对图像像素灰度值大小的分布等信息的分析，此方法没有反映出人眼视觉特性。人眼感官对两图像中像素的灰度值误差的敏感度并不准确，很多因素都会影响人眼对其的感知，因此 PSNR 评价方法有时无法对图像做出正确的评价，甚至存在评估结果与人眼主观评价结果完全不同的状况。另外，PSNR 评价方法的分值越高说明图像质量越好，其分值并没有明确的上限，因此这种方法无法准确地客观量化出图像质量的评估结果，无法与人眼主观评估结果建立起清晰的对应关系。

SSIM

SSIM 是一种基于结构相似度的评价方法，SSIM 方法根据图像间的结构相似度判定图像的相似程度，代替之前较常用的误差可见度。算法针对图像的亮度、

对比度和结构相似度的特性，对图像进行评价。

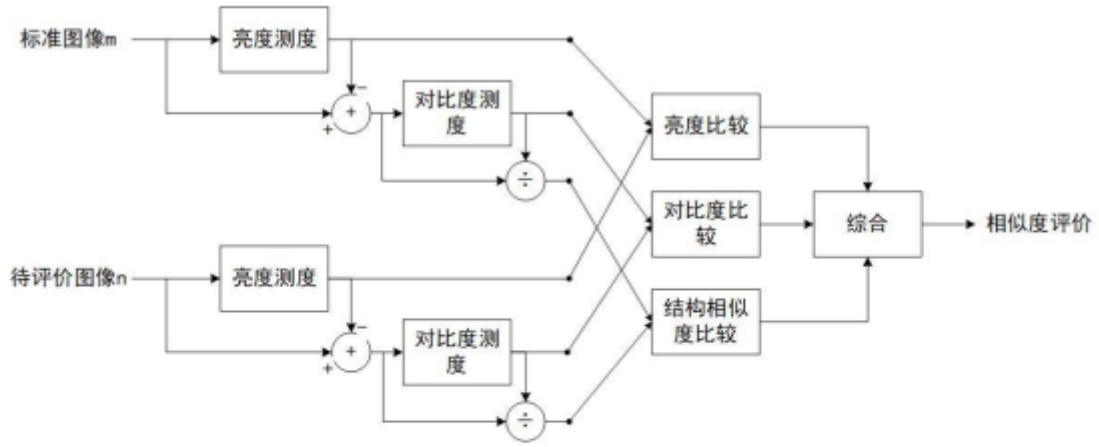


图 2.5 SSIM 评价流程示意图[18]

SSIM 的评价方法如下式所示：

$$S(m,n) = f\big(l(m,n), c(m,n), s(m,n)\big) \tag{2-7}$$

上式中的各项分别为 groundtruth 图像m和待评价图像n之间的亮度相似度$l(m,n)$、对比度相似度$c(m,n)$和结构性相似度$s(m,n)$，$S(m,n)$即为 SSIM 评分，f()表示某种函数关系。分别将μ_m，μ_n设置为两张图像的平均强度，σ_m，σ_n设置为标准差，σ_{mn}设置为两图像之间的互相关系数。

$$\begin{cases} \mu_m = \dfrac{1}{N}\sum_{i=1}^{N} m_i \\ \sigma_m = \left(\dfrac{1}{N-1}\sum_{i=1}^{N}(m_i-\mu_m)^2\right)^{\frac{1}{2}} \\ \sigma_{mn} = \dfrac{1}{N-1}\sum_{i=1}^{N}(m_i-\mu_m)(n_i-\mu_n) \end{cases} \tag{2-8}$$

上式中m_i、n_i分别表示了原始图像和待评价图像中第 i 个像素的灰度值。根据相关论文的定义[21]，SSIM 评价函数的三个相似度计算方法如下式所示：

$$\begin{cases} l(m,n) = \dfrac{2\mu_m\mu_n + C_1}{{\mu_m}^2 + {\mu_n}^2 + C_1} \\ c(m,n) = \dfrac{2\sigma_m\sigma_n + C_2}{{\sigma_m}^2 + {\sigma_n}^2 + C_2} \\ s(m,n) = \dfrac{\sigma_{mn} + C_3}{\sigma_m\sigma_n + C_3} \end{cases} \quad (2-9)$$

在上式中为了防止分母过小引起函数出现错位而分别引入常量C_1、C_2和C_3。依据上式，SSIM 评价方法为：

$$\mathrm{SSIM(m,n)} = [l(m,n)]^{\alpha}[c(m,n)]^{\beta}[s(m,n)]^{\gamma} \quad (2-10)$$

α、β、γ分别为三个分量对评分的权重值，相关论文依据经验设置权重值为$\alpha = \beta = \gamma = 1$，并令$C_3$=$C_2/2$，综上所述，SSIM 评价方法为:

$$\mathrm{SSIM(m,n)} = \frac{(2\mu_m\mu_n + C_1)(2\sigma_{mn} + C_2)}{({\mu_m}^2 + {\mu_n}^2 + C_1)({\sigma_m}^2 + {\sigma_n}^2 + C_2)} \quad (2-11)$$

根据上式便可以根据 groundtruth 图像与待评价图像的数据，实现图像质量评估。

SSIM 在高层次模拟了 HVS(human visual system) 的视觉特征，能够从图像中提取结构信息并对图像质量做了近似处理，算法根据人眼对图像的感知原理使质量评估结果在一定程度上较符合人眼主观评价，并且由于 SSIM 方法的计算相对简洁，从而在国内外得到广泛的关注和应用。

SSIM 评价方法虽然对 groundtruth 和待评价图像之间的亮度对比度的相似性进行了分析，但是经过相关实验的设置并根据实验结果，SSIM 评价方法的评价结果相对于亮度差异不够敏感，并且当出现图像失真、图像质量较差的情况下，SSIM 评价方法与主观评价相差较大。SSIM 评价方法为简单的线性建模处理，因此与人眼主观评价还存有误差，SSIM 评价方法主要考虑了图像间的结构信息的相似性并进行计算量化，但是没有对于图像的底层信息的误差进行深入研究。

国际无线电咨询委员会制订的CCIR500-1主观评价标准

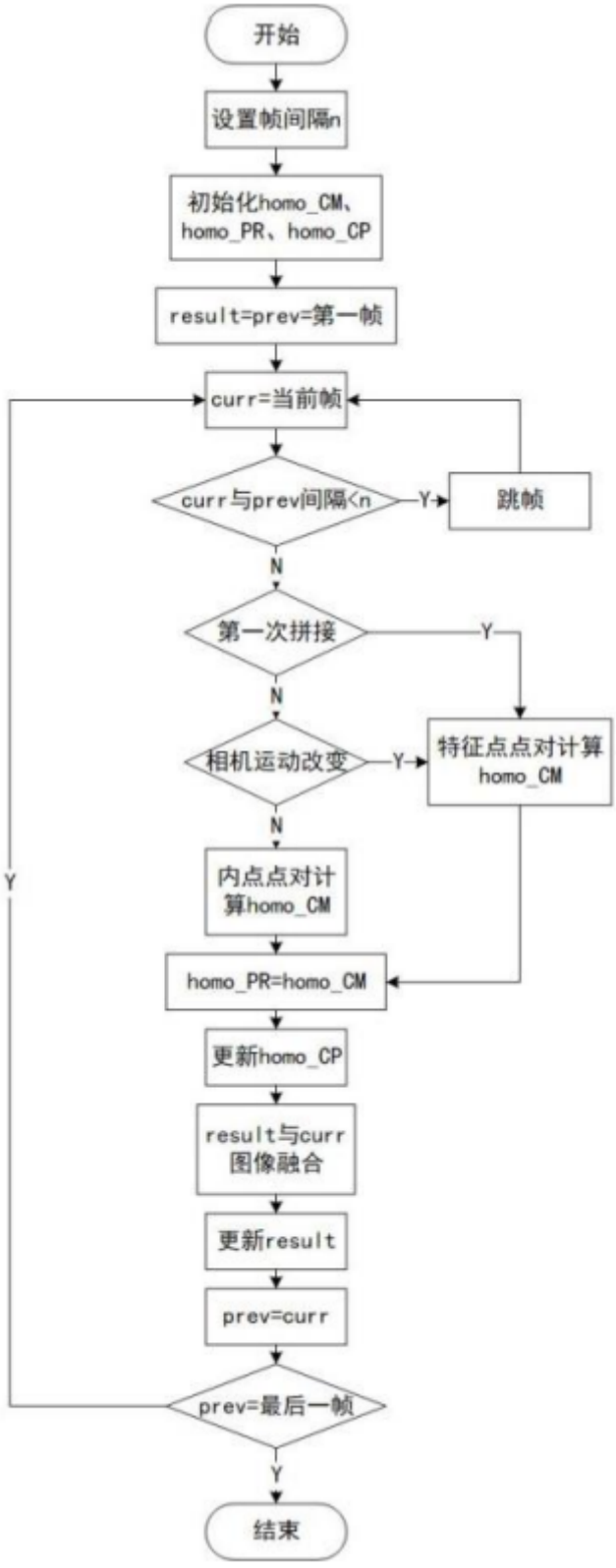
评价分数	主观质量尺度	主观妨碍尺度
0.9~1.0	非常好	微量错位/无错位，拼接痕迹不易察觉
0.8~0.9	很好	微量错位/亮度差异不明显
0.7~0.8	好	基本无错位/有稍许亮度差异
0.6~0.7	较好	存在很小错位/有较明显亮度差异 不妨碍信息获取
0.5~0.6	一般偏好	存在较小错位/有较明显亮度差异 基本不妨碍信息获取
0.4~0.5	一般偏差	存在错位/有明显亮度差异 稍妨碍信息获取
0.3~0.4	较差	存在较明显错位/有明显亮度差异 妨碍信息获取
0.2~0.3	差	错位严重/有明显亮度差异 较严重妨碍信息获取
0.1~0.2	很差	错位很严重/有明显亮度差异 严重妨碍信息获取
0.0~0.1	非常差	完全错位/亮度差异明显/拼接错误 几乎无法获取相关信息

图 2.6 主观评价和客观评分等级之间的映射关系

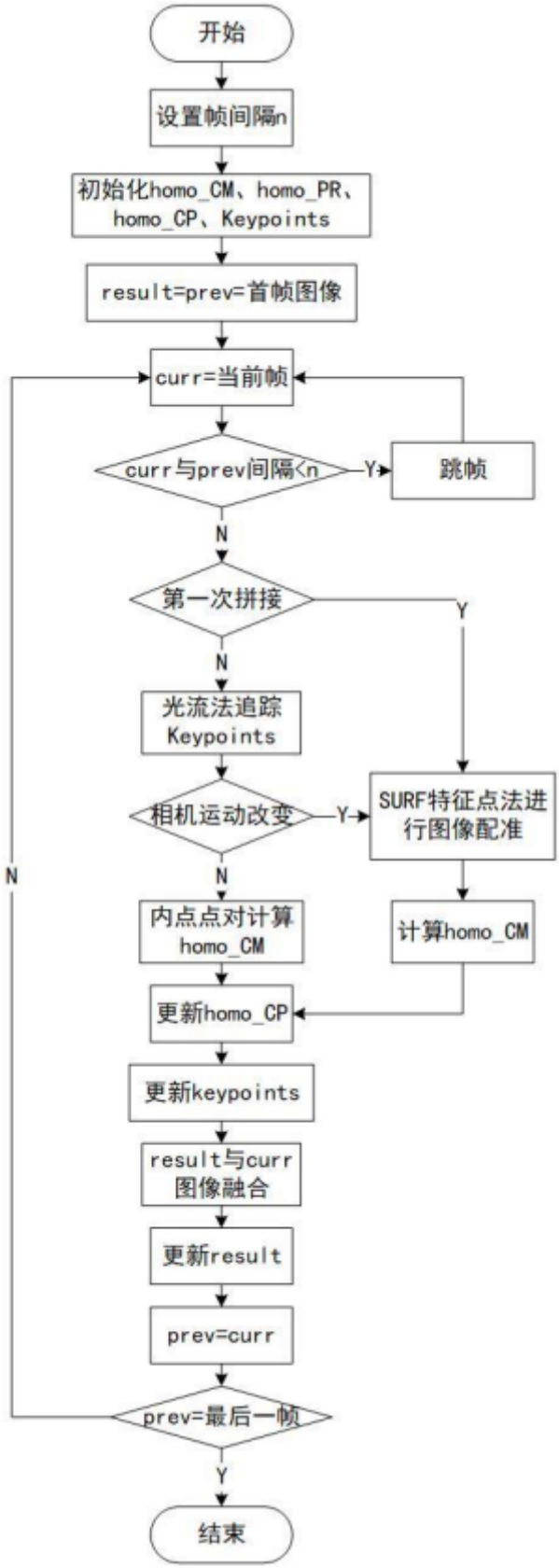
此客观评分为SSIM法的评分结果

程序流程

视频拼接直接法



视频拼接的快速实现算法


开始
设置帧间隔n
初始化homo_CM、homo_PR、
homo_CP、Keypoints
result=prev=首帧图像
curr=当前帧
curr与prev间隔<n
Y
跳帧
N
第一次拼接
Y
N
光流法追踪
Keypoints
相机运动改变
Y
SURF特征点法进
行图像配准
N
N
内点点对计算
homo_CM
计算homo_CM
更新homo_CP
更新keypoints
result与curr
图像融合
更新result
prev=curr
prev=最后一帧
Y
结束