

基于置信传播的双目匹配算法

Markov Random Field (MRF)

- 见 [Paper & Summary / MRF.md](#)

尚不清楚的表述

- ☒ spatial line process 线性函数
- ☒ spatial binary process 二值函数
- ☒ outlier process 离群点处理的robust function
- ☒ first-order neighbor 一阶邻接点 知某个像素块上下左右的四个点

双目相机模型建立

- 三层耦合的马尔可夫随机场
 - D: 双目平滑视差 (reference view)
 - L: 深度不连续情况 (同上)
 - O: 遮挡发生情况
- $P(D, L, O | I) = P(I | D, L, O) * P(D, L, O) / P(I)$ 公式解释
 - 左边: 在给定观测图像I的情况下, 视差/深度不连续/遮挡 的可能情况
 - 相当于给定观测图像, 有关深度划分的这些信息 (D, L, O) 的分布给出 (后验)
 - $P(I | D, L, O)$ 一个似然: - 给定D, L, O, 一个观测图像I符合给定的三个特征的似然
 - 使用 $\exp(-F(s, ds, I))$ 来建模 (对所有不属于遮挡点的像素s求连乘)
 - F为图像匹配的cost函数, 与像素的匹配差值有关
 - 注意 $P(I | D, O, L) = P(I | D, O)$
 - 这实际上是一个条件独立式, 似然与单个像素有关, 深度不连续属于像素间关系, 并且遮挡处的像素似然无法很好定义
 - **似然一般是后验这个条件式的条件, 结果互反**, 而本论文中, 似然的形式较为简单
 - $P(D, L, O)$ 一个先验:
 - 先验的作用是用于: 约束条件的加入, 约束条件一半存在于先验项中
 - 此处使用MRF理论, 使用卷积样的模型
 - 首先, $P(D, L, O) = P(D, L)P(O)$ (近似简化)
 - $P(D, L)$ 为深度差异与深度不联系的联合分布, 而联合分布在MRF中就是使用势函数的Gibbs分布表达的
 - G为所有像素的集合, N为s的邻接 (Neighbor) 点集合, t属于N
 - 则 $P(D, L)$ 使用 $ds(s处的D)$, $dt(t处的D)$, $Ist(s, t 确定的深度不连续情况)$ 三个变量的团势函数建模, 式子比较复杂
 - $P(O)$ 则为O单独的势函数:
 - 注意此处, O 单独的势函数包含两部分:
 - A: O本身决定的势函数值 (在s处是否存在遮挡0, 1数字值得函数)
 - B: O与其他位置的联系 (领域内)
 - 而第二部分 (B) 通常是可简化的, 那么实际上 $P(O)$ 只与像素s本身有关
 - 由此, 先验和似然全部建模完成, 只剩下归一化常数 $P(I)$, 我们无需对此建模

简洁化

这一部分比较难以理解，模型有一定的抽象度

- 一个MAP问题
 - MAP在此的作用是 MAP-MRF问题最终需要求解：
 - D, O, L 我们希望后验 $P(D, O, L | I)$ 最大
 - 即给定一个输入图像，计算出的 (D, O, L) 概率最大
 - 但我们实际上需要的是由原始的左右图像得到深度图划分（并进行相应的cost aggregation）
 - 问题也即求最佳的 D, O, L ，虽然论文的公式没有显式地写 $\arg \max$ ，个人认为这仍然是 $\arg \max$ 问题
 - 使用鲁棒代价函数 (Robust Function) 的鲁棒估计(Robust estimator)
 - 原来我们的MAP问题是求解 后验(上式) 对应的先验 / 似然 乘积的最大值对应 D, O, L 参数，其中存在几个问题

O, L 均是二值的（数字量，只有0, 1两种情况） 潜在的outliers

- 根据另一篇论文 [M.J.Black et al. The Outlier Process: Unifying Line Processes and Robust Statistics](#)
- 对于binary line process 与 analog line process 有这样的定义：
 - Binary line process 就是二值函数，比如discontinuity 关于像素位置 s, t 的值，存在不连续为0，不存在为1
 - 转化为analog line process 就相当于一种插值
- 论文做了更骚的事情：通过robust function 将binay line process -> analog line process -> elimination.
- 直接通过异常处理，建模了离群点观测模型，并且使用robust function代替了两个process
- 更加简洁了

• 置信传播

- 网络构建：将MRF转化成Markov网，这一步很好理解：
- 对于像素网络，原来存在的信息是：观测 I ，视差 ds ，像素本身 s
- 建模：对像素 s 及其对应的随机变量 ds ，存在一个隐结点 xs 与之连接
- 这个 xs 表达的意义就是 s 处的 ds 值
- “private”观测结点 ys ，代表的是 s 处的 I 上的信息
- 那么在双目模型建立过程中，势函数可以更改为与 xs, ys 有关的函数
- 比如 $\psi(s, ds, I)$ (即 $P(I | D, O)$)，可以改写为 $\psi(xs, ys)$
- 需要注意的是，此处换成了矩阵以及向量表达，比如 ys 就是一个向量
- ys 与对不同的 xs 观测有关，具体的还没搞清楚
- **消息更新规则**（不是很好理解）
 - 两种更新方式都介绍一下，文中使用的max product (最大化乘积)
 - sum product 乘积和：
 - summation($\phi(x_i, y_i)$) 是什么意思？(网上资料的第一个求和项)
 - 其中的 $\phi(\dots)$ 为似然函数
 - x_i 的取值为disparity，就是 d_i ，是可以变的
 - sum product 的似然项就是对所有的 x_i 可能取值的似然进行求和
 - max product 论文中的方法：

- 首先，除去因子 k 以外， $mst(i+1)$ 右边由三项的乘积合成：1. 极大似然 2. y_s 与 x_s 构成的自消息 3. 其他结点消息
- 可以知道，message的结构：其他消息 * 自消息 * 似然，这在sum product中也是一样的
- 不禁想问：message里的传递的东西到底是什么？首先肯定没有量纲
- 那么可以估计，要么就是势函数值（也没有量纲），要么就是概率（因为似然是一种特殊的概率）
- 由网上的资料，**置信度就是边缘概率，而置信度由消息直接得到**，那么可以推断，消息就是概率
- 我们对所有结点进行这样的消息更新，迭代 T 次（ T 为迭代次数上限），如果置信度收敛，则开始计算结点置信度
- 置信度就是边缘分布，那么由每个结点的边缘分布（关于 d_s 的边缘分布）
 - 关于这个边缘分布，再补充一下，由于 x_s 为 d_s 的隐结点，对 x_s 求MAP就是隐式地求 d_s 的MAP
 - 这个后验边缘分布表明了什麼？还记得一开始的建模吗？ $P(D, O, L | I)$
 - 在此处就是给定图像观测时，某一具体点的概率，而开始的建模是全图的联合分布
- 求置信度最大时（最大后验）对应的 x_s 取值，就是MAP估计值
- 这样求出的 d_s ，将会是一个非常贴近最优分配值的值

与其他的信息进行结合

- 与segment类的方法进行结合
 - 以segmentation的结果作为一个先验的 soft constraint
 - *** Mean-Shift *** 算法得出的 segmentation

对于BP算法的评价

- 非对称性，论文中提到高低置信度结点相互传递消息的信息熵的不同
 - 由于我没有学过信息论相关知识，只是粗略地了解信息熵的理论
 - 对于这个问题我个人的理解是：信息量与信息熵直接相关
 - 信息熵大将表示信息量大，而代表信息本身的概率较小，也就是说
 - 低置信度结点传递到高置信度结点的消息（代表的概率）很小
 - 也就是说，低置信度结点对高置信度结点的影响低（毕竟低置信度代表的出错概率大）
 - 而其他的算法，比如FW, SW等算法，其取的support一般是symmetric的
 - 也就是说，对于一个support下的所有结点，几乎对称的方式进行相关的讨论
 - 比如说汉明距离或是开方后的距离，是具有对称性的
- 自适应性：主要体现在对置信度低或者匹配差异大的结点存在更大力度的惩罚上

最后的总结与感想

- 后验概率的建模过程非常优雅，虽然刚开始看的时候让人感觉云里雾里的
- 看懂了之后便觉得妙极。妙主要妙在：
 - 后验概率的分解以及简化讨论（独立性近似以及 L 的消除）
 - Gibbs分布模型函数的建立（似然和先验的构建）
 - 二值 -> 模拟（连续化）-> 使用robust function 消除难以讨论的变量并且完成outliers的建模
 - MRF -> Markov Network，后验的进一步转化
- 本科生如何对复杂问题进行更加深入的建模？
 - 懂得的方法太少，概率论学的太浅

- 个人很喜欢贝叶斯学派的相关理论，但是在概率论中，这个并不作为学习的重点
- 这种结合其他信息（特别是人的知识）进行推理，对先验模型进行观测修正的思想非常优雅
- 但是在这方面的训练是欠缺的，如何进行概率建模
- 如何分解各变量的关系，如何简化问题，如何选择概率表达的方式，如何进行**信息融合**
 - 特别是信息融合，如何利用其他的已有知识，迁移到这个问题下，变成这个问题的信息或是约束？
 - 比如说单目视觉的深度估计，人之所以能做到只用一只眼睛的深度估计
 - 就是因为人脑中有许多其他的信息，比如所看到的物体---（判断物体的性质，估计物体的形态）
 - 人脑可以综合视觉信息，对所有的物体 / 脑中的先验知识 进行融合，得到一个较为准确的估计

TODO

- ☒ Conditional Random Field (CRF)
- ☒ Belief Propagation
- ☐ Hidden Markov Model (HMM)
- ☐ Boltzman Distribution
- ☐ Mean Shift