

文章编号: 1001-0920(2011)06-0911-05

视觉同步定位与地图重建——基于先验信息的 SIFT 匹配算法

孟旭炯, 蒋荣欣, 陈耀武

(浙江大学 数字技术及仪器研究所, 杭州 310027)

摘要: 鉴于尺度不变特征转换(SIFT)匹配算法存在计算效率不高且容易出现误匹配的问题, 针对视觉同步定位与地图重建, 提出了一种基于先验信息的 SIFT 匹配算法. 该算法首先根据机器人和特征点的相对距离变化来预测尺度空间的变化; 然后根据机器人和特征点的当前状态来预测特征点的图像位置; 最后在预测的图像位置进行 SIFT 匹配. 实验结果表明该算法能显著提高 SIFT 匹配的计算效率和准确性.

关键词: 视觉同步定位与地图重建; 特征匹配; 尺度空间; 扩展卡尔曼滤波

中图分类号: TP249

文献标识码: A

Prior information constrained SIFT matching algorithm for visual simultaneous localization and mapping

MENG Xu-jiong, JIANG Rong-xin, CHEN Yao-wu

(Institute of Advanced Digital Technologies and Instrumentation, Zhejiang University, Hangzhou 310027, China.

Correspondent: MENG Xun-jiong, E-mail: mxj@mail.bme.zju.edu.cn)

Abstract: The scale invariant feature transform(SIFT) algorithm has the problem of computational inefficiency and mismatch. Therefor, a prior information constrained SIFT matching algorithm is proposed for the visual simultaneous localization and mapping(vSLAM) applications. Firstly, the scale space is predicted according to the relative distance from the robot to the feature. Then the feature position is estimated according to the state of both the robot and the feature. Finally, sift matching is conducted within the predicted image region. The experiment results show that the proposed algorithm can achieve better computational efficiency and matching performance.

Key words: visual simultaneous localization and mapping; feature matching; scale space; extended Kalman filter

1 引言

视觉同步定位与地图重建(vSLAM)是移动机器人利用视觉测量装置在未知环境中创建地图, 同时也在所创建地图中估计自身位置的过程. 相对于传统的激光测距仪等装置, 视觉测量装置具有廉价、体积小、功耗低等特点, 并且能够提供更丰富的信息和更接近于人类的感知方式, 因此, vSLAM 在近几年受到了越来越多的关注^[1-3]. 特征匹配是 vSLAM 中的一大难点. 在机器人的移动过程中, 由于普遍存在着图像的旋转、缩放等情况, 基于固定模板的匹配算法容易失效. 而尺度不变特征转换(SIFT)匹配算法^[4]由于对尺度、光照、旋转等不敏感, 在 vSLAM 中得到了比较广泛的应用^[5-7]. 但是 SIFT 算法存在着以下两个问题: 一是需要在多个尺度空间上进行极值检测, 导致其计

算复杂度较高; 二是在图像中存在多个相似区域的情况下, 容易出现误匹配. 对于第 1 个问题, 可以通过降低描述符的维数^[8]或通过积分图像的 haar 求导^[9]来加以改善, 但是这些算法均以降低匹配的效果为代价. 对于第 2 个问题, 通常需要依赖 RANSAC 算法^[10]或极线约束条件^[11]. 此外, 基于彩色空间的 CSIFT 算法^[12]、考虑全局运动信息的算法^[13]也都可以取得较好的效果, 但是又会带来计算复杂度的进一步增加.

针对 SIFT 算法存在的问题和 vSLAM 这一特定应用, 提出了一种改进的 SIFT 匹配方法: 通过对尺度空间和图像位置进行预测, 可以避免在全部尺度空间范围内和整个图像空间上进行穷举式的特征匹配. 与此相关的工作主要有主动搜索技术^[1]和预测多分辨率描述符技术^[4].

收稿日期: 2010-03-30; 修回日期: 2010-06-01.

基金项目: 浙江省科技计划重大科技专项重点项目(2006C11200).

作者简介: 孟旭炯(1981—), 男, 博士生, 从事移动机器人、单目视觉 SLAM 的研究; 陈耀武(1963—), 男, 教授, 博士, 从事嵌入式系统、多媒体领域等研究.

2 SIFT 算法

SIFT 算法的基本思想是建立高斯差分尺度空间并从中提取极值点; 然后进行关键点位置的精确定位和方向参数的计算; 最后生成 SIFT 描述符。

2.1 尺度空间理论

为了得到多尺度的图像特征, 首先需要建立起多尺度图像空间, 可以由一个 O 组且每组 S 层的图像金字塔来表示. 下 1 组图像由上 1 组图像的降采样得到. 假设第 0 组的分辨率为 $M \times N$, 则第 i 组的分辨率为

$$\lfloor M/2^i \rfloor \times \lfloor N/2^i \rfloor. \quad (1)$$

同组内各层具有相同的分辨率, 下 1 层的图像通过对上 1 层图像的高斯卷积得到, 即

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}}, \quad (2)$$

$$L(x, y, \sigma) = G(x, y, \sigma)I(x, y), \quad (2)$$

其中 σ 与组坐标 o 和层坐标 s 相关, 可以看作尺度空间坐标

$$\sigma(o, s) = \sigma_0 \cdot 2^{o+s/S}, \quad (3)$$

σ_0 为基准尺度因子, 取值为

$$\sigma_0 = 1.6 \times 2^{1/S}. \quad (4)$$

在文献[4]中, 起始组坐标 O_{\min} 设为 -1 , 最大层坐标 S 设为 3. 为了得到稳定的关键点, 需要在差分尺度空间进行极值点的检测. 差分尺度空间可以由尺度空间中的相邻图像相减来生成, 有

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma))I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma). \quad (5)$$

2.2 关键点定位和方向分配

判断一个采样点是否为极值点, 需要将它与其周边的相邻点进行比较, 包括同一层上的 8 个相邻点和上下两层各 9 个相邻点. 得到极值点后, 通过对尺度空间函数 $D(x, y, \sigma)$ 的泰勒展开来拟合曲面的极值, 从而精确定位关键点的位置. 关键点的主方向由其相邻点的梯度方向分布决定, 梯度的计算公式为

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}, \quad (6)$$

$$\theta(x, y) = \tan^{-1} \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)}. \quad (7)$$

其中: $m(x, y)$ 为梯度的模, $\theta(x, y)$ 为梯度的方向.

2.3 描述符生成

为了保持旋转不变性, 首先需要将坐标轴转换到关键点的主方向; 然后计算以关键点为中心的 16×16 邻域内的每个点的梯度; 最后将这块区域划分为 16 个 4×4 的子域, 计算每个子域在 8 个方向上的梯度直方图, 即可得到 $16 \times 8 = 128$ 维的 SIFT 描述符。

3 基于先验信息的 SIFT 匹配算法

3.1 尺度空间预测

从第 2.1 节可以看出, SIFT 算法的复杂度与尺度空间的大小直接相关. 尺度空间越小, 所需要的下采样次数和高斯卷积的次数就越少, 计算效率就越高. 因此, 希望通过对特征点尺度空间的变化进行预测来避免生成没有特征点存在的尺度空间。

尺度空间的变化与特征点到摄像头之间的相对距离有关, 如图 1 所示. 设特征点 P 高度为 h , 在 i 时刻和 j 时刻, 摄像头与特征点 P 的距离分别为 d_i, d_j , 由此可以估算出 P 在 j 时刻相对于 i 时刻的图像尺度的变化为

$$\frac{u_j}{u_i} \approx \frac{d_i}{d_j}. \quad (8)$$

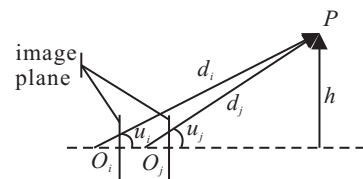


图 1 根据距离变化计算尺度空间变化

在 vSLAM 应用中, 由于机器人和特征点的位置均可以通过估计得到, 可以计算出机器人和特征点相对距离的变化, 根据式 (8) 即可知道尺度空间的变化情况。

假设在 i 时刻, 机器人检测到一个特征点, 用锚点齐次坐标^[15]表示其空间位置为

$$\begin{bmatrix} x_f \\ y_f \\ z_f \end{bmatrix} = \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} + \frac{1}{\omega_i} \begin{bmatrix} u_i \\ v_i \\ w_i \end{bmatrix}. \quad (9)$$

其中: (x_i, y_i, z_i) 为此时机器人的空间位置; $(u_i, v_i, w_i, \omega_i)$ 为特征点在摄像头坐标系中的齐次坐标表示, ω_i 为深度尺度因子. 此时机器人与特征点的最近距离 $d_{i-\min}$ 和最远距离 $d_{i-\max}$ 分别为

$$d_{i-\min} = \frac{1}{\omega_i + 2\sigma_i^a} - 2\sigma_i^r, \quad (10)$$

$$d_{i-\max} = \frac{1}{\omega_i - 2\sigma_i^a} + 2\sigma_i^r.$$

其中: σ_i^r 表示该时刻机器人空间位置的最大不确定度, σ_i^a 表示特征点空间位置的最大不确定度, 如图 2 所示。

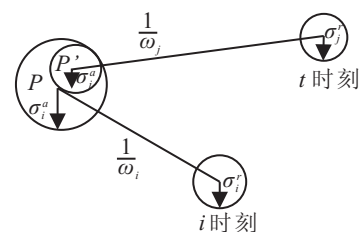


图 2 机器人和特征点的相对距离

在 j 时刻, 机器人运动到 (x_j, y_j, z_j) 点. 此时机器人与特征点的最近距离 $d_{j-\min}$ 和最远距离 $d_{j-\max}$ 分别为

$$\begin{aligned} d_{j-\min} &= \sqrt{(x_j - x_a)^2 + (y_j - y_a)^2 + (z_j - z_a)^2} - \\ &\quad 2\sigma_j^a - 2\sigma_j^r, \\ d_{j-\max} &= \sqrt{(x_j - x_a)^2 + (y_j - y_a)^2 + (z_j - z_a)^2} + \\ &\quad 2\sigma_j^a + 2\sigma_j^r. \end{aligned} \quad (11)$$

根据式(10)和(11)可以计算出特征点在 j 时刻相对于 i 时刻的图像尺度空间的变化范围 $[r_{\min}, r_{\max}]$, 即

$$\begin{aligned} r_{\min} &= d_{j-\min}/d_{i-\max}, \\ r_{\max} &= d_{j-\max}/d_{i-\min}. \end{aligned} \quad (12)$$

利用式(3)的反函数, 可以推导出此时特征点的图像组坐标 o 和组内层坐标 s .

3.2 特征点图像位置预测

从第2.3节可以看出, 每一个特征点描述符的生成均需要计算 16×16 邻域的梯度直方图, 因此特征点越多, 计算量就越大. 如果在生成描述符之前能够预先知道特征点的可能分布区域, 则只生成预测区域内的特征点的描述符即可达到减少计算量的目的.

假设机器人的状态向量为 x_r , 特征点的状态向量为 x_a , 测量方程为

$$z = h(x_r, x_a). \quad (13)$$

在 i 时刻, 机器人的状态估计为 \hat{x}_r , 特征点状态估计为 \hat{x}_a , 根据式(13), 特征点的图像位置为

$$\hat{z} = h(\hat{x}_r, \hat{x}_a). \quad (14)$$

由于机器人状态和特征点状态均是不确定的, 特征点的图像位置也是不确定的, 可以用协方差来表示这种不确定性. 假设机器人状态向量的协方差为 P_r , 特征点状态向量的协方差为 P_a , 对测量方程(13)在 (\hat{x}_r, \hat{x}_a) 处进行泰勒展开, 有

$$\Delta z \approx \frac{\partial h}{\partial x_r}(x_r - \hat{x}_r) + \frac{\partial h}{\partial x_a}(x_a - \hat{x}_a). \quad (15)$$

则 z 的协方差 P_z 为

$$\begin{aligned} P_z &= E[(z - \hat{z})(z - \hat{z})^T] \approx \\ &\frac{\partial h}{\partial x_r} E[(x_r - \hat{x}_r)(x_r - \hat{x}_r)^T] \frac{\partial h}{\partial x_r}^T + \\ &\frac{\partial h}{\partial x_r} E[(x_r - \hat{x}_r)(x_a - \hat{x}_a)^T] \frac{\partial h}{\partial x_a}^T + \\ &\frac{\partial h}{\partial x_a} E[(x_a - \hat{x}_a)(x_r - \hat{x}_r)^T] \frac{\partial h}{\partial x_r}^T + \\ &\frac{\partial h}{\partial x_a} E[(x_a - \hat{x}_a)(x_a - \hat{x}_a)^T] \frac{\partial h}{\partial x_a}^T = \\ &\frac{\partial h}{\partial x_r} P_r \frac{\partial h}{\partial x_r}^T + \frac{\partial h}{\partial x_r} P_{ra} \frac{\partial h}{\partial x_a}^T + \end{aligned}$$

$$\frac{\partial h}{\partial x_a} P_{ar} \frac{\partial h}{\partial x_r}^T + \frac{\partial h}{\partial x_a} P_a \frac{\partial h}{\partial x_a}^T. \quad (16)$$

考虑到测量噪声 M , 则 z 的协方差为

$$\begin{aligned} P_z &= \frac{\partial h}{\partial x_r} P_r \frac{\partial h}{\partial x_r}^T + \frac{\partial h}{\partial x_r} P_{ra} \frac{\partial h}{\partial x_a}^T + \\ &\frac{\partial h}{\partial x_a} P_{ar} \frac{\partial h}{\partial x_r}^T + \frac{\partial h}{\partial x_a} P_a \frac{\partial h}{\partial x_a}^T + M. \end{aligned} \quad (17)$$

由式(14)和(17)确定的区域即为特征点的预测图像位置. 图3表示对特征点的图像位置进行预测得到的结果. 图3(a)中的+表示上一帧时实际检测到的特征点位置, 图3(b)中的椭圆表示预测的特征点分布区域, 实心点表示预测位置的中心, +表示实际检测到的特征点位置, 可以看出预测区域能够覆盖实际特征点所在的位置.



(a) 上一帧特征点实际位置 (b) 当前帧特征点实际位置

图3 特征点图像位置预测

3.3 基于位置限制的特征点匹配

对于两个特征点 p_i 和 p_j , 其欧氏距离定义为

$$D_{i,j} = \|p_i - p_j\|. \quad (18)$$

如果 $D_{i,j}$ 满足

$$D_{i,j} < 0.6 \times D_{i,k}, \quad k \in N \text{ 且 } k \neq j, \quad (19)$$

则认为 p_i 和 p_j 是匹配的. 从以上判断标准可以看出, 在局部图像相似的情况下, 出现误匹配是不可避免的. 为此, 限定特征匹配只能在 p_i 和存在于其预测区域内的特征点之间进行, 以达到降低误匹配概率的目的.

4 仿真实验

设定摄像头的分辨率为 640×480 pixels, 帧率为 15 fps, 采用文献[16]提供的算法进行校正. 获取到图像序列后在台式机上验证, 台式机配置为 2.66 G 双核 CPU, 2 G 内存. 控制摄像头沿直线运动, 速度为 0.5 m/s, 为了抵消抖动等造成的干扰, 测量噪声设为 3 pixels. 以参考文献[1]提出的 EKF-SLAM 算法为基本框架, 分别实现本文提出的匹配算法(方法1)和文献[6]提出的匹配算法(方法2). 特征深度倒数的初始值及其标准差设置为 $(\omega, \sigma_\omega) = [0.5, 1]$. SIFT 相关参数设置为起始组坐标 $O_{\min} = 0$, 最大组坐标 $O_{\max} = 6$, 最大层坐标 $S = 3$.

不同尺度空间的特征点具有不同的匹配效果^[17], 为此设计两个实验, 分别选取 $3 < \sigma < 7$ 和 $\sigma > 5$ 两种类型的特征点作为匹配测试的基准特征点. 图4给出了两种条件下基准特征点的分布情况.

(a) $7 > \sigma > 3$ (b) $\sigma > 5$

图4 基准特征点分布

图5(a)和图5(b)给出了两种方法在两种环境下的匹配准确率对比. 特征匹配准确率由下式衡量:

$$p = \frac{\# \text{correct matches}}{\# \text{total possible matches}}. \quad (20)$$

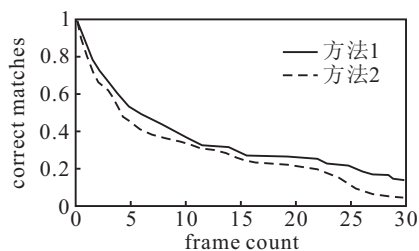
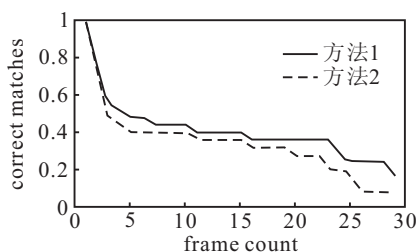
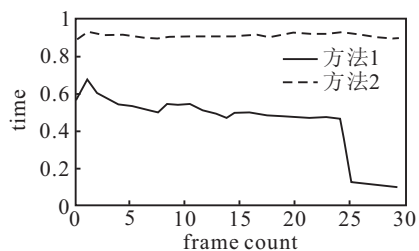
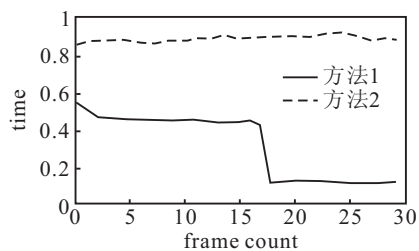
(a) 匹配准确性曲线($3 < \sigma < 7$)(b) 匹配准确性曲线($\sigma > 5$)(c) 匹配时间曲线($3 < \sigma < 7$)(d) 匹配时间曲线($\sigma > 5$)

图5 实验结果

可以看出, 本文提出的算法比文献[6]提出的算法具有更高的匹配准确率. 同时可以看出, 选取 σ 较大的特征点可以得到更好的匹配效果, 这与文献[17]的结论是一致的.

图5(c)和图5(d)给出了每一帧上特征点匹配所需时间的对比. 在初始的一段时间里, 需要在整个尺度空间内进行特征点匹配, 因此该阶段计算量的减少主要来自于特征点位置预测, 且减少的程度与特征点数相关: 对于 $3 < \sigma < 7$ 的情况, 它能够带来约400ms的计算量下降; 对于 $\sigma > 5$ 的情况, 它能够带来约450ms的下降. 在尺度空间发生变化后(分别对应图5(c)中的第25帧和图5(d)中的第18帧), 可以看到计算量再次出现跳跃式下降, 这是因为预测到特征点都将分布在第1个及其之上的图像组, 不再需要对第0个图像组进行处理.

5 结论

本文提出了一种改进的SIFT匹配算法. 为了降低计算复杂性, 首先利用vSLAM算法输出的机器人和特征点的位置信息来估计尺度空间的变化, 从而避免生成无特征点分布的尺度空间; 然后估计待匹配特征点在图像空间的可能分布区域, 从而避免在整个图像域内进行特征点的提取. 为了降低误匹配的概率, 对参与匹配的特征点进行限制, 每个待匹配特征点只与其预测位置内的特征点进行匹配. 实验结果表明本文提出的算法是有效的.

参考文献(References)

- [1] Davison A J, Reid I D, Molton N D, et al. Monoslam: Real-time single camera slam[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2007, 29(6): 1052-1067.
- [2] Paz L M, Pinies P, Tardos J D, et al. Large-scale 6-DOF SLAM with stereo-in-hand[J]. IEEE Trans on Robotics, 2008, 24(5): 946-957.
- [3] Solà J, Monin A, Devy M, et al. Fusing monocular information in multicamera SLAM[J]. IEEE Trans on Robotics, 2008, 24(5): 958-968.
- [4] Lowe D. Distinctive image features from scale-invariant keypoints[J]. Int J of Computer Vision, 2004, 60(2): 91-110.
- [5] Gil A, Reinoso O, Payá L, et al. Managing data association in visual SLAM using SIFT features[J]. Int J of Factory Automation, Robotics and Soft Computing, 2007: 179-184.
- [6] Chen C H, Chan Y P. SIFT-based monocular SLAM with inverse depth parameterization for robot localization[C]. IEEE Workshop on Advanced Robotics and its Social Impacts. Hsinchu: IEEE Press, 2007.
- [7] Pradeep V, Medioni G, Weiland J. Visual loop closing using multi-resolution SIFT grids in metric-topological SLAM[C]. IEEE Conf on Computer Vision and Pattern Recognition. Miami: IEEE Press, 2009: 1438-1445.

- [8] Ke Y, Sukthankar R. PCA-SIFT: A more distinctive representation for local image descriptors[C]. Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Washington DC: IEEE Press, 2004: 506-513.
- [9] Bay H, Tuytelaars T, Gool Van L. SURF: Speeded up robust features[C]. Proc of the European Conf on Computer Vision. Graz: Springer Press, 2006.
- [10] Fischler M A, Bolles R C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography[J]. Communications of the ACM, 1981, 24(6): 381-395.
- [11] Folkesson J, Christensen H. SIFT based graphical slam on a packbot[J]. Field and Service Robotics, 2008, 42: 317-328.
- [12] Abdel-Hakim A E, Farag A A. CSIFT: A sift descriptor with color invariant characteristics[C]. Proc of the IEEE Conf on Computer Vision and Pattern Recognition. New York: IEEE Press, 2006: 1978-1983.
- [13] Mortensen E N, Deng H L, Shapiro L. A SIFT descriptor with global context[C]. Proc of the IEEE Conf on Computer Vision and Pattern Recognition. San Diego: IEEE Press, 2005: 184-190.
- [14] Chekhlov D, Pupilli M, Mayol-Cuevas M, et al. Real-time and robust monocular SLAM using predictive multi-resolution descriptors[C]. Proc of the 2nd Int Symposium on Visual Computing. Lake Tahoe: Springer Press, 2006.
- [15] Solà J. Consistency of the monocular EKF-SLAM algorithm for 3 different landmark parametrizations[C]. Proc of the IEEE/RSJ Int Conf on Robotics and Automation. Anchorage: IEEE Press, 2010: 2570-2575.
- [16] Jean-Yves B. www.vision.caltech.edu/bouguetj/calib_doc/index.html[Z].
- [17] Wang X, Zhang H. Good image features for bearing-only SLAM[C]. Proc of the IEEE/RSJ Int Conf on Intelligent Robots and Systems. Beijing: IEEE Press, 2006: 2576-2581.

(上接第910页)

参考文献(References)

- [1] Sclaroff S, Liu L. Deformable shape detector and description via model-based grouping[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2001, 23(6): 475-488.
- [2] Kurugollu F, Sankur B, Harmanci A E. Color image segmentation using histogram multithresholding and fusion[J]. Image and Vision Computing, 2001, 19(13): 915-928.
- [3] Pham T D. Image segmentation using probabilistic fuzzy C -means clustering[J]. Proc of Int Conf on Image Processing, 2001, 21(3): 722-725.
- [4] Schoenemann T, Cremers D. A combinatorial solution for model-based image segmentation and real-time tracking[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2009, 32(7): 1153-1164.
- [5] Kass M, Witkin A, Terzopoulos D. Snakes — Active contour models[J]. Int J of Computer Vision, 1987, 1(4): 321-331.
- [6] 吕遐东, 黄心汉. 基于动态 Snake 模型的机械手运动轨迹视觉跟踪[J]. 控制与决策, 2006, 21(10): 1143-1147.
(Lv X D, Huang X H. Visual tracking of robotic manipulator motion trajectory using dynamic active contour model[J]. Control and Decision, 2006, 21(10): 1143-1147.)
- [7] Xu C Y, Prince J L. Snakes, shapes and gradient vector flow[J]. IEEE Trans on Image Processing, 1998, 7(3): 359-369.
- [8] Caselles V, Kimmel R, Sapiro G. Geodesic active contours[J]. Int J of Computer Vision, 1997, 22(1): 61-79.
- [9] Sethian J A. Evolution, implementation and application of level set and fast marching methods for advancing fronts[J]. J of Computational Physics, 2001, 169(2): 503-555.
- [10] Xu C Y, Yezzi A, Prince J. On the relationship between parametric and geometric active contours[C]. A silomar Conf on Signal, Systems and Computers. Pacific Grove, 2000: 483-489.
- [11] Osher S J, Sethian J A. Fronts propagation with curvature dependent speed: Algorithms based on Hamilton-Jacobi formulations[J]. J of Computational Physics, 1988, 79(1): 12-49.
- [12] Aubert G, Kornprobst P. Mathematical problems in image processing: Partial differential equations and the calculus of variations[M]. New York: Springer, 2002.
- [13] Zhao H K, Chan T, Merriman B, et al. A variational level-set approach to multiphase motion[J]. J of Computer Physics, 1996, 127(1): 179-195.
- [14] Kreyszig E. Differential geometry[M]. University of Dorendo Press, 1959: 2-11.
- [15] Li C M, Xu C Y, Gui C F, et al. Level set evolution without re-initialization: A new variational formulation[C]. IEEE(CVPR05). San Diego, 2005, 1: 430-436.