

〈图像处理与仿真〉

视频图像的 SIFT 特征点自适应提取算法

余宏生¹, 金伟其²

(1. 湖北理工学院数理学院, 湖北 黄石 435000;

2. 北京理工大学光电学院 光电成像技术与系统教育部重点实验室, 北京 100081)

摘要: 采用 SIFT 算法匹配视频图像帧前, 必须首先提取特征点。如果输入图像的大小和特性变化, 特征点的灰度阈值必须随之重新设置, 以避免过大的计算量和配准失败。提出了一种视频图像的特征点自适应提取算法。该算法能够将前一帧的视频图像的参数反馈到当前帧, 自动设置适当的特征点灰度阈值, 使得从当前帧提取的关键点的数量接近预期值。实验结果表明, 当输入图像改变时, 采用自适应设置阈值方法, 从视频帧提取的特征点的数量始终保持在预期值。该方法可以通过 SIFT 算法自适应地配准数字视频图像, 使特征点数量保持稳定, 避免配准失败, 减小计算量。

关键词: 视频图像; SIFT; 自适应; 关键点提取

中图分类号: TP391.4 文献标识码: A 文章编号: 1001-8891(2013)12-0768-05

SIFT Key-points Self-adaptive Extraction Algorithm for Video Images

YU Hong-sheng¹, JIN Wei-qi²

(1. School of Mathematics and Physics, Hubei Polytechnic University, Huangshi 435000, China; 2. MoE Key Laboratory of Photoelectronic Imaging Technology and System, School of Optoelectronics Beijing Institute of Technology, Beijing 100081, China)

Abstract: Before matching the video frames in Scale-Invariant Feature Transform (SIFT) algorithm, the key-points must be extracted firstly. If the size and characteristic of input images are changed, gray threshold of key-points must be reinstalled, to avoid extremely computation cost or failure in registration. In this paper, a self-adaptive SIFT key-points extraction algorithm for video images is developed. The algorithm can set appropriate gray threshold of key-points automatically by feeding parameter of previous frame back to present frame to make the number of key-points extracting from present frame close to the expected value. The experiments show that, when the input image is changed, the key-points number of the video frame always keep near the expected value by setting the threshold self-adaptively. The method makes it possible for digital video images to be registered self-adaptively by SIFT algorithm and the number of feature points remains stable so that the computation costs can be reduced while avoiding registration failure.

Key words: Video image, SIFT, self-adaptive, key-points extraction

0 引言

视频图像的自动配准时, 要求图像配准算法所采用的特征空间必须对存在多种几何失真的视频帧图像具有仿射不变性。David G. Lowe^[1-2]提出了一种基于尺度空间的、对图像缩放、旋转及仿射变换保持不变性的图像局部特征描述方法: 尺度不变特征

变换 (Scale-Invariant Feature Transform, SIFT) 算法。SIFT 算法提取的 SIFT 特征向量对平移、旋转、缩放等几何失真及亮度变化保持不变性; 对视差、仿射变换、压缩及噪声污染也保持一定程度的稳定性。SIFT 算法广泛应用于机器视觉^[3-4], 目标跟踪和目标检测^[5-6]。

采用 SIFT 算法提取特征点进行图像配准时,

收稿日期: 2013-08-29; 修改日期: 2013-11-02。

作者简介: 余宏生 (1968-), 男, 湖北黄冈人, 教授, 博士, 研究方向为光电成像与检测技术。E-mail: yhs99981@sina.com。

基金项目: 湖北省自然科学基金项目, 编号: 2011CDC107。

如果图像的像素大小和内容变化, 而特征点的灰度阈值不变, 在不同图像中提取的特征点的数量差别非常大。特征点过多将使得计算量呈几何级数增加; 特征点过少, 将减小配准精度, 甚至导致配准失败。因此, 当数字图像自动配准时, 特征点的灰度阈值必须根据图像的像素大小和图像特征适当调整, 尽可能通过一定数量的特征点以较低的计算量实现精确的配准。

本文提出一种视频图像的自适应 SIFT 特征点提取算法。首先, 介绍利用高斯金字塔图像及差分金字塔图像检测 SIFT 关键点的步骤; 然后, 介绍从关键点中提取特征点集的算法; 最后, 详细论述灰度阈值的自适应设置方法, 并采用实际视频图像验证该算法。

1 提取关键点的方法

SIFT 关键点提取是指在输入图像的差分高斯金字塔 (Difference of Gaussian Pyramid, DGP) 图像中找出满足灰度阈值要求的局部极值点的精确位置, 并生成特征向量集。首先, 在不同尺度的 DGP 图像中找出满足灰度阈值要求的局部极值点作为关键点 (Keypoints); 然后, 对关键点拟合 3-D 二次函数, 通过迭代运算剔除低灰度值的关键点, 并确定极值关键点所处的尺度和精确位置, 组成 SIFT 特征向量集。

1.1 构建高斯金字塔

要建立 DGP 图像, 首先必须建立高斯金字塔 (Gaussian Pyramid, GP) 图像。GP 图像是指通过将输入图像与不同的尺度变换核进行卷积运算, 得到在不同尺度下的图像金字塔图像。在不同尺度下进行关键检测, 有利于提取图像中具有不同尺度细节特征的特征点, 提高配准精度。Koenderink^[7]认为高斯卷积核是实现尺度变换的唯一变换核, 而 Lindeberg^[8]则证明在一系列合理的假设下高斯核是唯一的线性核。

假设可变尺度的高斯核函数为:

$$G(\sigma) = \frac{1}{2\pi\sigma^2} e^{-(i^2+j^2)/2\sigma^2} \quad (1)$$

则输入图像 $I(m,n)$ 在尺度 σ 下的图像为:

$$L(\sigma)(m,n) = G(\sigma)*I(m,n) \quad (2)$$

式中: “*”为卷积运算符。改变 σ 可获得 $I(m,n)$ 尺度空间的图像, 组成 GP 图像。

假设 GP 图像有 ξ 阶, 每阶有 ζ 层; p 、 q 分别表示阶、层的序号。第 0 阶第 0 层的尺度为 σ ; 每一阶第 q 层的尺度均为第 $q-1$ 层的尺度的 k 倍。为了使所有的特征点能够均匀分布在 DGP 图像的尺度

空间中, 要求第 p 阶第 0 层的尺度为第 $p-1$ 阶第 0 层的尺度的 2 倍, 即 $k^{\xi-2}\sigma=2\sigma$, 则 $k=2^{2-\xi}$ 。因此, 同一阶的第 q 层都是对第 $q-1$ 层进行方差为 $\sigma_{q-1 \rightarrow q}$ 的高斯平滑得到, 则第 p 阶 q 层的 GP 图像为:

$$\begin{aligned} L_{p,q}(\sigma_{p,q})(2^{1-p}m, 2^{1-p}n) = \\ G(\sigma_{q-1 \rightarrow q})*L_{p,q-1}(2^{1-p}m, 2^{1-p}n) \\ p=0,1,\cdots,\xi-1; \quad q=1,\cdots,\xi-1 \end{aligned} \quad (3)$$

第 0 阶第 0 层的 GP 图像为:

$$L_{0,0}(2m,2n)=G(\sigma)*\text{Interp}(I(m,n),2) \quad (4)$$

式中: $\text{Interp}(I,2)$ 表示采用双线性插值方法使图像行、列像素大小增大为原来的 2 倍。

第 p 阶第 0 层的 GP 图像为:

$$\begin{aligned} L_{p,0}(2^{1-p}m, 2^{1-p}n) = \\ \text{Downsample}(L_{p-1,q-2}(2^{2-p}m, 2^{2-p}n), 2) \\ p=1,2,\cdots,\xi-1 \end{aligned} \quad (5)$$

式中: $\text{Downsample}(I,2)$ 表示通过降采样使图像行、列像素大小减小为原来的 $1/2$ 。

第 p 阶、第 q 层 GP 图像的尺度为:

$$\begin{aligned} \sigma_{p,q} = k^{p(\xi-2)+q} \cdot \sigma \\ p=0,1,\cdots,\xi-1; \quad q=1,\cdots,\xi-1 \end{aligned} \quad (6)$$

高斯平滑函数的方差为:

$$\sigma_{q-1 \rightarrow q} = \sqrt{(k^q\sigma)^2 - (k^{q-1}\sigma)^2} = k^{q-1}\sigma\sqrt{k^2-1} \\ q=1,2,\cdots,\xi-1 \quad (7)$$

GP 图像可以采用式(3)~式(7)生成。图 1 为原始 Lena 图像, 图 2 为部分 GP 图像。



图 1 原始 Lena 图像 (像素大小 512×512)
Fig.1 Original Lena image (pixel size 512×512)



(a) $p=0, q=4$



(b) $p=1, q=4$



(c) $p=2, q=4$



(d) $p=3, q=4$

图 2 Lena 图像的部分 GP 图像

Fig.2 Some GP images of Lena image

1.2 建立差分高斯金字塔图像

由 ξ 阶，每阶 ξ 层的GP图像相邻层相减可以得到 ξ 阶，每阶 $\xi-1$ 层的DGP图像，如图3所示。

$$D_{p,q}(\sigma_{p,q})(2^{1-p}m, 2^{1-p}n) = L_{p,q+1}(\sigma_{p,q})(2^{1-p}m, 2^{1-p}n) - L_{p,q}(\sigma_{p,q})(2^{1-p}m, 2^{1-p}n) \quad p=0,1,\dots,\xi-1; q=0,1,\dots,\xi-2 \quad (8)$$

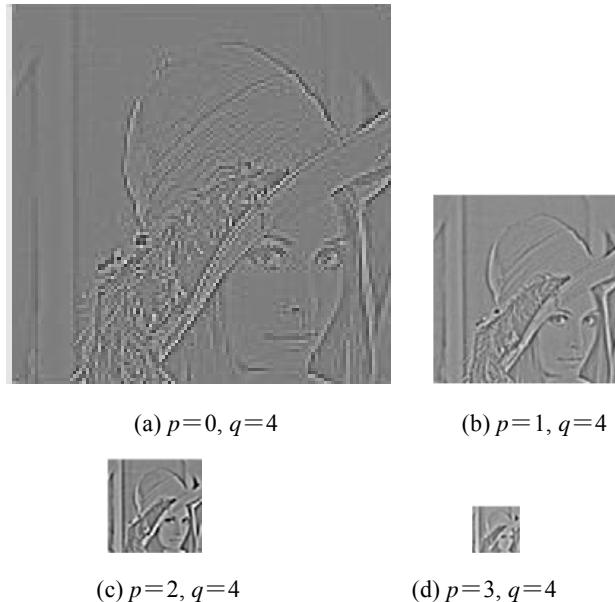


Fig.3 Some DGP images of Lena image

1.3 关键点检测

DGP图像的第 p 阶 q 层上的关键点是指在第 $p-1, p, p+1$ 阶、第 q 层的 3×3 邻域内的极值点，如图4所示。设第 p 阶、第 q 层的某像素点的灰度值为 $D(i,j,\sigma)$ ，若：

$$L_{\max} = \max(D(i \pm \tau, j \pm \tau, k^{\lambda} \sigma)) \quad \lambda, \tau = -1, 0, 1 \quad (9)$$

或者：

$$L_{\min} = \min(D(i \pm \tau, j \pm \tau, k^{\lambda} \sigma)) \quad \lambda, \tau = -1, 0, 1 \quad (10)$$

则 $D(i,j,\sigma)$ 为关键点。

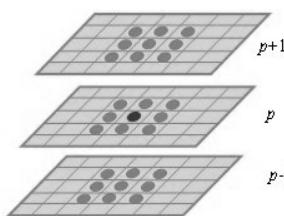


图4 DGP图像局部关键点检测

Fig.4 Local extrema detection in DGP images

1.4 关键点精确定位

2002年Brown和Lowe^[9-10]提出了对局部抽样点拟合3D二次函数，确定最大值的插值位置的方法。该方法采用Taylor级数展开方法将拟合的尺度空间函数 $D(i,j,\sigma)$ 展开到二次项，并对该函数求导，

求出关键点的极值精确位置。

$$D(i + \Delta i, j + \Delta j, \sigma + \Delta \sigma) = D(i, j, \sigma) + \frac{\partial D^T}{\partial X} X + \frac{1}{2} X^T \frac{\partial^2 D}{\partial X^2} X \quad (11)$$

式中： $X = (\Delta i, \Delta j, \Delta \sigma)^T$ 。

$$\frac{\partial D^T}{\partial X} = \begin{pmatrix} \frac{\partial D}{\partial i} & \frac{\partial D}{\partial j} & \frac{\partial D}{\partial \sigma} \end{pmatrix} \quad (12)$$

$$\frac{\partial D^2}{\partial X^2} = \begin{pmatrix} \frac{\partial^2 D}{\partial i^2} & \frac{\partial^2 D}{\partial i \partial j} & \frac{\partial^2 D}{\partial i \partial \sigma} \\ \frac{\partial^2 D}{\partial j \partial i} & \frac{\partial^2 D}{\partial j^2} & \frac{\partial^2 D}{\partial j \partial \sigma} \\ \frac{\partial^2 D}{\partial \sigma \partial i} & \frac{\partial^2 D}{\partial \sigma \partial j} & \frac{\partial^2 D}{\partial \sigma^2} \end{pmatrix} \quad (13)$$

要得到 $D(i + \Delta i, j + \Delta j, \sigma + \Delta \sigma)$ 的极值点，利用式(11)对 X 求导，并令：

$$\frac{\partial D(i + \Delta i, j + \Delta j, \sigma + \Delta \sigma)}{\partial X} = 0 \quad (14)$$

于是：

$$\hat{X} = -\frac{\partial^2 D}{\partial X^2} \frac{\partial D}{\partial X} \quad (15)$$

$$D(\hat{X}) = D + \frac{1}{2} \frac{\partial D^T}{\partial X} \hat{X} \quad (16)$$

使用公式 $X^{k+1} = X^k + \hat{X}$ 进行迭代运算，可以精确定确定极值关键点的位置，达到亚像素级。具体算法如下：

Step1：设置迭代次数初始值steps，最大迭代次数max_steps，设置像素灰度阈值为Gray_Threshold；设置极值点初始值为 $(i,j,\sigma) = (1,1,1)$ ；

Step2：若迭代次数 $steps > max_steps$ ，则将该关键点从特征点集合中剔除后结束；

Step3：根据式(16)，计算 \hat{X} 和极值点 X^{k+1} 处的函数值 $D(i^k + \Delta i, j^k + \Delta j, \sigma^k + \Delta \sigma)$ ；

Step4：若 $\Delta i < 0.5, \Delta j < 0.5$ ，且 $\Delta \sigma < 0.5$ ，则 X^k 就是关键点的精确位置；否则，转Step7；

Step5：断若Gray_Threshold，则将该关键点作为特征点；否则，从特征点集合中剔除该关键点；

Step6：保存 X^k 为特征点的亚像素坐标及尺度；

Step7： $X^{k+1} = X^k + [\hat{X}]$ ，其中 $[\hat{X}]$ 表示对 \hat{X} 进行取整操作；

Step8: 迭代次数 $steps = steps + 1$; 转至 Step2。

2 特征点阈值参数的自适应设置方法

假设 Gray_Threshold 最佳设置范围为 $[G_l, G_h]$ 。Gray_Threshold 为 G_k 时, 从第 k 帧视频提取的特征点的数量为 N_k , 特征点数量的期望值为 N_t ; 从第 $(k+1)$ 帧视频提取特征点的灰度阈值为 G_{k+1} , 从第 $(k+1)$ 帧视频提取特征点的数量为 N_{k+1} 。

如果 $N_k \geq N_t$, 则增大 G_{k+1} ; 否则, 则减小 G_{k+1} , 使得从第 $(k+1)$ 帧视频中提取的特征点的数量 N_{k+1} 逼近期望值 N_t 。令:

$$\Delta G = \sqrt{\frac{N_k}{N_k + N_t}} \quad \delta = 1/\sqrt{2} \quad (17)$$

$$G_{k+1} = (1 - \delta)^{-1}(\Delta G - \delta)(G_h - G_k) + G_k \text{ if } N_k \geq N_t \quad (18)$$

$$G_{k+1} = \Delta G(G_k - G_l)/\delta + G_l \text{ if } N_k < N_t \quad (19)$$

图 6 为特征点数量期望值 $N_t = 1000$ 时, G_{k+1} 随着 N_k 不同的变化的阈值设置曲线。图中横坐标采用对数坐标表示; 不同颜色的曲线表示 G_k 取值不同时的 G_{k+1} 取值曲线。

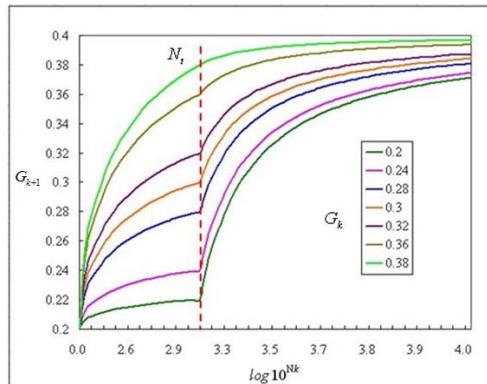


图 6 当前帧 Gray_Threshold 阈值的设置方法

Fig.6 The setting method of current frame's Gray_Threshold

3 实验结果与分析

取 20 帧航拍视频作为测试图像。实验中将特征点提取的灰度阈值 Gray_Threshold 初值设置为 0.4, 实验测试特征点的期望值 N_t 分别设置为 500, 800, 1000 时, 从视频帧提取的特征点数量 N_{k+1} 及灰度阈值 Gray_Threshold 的变化情况。图 7 为各视频帧的灰度阈值 Gray_Threshold; 图 8 为各视频帧的特征点数量。

从图 7 及图 8 中可以看出, 采用视频帧之间参数反馈的方法设置灰度阈值后, 特征点的灰度阈值能够根据特征点数量的变化自动调整, 使得特征点的数量始终保持在期望值附近。

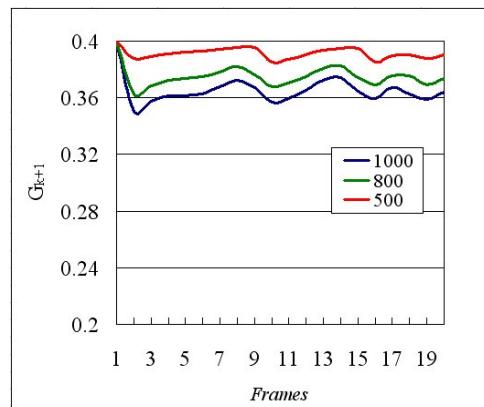


图 7 视频帧的灰度阈值

Fig.7 Gray_Threshold of video frames

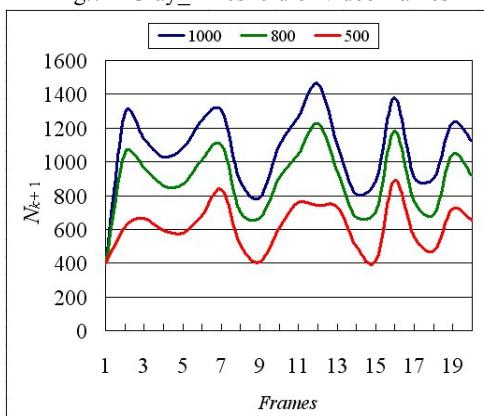


图 8 从视频帧提取的特征点数量

Fig.8 The numbers extracted from video frames

将图 7 中 $N_t = 1000$ 时的灰度阈值的平均值作为固定阈值, 测试 60 帧视频图像的特征点数量的变化, 并与本文算法进行对比。实验结果如图 9 所示。

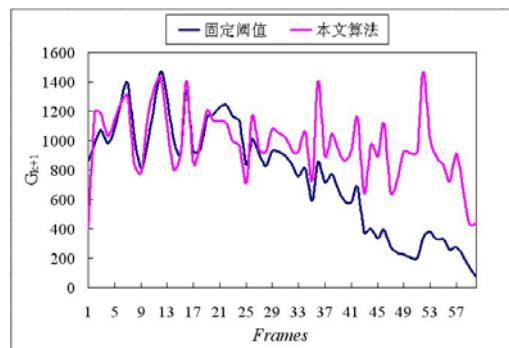


图 9 固定阈值方法与自适应算法的对比

Fig.9 Comparison of fixed threshold method with self-adaptive method

图 9 中, 由于固定阈值是采用本文算法时前 20 帧的平均值, 所以前 20 帧固定阈值与本文算法相差不大, 随着视频场景的变化, 在随后的视频中提取的特征点越来越少; 但是, 本文算法始终保持在设定的期望值附近。采用固定阈值方法和本文算法从 60 帧视频图像中提取的 SIFT 特征点数量的平均值

分别为763, 976。因此,采用自适应阈值方法时,SIFT特征点的数量更加接近于预设值。

4 结论

本文针对基于SIFT特征点的配准算法需要人工设置阈值参数的缺陷,提出了一种视频图像的自适应SIFT特征点提取算法。该方法采用视频帧之间参数反馈的方法,比较预设的特征点数量的期望值和前一帧视频的特征点数量的大小,根据比较结果,选用不同的非线性函数,自动计算当前帧的阈值参数,使得从当前帧提取的特征点数量趋近于期望值。

实际视频图像的实验表明:在自动配准过程中,该方法能够适应不同图像大小和图像内容的视频,能够根据预设的特征点数量的期望值自动设置特征点的阈值参数,使得特征点数量在图像大小和图像内容变化时,保持相对的稳定性;这样,在保证图像配准精度的前提下,大大地减小了后续的特征点匹配及优选、变换矩阵的计算及优化等数据处理计算量。该算法可以应用于自动图像配准,对于自动图像拼接具有重要的意义。

参考文献:

- [1] Lowe D G. Distinctive image features from scale-invariant key-points[J]. *International Journal of Computer Vision*, 2004, **60**: 91-110.

(上接第750页)

- [40] John Harchanko, David Chenault, Craig Farlow, et al. Detecting a surface swimmer using long wave infrared imaging polarimetry [C]//SPIE, 2005, **5780**: 138-144.
[41] John S Harchanko, Larry Pezzaniti, David Chenault, et al. Comparing a MWIR and LWIR polarimetric imager for surface swimmer detection[C]//SPIE, 2008, **6972**: 697211-1-697211-11.
[42] Barry Connor, Iain Carrie, Robert Craig, et al. Discriminative

- [2] Lowe D G. Local feature view clustering for 3D object recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition 2001, Kauai, Hawaii: 682-688.
[3] Krzysztof Slot, Hyongsuk Kim. Keypoints derivation for object class detection with SIFT algorithm[J]. *Lecture Notes in Computer Science*, 2006, **4029**: 850-859.
[4] Jo Yong-Gun, Lee Ja-Yong, Kang Hoon. Segmentation tracking and recognition based on foreground-background absolute features simplified SIFT and particle filters[C]//IEEE Congress on Evolutionary Computation, 2006, 1279-1284.
[5] Yoon-Sik Tak, Eenjun Hwang. Combining shape and sift features for 3-D object detection and pose estimation[C]//Transactions of the Korean Institute of Electrical Engineers, 2010, **59**: 429-435.
[6] Rattani A, Kisku D R, Lagorio A, et al. Facial template synthesis based on SIFT features[C]//Proceedings of IEEE Workshop on Automatic Identification Advanced Technologies, 2007, 69-73.
[7] Koenderink J J. The structure of images[J]. *Biological Cybernetics*, 1984, **50**: 363-396.
[8] Lindeberg T. Scale-space theory: A basic tool for analysing structures at different scales[J]. *Journal of Applied Statistics*, 1994, **21**: 224-270.
[9] Brown M, Lowe D G. Invariant features from interest point groups[C]//British Machine Vision Conference, 2002, Cardiff, Wales: 656-665.
[10] K. Mikolajczyk, C. Schmid. An affine invariant interest point detector[C]//European Conference on Computer Vision(ECCV), 2002, Copenhagen, Denmark: 128-142.

imaging using a LWIR polarimeter[C]//SPIE, 2008, **7113**: 71130K-1 - 71130K-11 .

- [43] Nathan J Pust, Joseph A Shaw. All-sky polarization imaging[C] //SPIE, 2007, **6682**: 668204 -1-668204-6.
[44] Nathan J Pust, Joseph A Shaw. How good is a single-scattering model of visible-NIR atmospheric skylight polarization?[C]//SPIE 2009, **7461**: 74610B-1-74610B-6.
[45] University of Arizona Polarization Lab. Polarization RayTracing with Polaris[EB/OL]. 2011. <http://www.optics.arizona.edu/chipman/>.