

# 基于 SURF 和 CamShift 的物体跟踪方法

路 宁

(中国电子科技集团公司第 15 研究所,北京 100083)

**摘 要:** CamShift 算法是一种常用的跟踪算法,但是它只利用颜色特征来进行目标跟踪,当背景颜色与跟踪目标的颜色相似时会出现跟踪失败的情况。结合 SURF 设计了一种基于 SURF 和 CamShift 算法的创新的跟踪方法。利用反馈机制计算 CamShift 跟踪窗口与初始窗口的特征相似度以及 SURF 跟踪窗口与初始窗口的色彩相似度;同时抑制大的位移变化,为两种跟踪算法动态分配位移的权重,以实现对该物体的跟踪。测试实验表明,该跟踪方法对不同的物体、在不同的环境下有较好的效果。

**关键词:** 目标跟踪;SURF;CamShift;权值分配;反馈

中图分类号: TP391.41

文献标识码: A

文章编号: 1674-7720(2012)21-0040-04

## A tracking method using SURF and CamShift

Lu Ning

(China Electronic Science and Technology Group Corporation No.15 Institute, Beijing 100083, China)

**Abstract:** CamShift tracking algorithm is a well-known tracking algorithm. Since it only adopts color as feature, which would fail to track when the object is similar to the background. A novel object tracking method based on SURF and CamShift is proposed. A feedback mechanism is used to calculate the feature similarity between the tracking windows of CamShift, so does the color similarity between the tracking windows of SURF. Large change of displacement is restrained and the weights of both results are allocated dynamically to achieve robust tracking. The experimental results show that the proposed method has good performance on different objects and environments.

**Key words:** object tracking; SURF; CamShift; weight allocated; feedback

物体跟踪是计算机视觉领域一个重要的研究方向。近年来,对于物体跟踪的研究中有两种方法:基于颜色特征的跟踪算法和基于局部特征描述子的跟踪算法。在 1997 年 COMANICIU D 和 MEER P 就提出了基于色彩的 MeanShift 算法<sup>[1]</sup>,它是利用图像的颜色信息,在反向投影图中迭代搜索,直到其收敛或达到最大迭代次数,并保存零次矩和质心位置;BRADSKI G R 等人在 MeanShift 算法的基础上提出了 CamShift (Continuously Adaptive Mean Shift) 算法<sup>[2]</sup>,并将其应用在人脸的跟踪上,取得了不错的效果。CamShift 算法利用颜色的概率信息进行物体的跟踪,运行效率比较高。CamShift 算法在简单场景下具有较好的应用效果,但在图像背景与物体本身色彩相似度比较大时,算法的跟踪效果会大大降低,甚至导致跟踪失败。SURF (Speeded Up Robust Features)<sup>[3]</sup>是对 SIFT (Scale-Invariant Feature Transform)

的一种改进特征描述,它在性能较小程度降低的情况下,极大地提升了算法的执行效率,使其能够适应对实时性要求比较高的场合。SURF 描述子具有尺度和旋转不变性的特点,在物体的匹配、查找方面有较高的准确度,近年来广泛应用在物体检测、姿态估计等方面。但 SURF 特征和 SIFT 特征一样,不包含图像的颜色信息,在纹理等特性相似的物体的干扰下,物体匹配的准确度会大幅降低。

本文基于 SURF 特征描述子和 CamShift 算法实现了一种物体跟踪方法。它利用交叉反馈的方式,通过为两种跟踪算法分配权值以实现 SURF 跟踪和 CamShift 跟踪的均衡调节,并将两种算法的结果结合在一起实现跟踪,提高了物体跟踪算法的鲁棒性。在调整反馈的权值的计算中,使用了抑制剧烈运动的方法。本文首先分别介绍了基于 SURF 的跟踪算法和基于 CamShift 的跟踪算

法,然后介绍了如何将两种算法融合在一起,以实现物体的跟踪,最后通过实验进行了测试分析,并对效果作了比较分析。

## 1 基于 SURF 的跟踪算法

### 1.1 SURF 描述子的提取

SURF 描述子的提取步骤如下:

#### (1) 构建图像 Hessian 矩阵

SURF 利用 Hessian 矩阵对图像中的像素点进行分类,根据 H 矩阵判别式取值的正负,来判定该点是或不是极值点。在 SURF 中,选用二阶标准高斯函数作为滤波器,通过特定核间的卷积计算二阶偏导数,计算出 H 矩阵的三个矩阵元素  $L_{xx}(x, \sigma)$ 、 $L_{xy}(x, \sigma)$ 、 $L_{yy}(x, \sigma)$ ,从而计算出  $H(x, \sigma)$ ,即为:

$$H(x, \sigma) = \begin{bmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{xy}(x, \sigma) & L_{yy}(x, \sigma) \end{bmatrix}$$

为了提高算法的执行效率,选用盒式滤波器代替 Laplacian-Gaussians 滤波器,在计算像素点响应的时候使用积分图像来减少运算量,以便在滤波器尺度增加的同时能够保持计算量不变。图 1 为滤波器及计算时采用的近似滤波器,其中,上面一行为 Laplacian-Gaussians 滤波器,下面一行为其相对应的近似盒式滤波器。

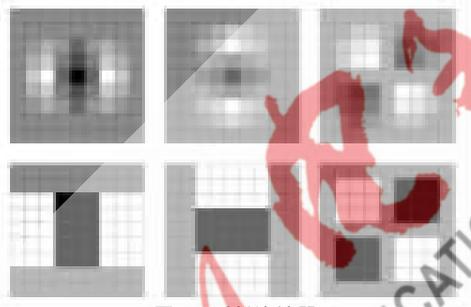


图 1 近似滤波器

为图像中每个像素计算出其 H 行列式的决定值,并用这个值来判别特征点。为方便应用,Herbert Bay 提出用近似值代替  $L$ 。同时为了平衡准确值与近似值间的误差而引入权值,权值随尺度而变化,则  $H$  矩阵判别式可表示为:

$$\det(H_{\text{approx}}) = D_{xx}D_{yy} - (0.9D_{xy})^2$$

#### (2) 构建尺度空间

图像的尺度空间是一幅图像在不同解析度下的表示,SURF 中对一幅图像  $I(X)$  在不同解析度下的表示利用高斯核  $G(\ell)$  的卷积来实现,图像的尺度大小一般用高斯标准差来表示。SURF 算法通过增加图像核的尺寸,允许尺度空间多层图像同时被处理,不需对图像进行二次抽样,从而提高算法性能。SURF 算法能够保持原始图像不变而只改变滤波器大小来构建尺度空间。

#### (3) 精确定位特征点

在计算出的极值点中,所有小于预设极值的取值都被丢弃,因此,通过增大预设极值使检测到的特征点的

数量减少,最终只有少量特征最强的点被检测出来。检测过程中使用与该尺度层图像解析度相对应的滤波器。以  $3 \times 3$  的滤波器为例,该尺度层图像中检测特征点与自身尺度层中其余 8 个点和在其之上及之下的两个尺度层各 9 个点,共 26 个点进行比较,若该像素点的特征值大于周围像素的特征值则可确定该点为该区域的特征点。

#### (4) 确定特征点主方向

为保证旋转不变性,首先以特征点为中心,计算半径为  $6s$  ( $s$  为特征点所在的尺度值)的邻域内的点在  $x$ 、 $y$  方向的 Haar 小波 (Haar 小波边长取  $4s$ ) 响应,并给这些响应值赋高斯权重系数,使得靠近特征点的响应贡献大,而远离特征点的响应贡献小。然后将  $60^\circ$  范围内的响应相加以形成新的矢量,遍历整个圆形区域,选择最长矢量的方向为该特征点的主方向。

#### (5) 生成特征点描述子

以特征点为中心,选取  $20\sigma$  的正方形区域,其中  $\sigma$  是特征点所在的尺度空间,同时,正方形窗口的方向与特征点的主方向一致,然后将区域分成  $4 \times 4$  的子域,计算每个子域大小为  $2\sigma$  的 Haar 小波,每个子域得到 1 个表征该子区域的 4 维的向量,即为:

$$v_{\text{subregion}} = [\sum dx, \sum dy, \sum |dx|, \sum |dy|]$$

因此,每个特征点能够得到一个  $4 \times 4 \times 4$  维的描述向量。

### 1.2 SURF 描述子匹配与位置估计

为实现图像中特征点的匹配,选用 FLANN (Fast Approximate Nearest Neighbor Search Library) 方法<sup>[4]</sup>,FLANN 是当今普遍使用的一种求最小临近值的方法,利用它可以实现高效快速的特征点匹配。

本方法中,采用对前后两帧之间特征点坐标值的位移量进行加权计算的方法,表达如下:

$$\Delta d = \frac{\sum_{i=1}^n \text{dis}(p_i, p_i') [p_i(x, y) - p_i'(x, y)]}{n \sum_{i=1}^n \text{dis}(p_i, p_i')}$$

其中  $\text{dis}(p_i, p_i')$  为匹配特征点对的相似距离。

## 2 CamShift 跟踪算法

CamShift 算法是一种基于目标的颜色特征,通过迭代算法进行目标定位的算法。算法过程如下:

- (1) 初始化搜索窗;
- (2) 计算搜索窗的颜色概率分布 (反向投影);
- (3) 运行 MeanShift 算法,获得新搜索窗口的大小和质心位置;
- (4) 在下一帧视频图像中用 (3) 中的值重新初始化搜索窗的大小和质心位置,再跳转到 (2) 继续进行。

### 2.1 MeanShift 算法

MeanShift 算法是一种密度函数梯度估计的非参数

方法,它通过迭代寻优找到概率分布的极值来定位目标。

MeanShift 算法过程为:

- (1) 在颜色概率分布图中选取搜索窗。
- (2) 计算零阶距:

$$M_{00} = \sum_x \sum_y I(x, y)$$

计算  $x, y$  的一阶距:

$$M_{10} = \sum_x \sum_y xI(x, y)$$

$$M_{01} = \sum_x \sum_y yI(x, y)$$

计算搜索窗的质心位置:

$$x_c = \frac{M_{10}}{M_{00}}$$

$$y_c = \frac{M_{01}}{M_{00}}$$

- (3) 根据  $M_{00}$  调整搜索窗大小。

(4) 移动搜索窗的中心到质心,如果移动的距离大于预设的固定阈值,则重复算法,直到搜索窗的中心与质心间的移动距离小于预设的固定阈值,或者当循环运算的次数达到某一最大值时,就停止计算。

## 2.2 CamShift

将 MeanShift 算法扩展到连续图像序列,就是 CamShift 算法。它将视频的所有帧做 MeanShift 运算,并将上一帧的结果,即搜索窗的大小和中心,作为下一帧 MeanShift 算法搜索窗的初始值,如此迭代下去,就可以实现对目标的跟踪。

## 3 基于交叉反馈的跟踪方法的实现

基于交叉反馈跟踪方法的设计思想是在选定跟踪目标后,分别利用 SURF 跟踪算法<sup>[5-6]</sup>和 CamShift 跟踪算法<sup>[7]</sup>来进行常规方式的目标跟踪,然后对 SURF 跟踪算法和 CamShift 跟踪算法的预测的窗口分别与初始跟踪窗口进行色彩相似度和特征相似度的计算,将计算的结果根据运动连续性的原则,抑制运动中的剧烈变化,并根据运动位移变化量进行动态权重的分配。

跟踪方法的具体流程如图 2 所示。首先标定出要跟踪的物体目标;然后算法分别进行目标窗口图像的特征描述子的提取和计算目标窗口图像的颜色概率查找表;再分别利用 SURF 跟踪算法和 CamShift 算法进行下一帧物体位置的估算,在两者估算窗口的基础上进行特征相似度和色彩相似度的计算;最后根据特征相似度和色彩相似度,计算两种算法的权值,并计算出物体的实际估计位移量,更新跟踪窗口,并将其作为下一帧物体跟踪过程的初始值,实现对连续图像帧序列中的物体的跟踪。

进行物体跟踪时,通过交叉反馈来调整物体的跟踪

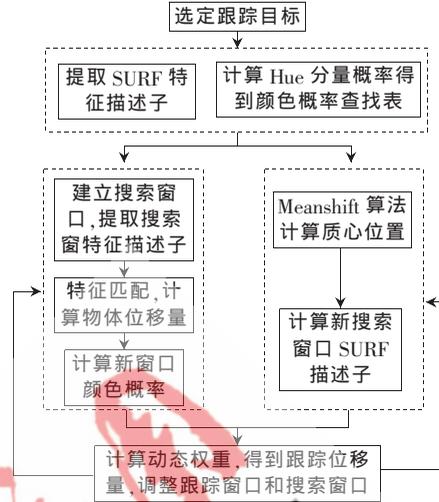


图 2 算法流程

窗口,其中主要涉及特征相似度的计算、色彩相似度的计算和动态权重的分配。

### 3.1 计算特征相似度

特征相似度表示的是 CamShift 跟踪算法计算得到的预测窗口与初始目标窗口之间的特征相似程度。在计算过程中,首先将特征点的距离(相似度)进行量化,取匹配特征点距离  $dis[P(W_{CamShift}, n), P(W_o, n)]$  与设定值  $D$  的比值  $DIS(n)$ ;然后采用取距离值的  $1/e$  的指数的方法,加大对距离小(相似度高)的特征点在相似度计算中的权重,计算式为:

$$S_f(W_{CamShift}, W_o) = \frac{\sum_{n=1}^N e^{-DIS(n)}}{N} \times 100\%$$

$$DIS(n) = \frac{dis[P(W_{CamShift}, n), P(W_o, n)]}{D}$$

用  $S_f(W_{CamShift}, W_o)$  表示预测窗口图像和初始窗口图像的特征相似度结果,  $P(W_{CamShift}, n)$ 、 $P(W_o, n)$  分别为预测窗口与初始目标窗口中相匹配的特征点,  $N$  为匹配特征点的个数。

### 3.2 计算色彩相似度

色彩相似度表示的是 SURF 跟踪算法计算得到的预测窗口与初始目标窗口之间色彩的相似程度。在计算过程中,首先计算出 SURF 跟踪算法获取的预测跟踪窗口的颜色概率,然后将其与 CamShift 算法计算出的跟踪目标的颜色概率查找表进行比较,本文用以下公式<sup>[9]</sup>来计算两副图像的色彩相似度,  $H_1$ 、 $H_2$  分别为两幅图像的色彩概率直方图。

$$S_c(H_1, H_2) = \frac{\sum_i [H_1(I) - \bar{H}_1][H_2(I) - \bar{H}_2]}{\sqrt{\sum_i [H_1(I) - \bar{H}_1]^2 \sum_i [H_2(I) - \bar{H}_2]^2}}$$

其中,  $\bar{H}_k = \frac{1}{N} \sum_j H_k(J)$ ,  $N$  为颜色直方图色彩域的总数。

### 3.3 动态权重的分配

在不同的场景和干扰下, SURF 跟踪算法和 CamShift 算法的效果并不能保持不变, 因此需要计算两种算法在各帧中的权值。本文按照物体运动的规律, 物体在运动过程中具有保持自身运动状态不变的特性, 因此可以对预测得到的较剧烈的运动状态的变化进行抑制, 为位移变化剧烈的预测算法分配相对较小的权值。由于在初始状态两种算法预测效果一致, 初始权值设为  $(0.5, 0.5)$ 。另外, 考虑到权值变化的连续性, 将前一帧的权值  $\delta_{CamShift}(i-1)$  引入到当前权值  $\delta_{CamShift}(i)$  的计算过程中。权值的计算采用以下公式:

$$\delta_{CamShift}(i) = \frac{\delta_{CamShift}(i-1)}{1 - \delta_{CamShift}(i-1)} \times \frac{(1 - \sigma(i)) S_f}{(1 - \sigma(i)) S_f + \sigma(i) S_c}$$

$$\delta_{SURF}(i) = 1 - \delta_{CamShift}(i)$$

其中,  $\sigma(i)$  是为抑制剧烈的位移变化, 而为两种相似度分配比例参量, 计算如下:

$$\sigma(i) = \frac{2\theta}{\pi}$$

$$\theta = \arctan\left(\frac{L_{SURF}(i)}{L_{CamShift}(i)}\right) \quad \theta \in (0, \frac{\pi}{2})$$

## 4 效果和分析

### 4.1 物体跟踪效果

采用的数据集 1 为道路汽车的运动视频, 数据集 2 是街道上的行人视频。对两种不同场景不同物体的数据集进行性能测试。

取实验数据中的部分有代表性的帧的跟踪结果, 如图 3 所示。



图 3 不同数据集的跟踪效果

从图 3 可以看出, 在简单场景下对于不同的物体具有较好的稳定性, 在复杂场景下, 虽然稳定性和效果上有一定的降低, 但仍能跟踪到物体而不丢失。

### 4.2 跟踪效果的比较

图 4 是 3 种方法分别第 120、130、140、150、160 帧处的跟踪效果。其中, 图 4(a) 为 SURF 跟踪算法的效果, 可以看到, 在第 140 帧处跟踪窗口开始偏离跟踪目标, 在第 160 帧处已经完全丢失跟踪目标; 图 4(b) 为 CamShift 算法的跟踪效果, 在第 120、140 帧, 算法受到路边停靠的汽车的干扰, 跟踪窗口产生较大的偏移; 图 4(c) 为本文跟踪方法的跟踪效果, 整个过程中, 能够较准确地跟踪到目标, 有更好的稳定性。

从图 4 可以看出, 单用 SURF 和 CamShift 算法, 在某些帧, 由于背景等原因, 能够准确地跟踪到物体, 但在干扰较大的时候, 会出现跟丢了的情况, 而基于 SURF 和 CamShift 的跟踪方法能够在整个跟踪过程当中不丢失, 同时又能够保证一定的跟踪精度。



图 4 跟踪效果

基于 SURF 和 CamShift 的跟踪方法利用了交叉反馈的方式, 实现了对 SURF 跟踪算法和 CamShift 算法的融合。实验表明, 本方法在准确度与鲁棒性上比单一方法的跟踪都有了一定提高, 在实际应用中能够起到较好的效果。但在外界干扰较大时, 算法的稳定性仍然有待进一步提高。在 CamShift 算法中只是取了一维的直方图, 在难以利用颜色区分背景和跟踪物体的时候, 跟踪效果难以保证, 可以尝试利用多个通道的颜色信息进行跟踪。

参考文献

- [1] COMANICIU D, MEER P. Mean Shift: a robust approach toward feature space analysis [J]. Pattern Analysis and Machine Intelligence, 2002, 24(5):603-619.
- [2] BRADSKI G R. Computer vision face tracking for use in a perceptual user interface [J]. Intel Technology Journal (2nd Quarter), 1998.
- [3] BAY H, TUYTELAARS T, VAN GOOL L. SURF: speeded up robust features [J]. Proceedings of 9th ECCV, 2006 (3591):404-417.
- [4] MUJA M, LOWE D G. Fast approximate nearest neighbors with automatic algorithm configuration [J]. International Conference on Computer Vision Theory and Applications, 2009.
- [5] TA D N, Chen Weichao, GELFAND N, et al. SURFrac: efficient tracking and continuous object recognition using local feature descriptors [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2009:2937-2944.
- [6] He Wei, YAMASHITA T, Lu Hongtao, et al. SURF tracking[C]. IEEE 12th International Conference on computer Vision, 2009:1586-1592.
- [7] ALLEN J G, RICHARD Y D, JIN J S. Object tracking using CamShift algorithm and multiple quantized feature spaces [C]. Proceedings of the Pan-Sydney Area Workshop on Visual Information Processing, 2004:3-7.

(收稿日期: 2012-07-11)

### 作者简介:

路宁, 男, 1986 年生, 硕士研究生, 主要研究方向: 计算机视觉与图像处理。