

4-D 尺度空间中基于 Mean-Shift 的目标跟踪

王宇雄^① 章毓晋^① 王晓华^②

^①(清华大学电子工程系 北京 100084)

^②(北京理工大学信息与电子学院 北京 100081)

摘要: 在基于 Mean-Shift 的目标跟踪算法中, 尺度自适应机制是算法研究的一个重要方向。一种典型的方法采用 Lindeberg 的尺度空间理论以获取目标尺度信息。但现有算法中将尺度由 2-D 矢量压缩为 1-D 量, 未能精细地刻画目标仿射变换时的尺度变化, 从而限制了算法的适用范围。为此, 该文将尺度维 1-D 滤波推广至 2-D, 构造得到了相应的 4-D 尺度空间, 并利用空间维和尺度维的 Mean-Shift 交替迭代, 实现了同时在空间位置和尺度方向对目标的有效跟踪, 提高了算法在目标尺度变化时的自适应性, 并扩大了算法的适用范围。

关键词: 目标跟踪; 均值偏移; 尺度空间理论; 尺度自适应

中图分类号: TP391

文献标识码: A

文章编号: 1009-5896(2010)07-1626-07

DOI: 10.3724/SP.J.1146.2009.01000

Mean-Shift Object tracking through 4-D Scale Space

Wang Yu-xiong^① Zhang Yu-jin^① Wang Xiao-hua^②

^①(Department of Electronic Engineering, Tsinghua University, Beijing 100084, China)

^②(School of Information and Electronic, Beijing Institute of Technology, Beijing 100081, China)

Abstract: The scale self-adaptive mechanism is one of the promising research directions in the object tracking issue based on mean-shift. A typical method is to adopt Lindeberg's scale-space theory to obtain the scale information of the target. However, the 2-D scale vector is reduced to 1-D in the existing algorithm. So it is not precise enough to portray the changing state of the object scale under affine transformation, which greatly limits the scope of application of the method. Hence, the 1-D filter in the scale dimension is extended to 2-D in this paper. Correspondingly, the 4-D scale space is constructed, then the problem is mapped into two interleaved mean-shift procedure in the spatial and scale dimension through scale space. The modified algorithm achieves the effective object tracking in spatial position and scale direction simultaneously, which enhances the self adaptability when the target scale is changing, and expands the scope of application of the algorithm.

Key words: Object tracking; Mean-Shift; Scale-Space theory; Scale adaptive

1 引言

Mean-Shift算法是一种非参数统计方法, 可经过有限次迭代快速有效地搜索到样本点分布的最近模式^[1]。在“目标模式搜索匹配”的思想之下, 引入 Mean-Shift 自适应梯度上升的特性, 可以避免穷尽搜索, 该算法被成功地应用到了对实时性要求较高的目标跟踪领域^[2]。

在基于 Mean-Shift 的目标跟踪算法中, 如何得到自适应的目标尺度估计是算法研究的一个重要方向。本质上, 尺度问题来源于核密度估计中自适应的核宽选择机制。而在跟踪问题中, 核宽直接决定了迭代计算中的样本数量和跟踪窗口的大小。

理论上对于 Mean-Shift 中核宽选择机制的分析无法应用于实时性较高的目标跟踪领域。此后产生的各种改进算法, 大体可分为3种。在文献[3]中, 对位置、尺度和方向建立联合的更高维特征空间, 通过并行迭代, 在所有维度上同时进行 Mean-Shift, 以减小某一维度对其它维度的影响。第2种方式是最常用的思路, 即采用空间维与尺度维分离迭代 (separated iteration)。首先利用 Mean-Shift 在空间维迭代, 定位新的目标中心位置; 在此基础上, 基于某种测度得到最佳尺度的估计^[4-6]。第3种方式采用 Lindeberg 的尺度空间理论以获取目标的尺度信息, 通过在权重图像上依次进行空间维卷积和尺度维卷积来建立特征空间, 将目标位置和尺度信息映射为尺度空间中的极值点, 最后通过空间维和尺度维 Mean-Shift 的交替迭代 (interleaved iteration), 找

2009-07-14 收到, 2009-11-14 改回

国家自然科学基金(60872084)资助课题

通信作者: 王宇雄 albertwyx@gmail.com

到该极值点^[7]。

本质上, 尺度如同空间坐标一样, 是一个2-D的矢量。然而, 在文献[7]中它被压缩为一个1-D量, 因此构造得到的尺度空间是3-D的, 这限制了算法仅能适应于目标在图像平面中的投影满足相似变换的情况, 而在实际场景中, 仿射变换是普遍存在的。为了扩大算法的适用范围, 本文基于文献[7]的框架结构, 进一步讨论目标跟踪中的尺度问题。经过对已有算法的分析、比较, 以及对物体运动模式的分类, 本文利用各向异性的核函数将原始的尺度维1-D滤波推广至2-D, 从而得到了改进的4-D Mean-Shift跟踪算法, 进一步提高了算法在目标尺度变化时的自适应性。

本文的后续部分如下组织。第2节为理论分析。首先对分离迭代和交替迭代进行比较, 得出采用前述第3种方式解决目标跟踪尺度问题的原因。继而, 通过对目标运动模式的分解与分类, 剖析原始3-D算法的适用范围, 并指出其不足。在此基础上, 第3节给出改进的4-D算法; 同时详细给出维度推广的基本思想、尺度空间的构建方式以及核函数的选取机制。第4节给出完整的算法实现并通过一个典型例子展示了改进算法的有效性。最后一节以一个简单的评述作结。

2 目标跟踪中尺度问题的分析

2.1 分离迭代与交替迭代的比较

文献[8]讨论了采用固定核窗宽的Mean-Shift跟踪算法在跟踪存在尺度变化的目标时的特点, 指出物体尺度增大和减小两种情况下问题具有不对称性。事实上, 当在原始核宽尺寸邻域内定义一系列的核宽参数, 即在不同尺度层次上进行评估时, 尺度问题更加复杂。

通过对目标模型与不同尺度层次上候选目标的相似性测度的实验, 可以看到在每一个尺度层次上都出现了一些新产生的空间位置并不重合的局部极大值, 分离迭代进程的贪婪性使得迭代受这些局部极值点干扰, 将阻止其收敛于希望达到的真正峰值点。相反, 基于联合空间-尺度空间(joint spatial-scale space)的交替迭代没有前者那么贪婪, 有可能避开这些噪声点达到真正的最大值。另外在某些情况下, 由于摄像机视角的变换, 目标的中心自身可能存在着漂移; 这时将导致更多局部极值点的产生, 从而使得基于分离迭代的Mean-Shift算法所获得的跟踪性能变得更差。这表明在尺度问题中, 某种意

义上类似于海森堡的“测不准原理”, 空间位置与尺度并不是相互独立的。相反, 二者紧密地耦合在一起, 相互制约。分离迭代进程人为地将这种耦合性拆分开, 但分界面的选择并不是二者耦合度最低的层次; 交替迭代则很好地反应了这一现象, 利用空间维的卷积结果继而进行尺度维的卷积, 使得每一样本点同时在平行的尺度层次中发挥了作用。同理, 文献[3]特征空间的建立中, 将尺度变量与空间坐标置于同等位置, 也在一定程度上降低了二者的耦合性与差异性。

基于上述分析, 为了避免分离迭代所导致的过分贪婪性以得到更好的跟踪效果, 应该使用交替迭代流程。当然, 可以仅仅在空间维使用均值偏移迭代; 至于尺度维, 可以基于不同的测度去获得适当的核宽估计。否则, 当跟踪目标的尺度增大时, 由于相互制约性必然导致错误的中心位置定位, 结果将使得在尺度估计中产生更多的误差^[8]。故本文采用文献[7]中映射于尺度空间Mean-Shift算法作为尺度问题的分析框架。

2.2 尺度问题的分析及原始算法的缺陷

为了更有效地处理目标跟踪中的尺度问题, 有必要将问题还原转化至原始的3-D世界坐标系中物体的运动形式分析。首先假设讨论的物体是刚性的。此处, 不失一般性, 忽略二者的比例关系, 并假设摄像机坐标系与世界坐标系是等同的, 则下文一律用世界坐标系。考虑世界坐标系XYZ中XY平面与视平面, 即投影 xy 图像平面相重合。X轴代表了高度维, Y轴代表了宽度维, 而Z轴沿着视平面的法向, 则代表了深度维。在该坐标系中, 物体的运动可以分为两大类6种。若考虑物体是沿着坐标轴运动, 还是以坐标轴为转轴进行旋转运动, 可将其运动模式分为两类。而每一类中根据坐标轴的不同又包含有3种形式。首先分析沿着坐标轴的运动。若物体沿着Z轴运动, 则其与摄像机相向运动, 称之为纵向运动。若物体沿着X轴或Y轴运动, 则其在视平面内运动, 统称为横向运动。现在分析物体以坐标轴为转轴进行的旋转运动。若转轴为Z轴, 则物体在视平面内的投影发生旋转运动, 即为通常意义上所研究的旋转运动, 此处不予以考虑。若转轴为X, Y轴, 则物体投影于视平面内, 其深度信息会转化为高度或宽度维的信息。

在3-D空间中物体的运动模式对应了其在2-D图像平面中投影的变换模式。此处, 着重考虑前者的运动对于后者变换的影响。进一步分析在3-D空间内这两类运动投影于2-D图像平面中物体的尺度结构

易知,第1类运动中投影发生相似变换,其尺度各向同性放缩;换言之,随着帧流的进行,目标的宽高比保持恒定。而第2类运动中投影发生仿射变换,增加了一个自由度,其尺度非各向同性放缩,目标的宽高比在不断发生着改变^[9]。

经过如上分析,可以得出以下结论:文献[7]的原始算法将尺度由2维向量压缩为一个1维量,仅仅当物体在2-D图像平面中的投影进行相似变换时适用,这一缺陷同样存在于文献[3]的算法之中。实际场景中问题要复杂得多,比如体育比赛,此时原始算法是无能为力的。为了在更高变换层次——仿射变换下,更精细地估计目标的尺度结构,使得算法有效,有必要将其进行推广和改进。

3 尺度空间中尺度维维数的推广

3.1 基本思想

为较为彻底地讨论尺度问题,需要回至Mean-Shift算法的原始出发点,即从一般的核函数密度估计式开始。首先,使用马氏距离(Mahalanobis distance)

$$r = (\mathbf{s} - \mathbf{x})^T \mathbf{H}^{-1} (\mathbf{s} - \mathbf{x}) \quad (1)$$

取代欧式距离(Euclid distance)得到

$$S = \{\mathbf{s} \mid (\mathbf{s} - \mathbf{x})^T \mathbf{H}^{-1} (\mathbf{s} - \mathbf{x}) < r_0^d\} \quad (2)$$

其中 \mathbf{H} 是带宽矩阵(bandwidth matrix)。上式定义了一个 d 变量空间的超椭圆球,在这种情况下, x 方向的尺度和 y 方向的尺度方能够分离。通常, \mathbf{H} 是一正定对称矩阵。

应用于多变量核密度估计的 d 变量核函数的一般形式为

$$G_H(\mathbf{x}) = |\mathbf{H}|^{-1/2} G(\mathbf{x}^T \mathbf{H}^{-1} \mathbf{x}) \quad (3)$$

其剖面函数 $G(r) : [0, \infty) \rightarrow R$ 。使用上述 d 变量核函数的加权概率密度估计为

$$q(\mathbf{x}) = \sum_{\mathbf{a} \in \Omega} G_H(\mathbf{a} - \mathbf{x}) w(\mathbf{a}) \quad (4)$$

相应的样本Mean-Shift核函数定义为

$$K_H(\mathbf{x}) = |\mathbf{H}|^{-1/2} K(\mathbf{x}^T \mathbf{H}^{-1} \mathbf{x}) \quad (5)$$

且满足 $K(r) = -cG'(r)$, $c > 0$ 。故样本均值表征为

$$m(\mathbf{x}) = \left(\sum_{\mathbf{a} \in \Omega} K((\mathbf{a} - \mathbf{x})^T \mathbf{H}^{-1} (\mathbf{a} - \mathbf{x})) w(\mathbf{a}) \right)^{-1} \cdot \sum_{\mathbf{a} \in \Omega} K((\mathbf{a} - \mathbf{x})^T \mathbf{H}^{-1} (\mathbf{a} - \mathbf{x})) w(\mathbf{a}) \mathbf{a} \quad (6)$$

对于类似Blob一样的特征,仍然使用具有各向异性的DOG核函数,即高斯滤波器(Gaussian filters,用Ga表示)的线性组合,来近似代替LOG滤波器作为微分算子。其形式为

$$\begin{aligned} \text{DOG}(x, y; \sigma_x, \sigma_y) &= \text{Ga}(x, y; \sigma_x / \sqrt{1.6}, \sigma_y / \sqrt{1.6}) \\ &\quad - \text{Ga}(x, y; \sigma_x \times \sqrt{1.6}, \sigma_y \times \sqrt{1.6}) \\ &= \frac{|\mathbf{H}|^{-1/2}}{2\pi/1.6} \exp \left[-\frac{1}{2/1.6} \begin{bmatrix} x \\ y \end{bmatrix}^T \mathbf{H}^T \begin{bmatrix} x \\ y \end{bmatrix} \right] \\ &\quad - \frac{|\mathbf{H}|^{-1/2}}{2\pi \times 1.6} \exp \left[-\frac{1}{2 \times 1.6} \begin{bmatrix} x \\ y \end{bmatrix}^T \mathbf{H}^T \begin{bmatrix} x \\ y \end{bmatrix} \right] \\ &= \frac{1}{2\pi\sigma_x\sigma_y/1.6} \exp \left[-\frac{1}{2/1.6} \left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) \right] \\ &\quad - \frac{1}{2\pi\sigma_x\sigma_y \times 1.6} \exp \left[-\frac{1}{2 \times 1.6} \left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) \right] \quad (7) \end{aligned}$$

其中 σ_x 表征了 x 方向的尺度, σ_y 表征了 y 方向的尺度,因子1.6基于与文献[7]中相同的考虑。显然,该形式是式(5)一般形式核函数的简化版本,其中采用了对角协方差矩阵作为带宽矩阵。因此,据尺度空间理论^[10],以 (x_0, y_0) 为中心,具有尺度 σ_{x0} 和 σ_{y0} 的目标物体可以被检测为图像 $f(x, y)$ 与可变尺度的DOG滤波器组卷积所生成表面的极值点。该卷积定义为

$$D(x, y; \sigma_x, \sigma_y) = \sum_{\mathbf{a} \in \Omega} \text{DOG}(a_x - x, a_y - y; \sigma_x, \sigma_y) f(a_x, a_y) \quad (8)$$

现在需要证明上述操作的尺度不变性,即

$$D(x, y; \sigma_x, \sigma_y) = D(s_x x, s_y y; s_x \sigma_x, s_y \sigma_y) \quad (9)$$

考虑一新的图像 $g(x_2, y_2) = f(x_1 / s_x, y_1 / s_y)$ 为 f 空间尺度变化的结果。由于DOG算子的特殊性,只要证明高斯滤波器的尺度不变性即可。不失一般性,假设目标位于原点,进行一变量替换 $x_2 = s_x x_1, y_2 = s_y y_1$,在 $f(x, y)$ 与高斯滤波器组卷积表面 $G(0, 0; \sigma_x, \sigma_y)$ 上,利用雅可比变换,进行类似文献[7]中的运算可以证明,经过位置平移和尺度变换后,结果依旧为局部的极大值。然而,在此种情况下,原始算法中的各向同性DOG滤波器并不满足尺度不变性的要求。

3.2 尺度空间的构建与核函数的选取

基于上述考虑,可以得到相应的核函数选择机制以及Mean-Shift操作。毫无疑问,该方式是在联合空间-尺度空间中建模,通过分离的空间位置与尺度迭代来实现的。

首先定义相应于真实目标的4-D阴影核函数 $H(x, y; \sigma_x, \sigma_y)$,其中两维为空间维的 x, y ,另两维为尺度维的 σ_x, σ_y 。在任意给定的尺度 $(\sigma_{x0}, \sigma_{y0})$ 处,2-D的边缘核函数 $H(x, y; \sigma_{x0}, \sigma_{y0})$ 为空间各向异性的滤波器 $\text{DOG}(x, y; \sigma_{x0}, \sigma_{y0})$ 。更确切地说,在当前尺度 $(\sigma_{x0}, \sigma_{y0})$ 的邻域附近,据预先约定的方式,定义一

个核函数尺度的集合, 或者说一个尺度的分布。例如^[4, 7]

$$\{\sigma_{xs} = \sigma_{x0} \times b^{s_x}, \sigma_{ys} = \sigma_{y0} \times b^{s_y}, \\ -n_x \leq s_x \leq n_x, -n_y \leq s_y \leq n_y\} \quad (10)$$

亦或

$$\{\sigma_{xs} = \sigma_{x0} \times s_x, \sigma_{ys} = \sigma_{y0} \times s_y, s_x = 1 \pm x^{n_x}, \\ s_y = 1 \pm x^{n_y}, 0 < x < 1\} \quad (11)$$

x 方向与 y 方向尺度以笛卡尔乘积的方式组合。此处采用第1种集合, 且选择参数 $b=1.1$ 与 $n_x = n_y = 2$ 。在视频跟踪中, 通常假定相邻两帧目标尺度的变化比例不超过1.1, 因此为降低计算复杂度, 亦可以选取 $n_x = n_y = 1$ 。正如尺度空间理论所指出的, 这一DOG滤波器组, 包含有 $(2n_x + 1) \times (2n_y + 1)$ 个滤波器, 将与样本权重图像 $w(x, y)$ 相卷积。然后, 将多尺度之上的卷积结果与尺度维上的2-D伊氏阴影核函数相卷积。显然, 参数 s_x 与 s_y 表征了相应方向的尺度值。故尺度核函数为

$$H_s(s_x, s_y) = 1 - \frac{s_x^2 + s_y^2}{n_x^2 + n_y^2} \quad (12)$$

结果为一个4-D的尺度空间, 紧密地将空间与尺度信息耦合在了一起。特征尺度空间为

$$E(x, y; s_x, s_y) = \sum_s H_s(s_x, s_y) \sum_{xy} H_{xy}(x, y; s_x, s_y) w(x, y) \quad (13)$$

可以从另一角度理解上述过程。滤波器组表征了目标尺度可能变化的基元。Mean-Shift迭代进程的结果将得到这些基元的组合, 从而决定物体真实的尺度。然而, 在文献[7]原始算法中这些基是不完备的, 仅由沿着同一方向尺度变化的基元组成, 甚至在高度和宽度方向的尺度变化比例系数亦相同。因此修正后的算法提高了基的完备性, 包含了沿着高度和宽度方向以不同速率增大的尺度基元, 以及沿一个方向增大而在另一个方向减小的尺度基元。

在得到详细的Mean-Shift迭代流程之前, 给出相应于尺度空间中阴影核函数的核函数。DOG算子的2-D核函数为

$$K_{xy}(x, y; s_x, s_y) = \text{Ga}(x, y; \sigma_x / \sqrt{1.6}, \sigma_y / \sqrt{1.6}) / (1.6^{-1}) \\ - \text{Ga}(x, y; \sigma_x \times \sqrt{1.6}, \sigma_y \times \sqrt{1.6}) \times (1.6^{-1}) \\ = \frac{1}{2\pi\sigma_x\sigma_y / 1.6^2} \exp\left[-\frac{1}{2/1.6}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right] \\ - \frac{1}{2\pi\sigma_x\sigma_y \times 1.6^2} \exp\left[-\frac{1}{2 \times 1.6}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right] \quad (14)$$

同时易知, 在尺度维中的伊氏阴影核函数对应于均匀核函数。

3.3 迭代流程

在上述构造的尺度空间中进行Mean-Shift, 得到了如下的交替Mean-Shift迭代流程。假设原始的参数为 $(x_0, y_0; \sigma_{x0}, \sigma_{y0})$, 则详细的迭代流程如下:

首先, 保持 σ_{x0} 和 σ_{y0} 固定, 使用下式进行一次空间维Mean-Shift操作

$$\begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \frac{\sum_s H_s(s_x, s_y) \sum_{a \in \Omega} K_{xy}(a_x - x_0, a_y - y_0; s_x, s_y) w(a_x, a_y)}{\sum_s H_s(s_x, s_y) \sum_{a \in \Omega} |K_{xy}(a_x - x_0, a_y - y_0; s_x, s_y) w(a_x, a_y)|} \cdot \begin{bmatrix} a_x - x_0 \\ a_y - y_0 \end{bmatrix} \quad (15)$$

其次, 令 (x_0, y_0) 表示上述空间维均值偏移进程的收敛点。固化 x_0 和 y_0 , 类似空间维一样, 使用下式运行一次尺度维Mean-Shift程序

$$\begin{bmatrix} \Delta s_x \\ \Delta s_y \end{bmatrix} = \frac{\sum_s \sum_{a \in \Omega} H_{xy}(a_x - x_0, a_y - y_0; s_x, s_y) w(a_x, a_y) \begin{bmatrix} s_x \\ s_y \end{bmatrix}}{\sum_s \sum_{a \in \Omega} |H_{xy}(a_x - x_0, a_y - y_0; s_x, s_y) w(a_x, a_y)| \begin{bmatrix} s_x \\ s_y \end{bmatrix}} \quad (16)$$

其中 $S_x = \{s_x \in Z \mid -n_x \leq s_x \leq n_x\}$, $S_y = \{s_y \in Z \mid -n_y \leq s_y \leq n_y\}$, 集合 $S = S_x \times S_y$, 由 S_x 与 S_y 的笛卡尔乘积组成。执行上述操作直至收敛。设置 σ_{x0} 和 σ_{y0} 的更新值为 $\sigma_{x0} \leftarrow \sigma_{x0} \times b^{\Delta s_x}$ 和 $\sigma_{y0} \leftarrow \sigma_{y0} \times b^{\Delta s_y}$ 。

4 算法总结

综合如上的理论分析, 得到了交替迭代的Mean-Shift操作流程, 以跟踪映射于4-D尺度空间中, 表征图像中目标空间位置和尺度的模式。4-D算法框架简要阐述如下:

在每一次的迭代循环中, 假定输入参数为 $(\mathbf{x}_0, \boldsymbol{\sigma}_0, \Gamma_0)$, 表征了第 t 帧的跟踪结果。此处 $\mathbf{x}_0 = (x_0, y_0)$ 为一个2维矢量, 代表了在空间维的目标中心位置坐标, $\boldsymbol{\sigma}_0 = (\sigma_{x0}, \sigma_{y0})$ 也是一个2维矢量, 代表了目标的尺度。 Γ_0 是跟踪窗口; 且 Γ 由尺度参数 $\boldsymbol{\sigma}$ 所唯一决定, 而与窗口中心相独立。

在第 $i+1$ 帧, 需要进行如下操作:

(1) 在 $(\mathbf{x}_0, \boldsymbol{\sigma}_0, \Gamma_0)$ 处产生目标候选。

(2) 在 $(\mathbf{x}_0, \boldsymbol{\sigma}_0, \Gamma_0)$ 处计算所谓的权重图像或称权值函数 $\omega(\mathbf{a})$ 。

(3) 保持 $\boldsymbol{\sigma}_0$ 固定, 使用式(15)运行一次空间维Mean-Shift操作以得到新的中心位置 $\mathbf{x}_1 = \mathbf{x}_0$

$+\Delta x$ 。

(4)保持 \mathbf{x}_1 固定,即以 \mathbf{x}_1 为中心,使用式(16)在 Γ_0 中运行一次尺度Mean-Shift操作。则得到新的尺度参数 $\Delta \mathbf{s}$,然后设置 σ_0 的新值为 $\sigma_1 = (\sigma_{x0} \times b^{\Delta s_x}, \sigma_{y0} \times b^{\Delta s_y})$,相应的跟踪窗口为 Γ_1 。

(5)上述迭代的结果为 $(\mathbf{x}_1, \sigma_1, \Gamma_1)$,重复步骤1~步骤4直至空间维和尺度维均收敛,即 $\|\Delta \mathbf{x}\| < \varepsilon_x$ 且 $\|\Delta \mathbf{s}\| < \varepsilon_s$ 。

(6)若必要的话,更新目标模型。

5 实验和分析

为了检验改进的4-D算法的有效性,进行了一系列测试。测试的平台为AMD Turion(tm) 64×2, CPU 1.81 GHz, 1.00 G内存配置的计算机。将3-D与4-D算法分别用于一汽车运动的视频序列中。该视频由252帧组成,每帧大小为640×272像素,共持续8.4084 s,速率为30 fps。这里,选取汽车整体作为跟踪的目标。

首先给出目标运动的特点。整个视频序列中,汽车的运动由首尾的两段近似直线运动和中部的转角曲线运动组成;其运动形式包含了前述的两类运动。第1类运动为:沿着Z轴负向的纵向运动与沿着Y轴正向的横向运动;而第2类运动为:以X轴为转轴的顺时针旋转运动(沿着X轴正向观察)。在这三段基本运动的叠加之下,目标的投影发生仿射变换。

跟踪从第1帧直至最后一帧。伊氏函数用于核直方图计算,目标直方图通过16×16×16bin量化RGB空间求得。迭代的终止条件为相邻两次计算得到的中心位置在一个像素点内,同时上限了迭代的最大次数不超过10次。在初始帧手动选择目标中心 $(x, y) = (180, 28)$,及窗口半径 $(h_x, h_y) = (10, 21)$ 。滤波参数的初始值为 $\sigma = 0.79$ 。图1显示了第205, 210, 215帧的跟踪结果,其中(a1)~(c1)为原始3-D算法的结果,(a2)~(c2)则为改进后4-D算法的结果。

由于前195帧中目标处于第1阶段的直线运动,旋转运动分量不甚明显,投影近似发生相似变换,故改进后的4-D算法还原为原始的3-D算法,此处不再给出跟踪的结果。但是观察图(a)至(c),目标处于第2阶段的转角曲线运动,旋转运动分量的权重逐渐增大。此时尽管在目标的高度方向3-D算法跟踪器能很好地对尺度予以定位,但是宽度方向则未能全部覆盖目标,车头部分未能被跟踪器所检测出来,这反映了算法在宽度与高度两个尺度方向上的制约性。而改进后的4-D算法的优越性就体现了出来,跟踪器在高度与宽度两个方向上均有效地覆盖了跟踪目标。

为了更好地检验改进的4-D算法的性能,还采用定量化的指标进行了比较。手工标定出各帧中汽车中心与尺度参数的真值,记为 $(x_{0i}, y_{0i}, h_{x0i}, h_{y0i})$,其中 i 为帧索引。注意到,跟踪的性能是与滤波参数 σ 的初始值密切相关的。令 σ 从0.77开始,以0.002为步长增加至0.83。在每一个特定的 σ 之下,对目标模型不进行更新,分别运行原始算法和改进后算法的相应程序,记录每帧跟踪后得到的目标中心位置与尺度值,分别记为 $(x_{3di}, y_{3di}, h_{x3di}, h_{y3di})$ 与 $(x_{4di}, y_{4di}, h_{x4di}, h_{y4di})$ 。计算得到每帧的跟踪误差,并求得整个跟踪过程中误差的均值与方差。计算公式如下:

$$\text{mean}_{3d}(t) = \frac{1}{n} \sum_{i=1}^n |t_{3di} - t_{0i}| \quad (17)$$

$$\text{mean}_{4d}(t) = \frac{1}{n} \sum_{i=1}^n |t_{4di} - t_{0i}| \quad (18)$$

$$\text{var}_{3d}(t) = \frac{1}{n} \sum_{i=1}^n (|t_{3di} - t_{0i}| - \text{mean}_{3d}(t))^2 \quad (19)$$

$$\text{var}_{4d}(t) = \frac{1}{n} \sum_{i=1}^n (|t_{4di} - t_{0i}| - \text{mean}_{4d}(t))^2 \quad (20)$$

其中 n 为计算的帧数, t 代表了参量 x, y, h_x, h_y 之一。误差均值和方差随 σ 分布的曲线如图2~图5。

各图中的虚线为原始3-D算法的结果,实线为改

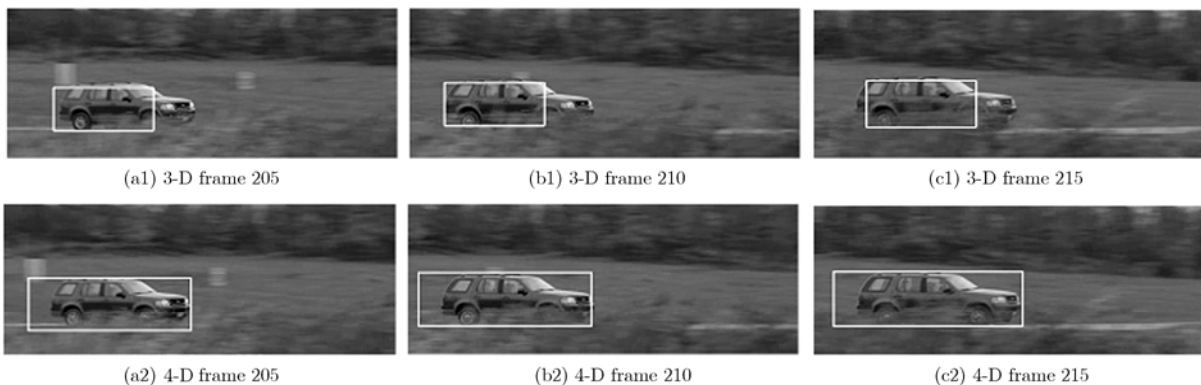


图1 汽车视频序列映射于尺度空间3-D与4-D Mean-Shift目标跟踪结果的比较

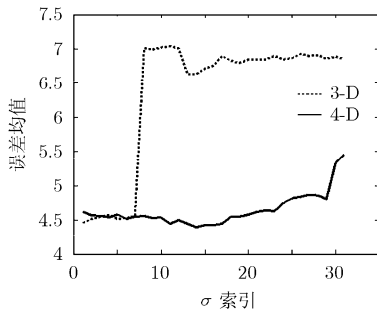


图2 x 方向中心坐标误差均值

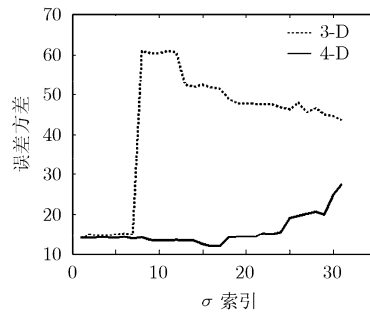


图3 x 方向中心坐标误差方差

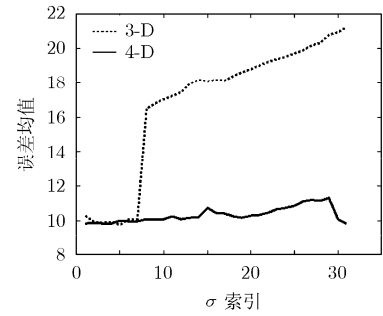


图4 x 方向尺度误差均值

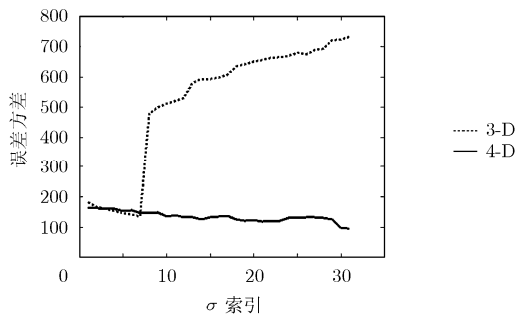


图5 x 方向尺度误差方差

进后4-D算法的结果；此处选取了 x 方向结果作为代表， y 方向结果与之类同。横向观察图2-图5，易知跟踪的性能是对参数 σ 敏感的，较好的跟踪结果需要限制 σ 在一定范围之内。纵向比较两条曲线，可见基本上虚线位于实线之上，且二者的距离较大。因此，验证了改进算法的合理性。

以上的测试实验验证了对于刚性目标，改进后的算法提高了目标相对于摄像机视角发生改变时导致其在2-D图像平面中的投影发生仿射变换时尺度跟踪定位的自适应性，将算法的适用范围由目标的第1类运动形式扩大至两类运动。

6 结论

本文在对物体运动模式的分类和目标跟踪尺度问题的分析基础之上，结合Lindeberg的尺度空间理论，应用Mean-Shift以解决目标跟踪中尺度自适应问题。通过将文献[7]原始算法中的1-D尺度维滤波推广至2-D，更精细地描述了物体的尺度结构，提高了目标尺度变化时跟踪算法的自适应性，相应地也使算法的应用范围由目标的相似变换一般化为仿射变换。

当然，将特征空间由3-D升维为4-D以提高跟踪算法的适用范围和准确性，会牺牲一定程度的快速性。这里计算耗时的增加主要是由于滤波器数目的增加。从理论上分析，文献[7]采用 $(2n + 1)$ 个滤

波器进行滤波卷积，而本文采用 $(2n_x + 1) \times (2n_y + 1)$ 个滤波器，实际中取 n 值为2，则计算量成为原来的5倍，仍为同一量级。

从物体的运动形式考虑，注意到原始3-D算法适用于前述的物体第1类运动的3种形式，改进后的4-D算法则适用于物体的两类运动的5种形式。可以推测，若引入高斯函数的相关系数，或者说将带宽矩阵由对角矩阵更一般化为正定对称矩阵，则在上述的处理框架之下，应能够同时实现对目标尺度与方位旋转变化的自适应，此时算法将对物体的两类运动6种形式有效。

参考文献

- [1] Cheng Y. Mean shift, mode seeking, and clustering [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1995, 17(8): 790-799.
- [2] Comaniciu D, Ramesh V, and Meer P. Real-time tracking of non-rigid objects using mean shift [C]. *IEEE Conference on Computer Vision and Pattern Recognition*, Hilton Head, SC, USA, 2000, II: 142-149.
- [3] Yilmaz A. Object tracking by asymmetric kernel mean shift with automatic scale and orientation selection [C]. *IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN, USA, 2007: 1-6.
- [4] Yao An-bang, Wang Gui-jin, Lin Xing-gang, and Wang Hao. Kernel based articulated object tracking with scale adaptation and model update [C]. *IEEE International Conference on Acoustics, Speech and Signal Processing*, Las Vegas, NV, USA, 2008: 945-948.
- [5] Chen Xiao-peng, Zhou You-xue, Huang Xiao-san, and Li Cheng-rong. Adaptive bandwidth mean shift object tracking [C]. *IEEE Conference on Robotics, Automation and Mechatronics*, Chengdu, China, 2008: 1011-1017.
- [6] Jiang Zhuo-lin, Li Shao-fa, Jia Xi-ping, and Zhu Hong-li. An improved mean shift tracking method based on nonparametric clustering and adaptive bandwidth [C]. *Proceedings of the Seventh International Conference on*

- Machine Learning and Cybernetics, Kunming, China, 2008: 12-15.
- [7] Collins R T. Mean-shift blob tracking through scale space [C]. IEEE Conference on Computer Vision and Pattern Recognition, Baltimore, USA, 2003, 2: 234-240.
- [8] 彭宁嵩, 杨杰, 刘志, 张凤超. Mean Shift 跟踪算法中核函数窗宽的自动选取 [J]. 软件学报, 2005, 16(9): 1542-1550.
Peng Ning-song, Yang Jie, Liu Zhi, and Zhang Feng-chao. Automatic selection of kernel-bandwidth for Mean-Shift object tracking [J]. *Journal of Software*, 2005, 16(9): 1542-1550.
- [9] 章毓晋. 图像工程. 上册, 图像处理 [M]. 第 2 版, 北京: 清华大学出版社, 2006: 62-71.
- Zhang Yu-jin. Image Engineering (I), Image Processing [M]. Second Edition, Beijing: Tsinghua University Press, 2006: 62-71.
- [10] Lindeberg T. Scale-Space Theory in Computer Vision [M]. Netherlands: Kluwer Academic Publisher, 1994, Chapter 13.
- 王宇雄: 男, 1987 年生, 硕士生, 研究方向为目标跟踪、目标识别.
- 章毓晋: 男, 1954 年生, 教授, 博士生导师, 研究方向为图像工程(图像处理、分析、理解).
- 王晓华: 女, 1971 年生, 副教授, 研究方向为图像处理、信号处理、非线性系统控制.