

多尺度时空上下文目标跟踪

李泽仁，纪 峰，常 霞，吴仰玉

(北方民族大学 数学与信息科学学院，宁夏 银川 750021)

摘要：相关滤波器在视觉目标跟踪中得到了广泛应用，针对复杂场景下目标跟踪容易出现跟踪漂移的问题，以及现有多尺度跟踪算法计算量大的问题，本文提出一种实时的多尺度目标跟踪方法。首先由时空上下文模型输出目标位置置信图完成目标定位，再在尺度空间上训练相关滤波器完成目标尺度估计，最后基于目标位置和尺度提出了一种新的时空上下文模型更新机制，避免了模型更新错误。实验表明：该方法在尺度变化、局部遮挡、目标姿态变化等情况下均能完成鲁棒跟踪，跟踪正确率较原始时空上下文跟踪算法提高了38.4%。

关键词：目标跟踪；时空上下文；多尺度

中图分类号：TP391 文献标识码：A

文章编号：1001-8891(2017)06-0535-06

Multi-scale Object Tracking Based on Spatio-temporal Context

LI Zeren, JI Feng, CHANG Xia, WU Yangyu

(School of Mathematics and Information Science, North Minzu University, Yinchuan 750021, China)

Abstract: Correlation filter has been great applied in visual tracking. Aiming at the drift problem in complex situations, a multiple scale tracking method is proposed. Firstly, spatio-temporal context model is used to output the precise location of object by the confidence map. Secondly, the scale estimation is obtained by a trained correlation filter. Finally, based on the new location and scale, a new update mechanism of the spatio-temporal context model is proposed. Experimental results show that the proposed algorithm can complete the robust tracking under the condition of scale changes, partial occlusion, pose variations, etc. Tracking precision is improved by 38.4% compared with the original spatio-temporal context tracking method.

Key words: object tracking, spatio-temporal context, scale update

0 引言

目标跟踪是当前计算机视觉领域内的一个基础性研究课题，其任务是在视频图像中的每一帧中估计出目标位置^[1]，尽管近几年目标跟踪取得了很大进展，但诸如光照变化、目标尺度变化、目标被遮挡等各种因素的干扰，使得设计一个能适应复杂场景的目标跟踪方法依然是一个非常具有挑战性的课题。根据目标在线模型，当前目标跟踪可分为生成式跟踪算法^[2-4]和判别式跟踪算法^[5-8]。

生成式跟踪算法通过特征提取完成目标模型构建，然后利用在线学习到的目标模型基于最小重构误差，在目标局部区域搜索完成目标定位。判别式跟踪

算法主要利用当前的检测技术，利用目标作为正样本、背景作为负样本训练一个分类器，通过在线更新的分类器判别输出置信度最高的样本作为目标在当前帧的位置。

近年来相关滤波器技术^[9]在目标跟踪中得到了广泛应用，基于此文献[10]在贝叶斯框架下利用时空上下文(spatio-temporal context, STC)来完成目标跟踪，在实时性和准确性上都取得了不错效果。然而，STC跟踪器采用全局模板更新目标在线模型，在目标尺度发生变化以及目标被遮挡等情况下容易发生更新错误，进而导致跟踪漂移。

针对以上问题，本文在利用时空上下文模型完成目标定位的基础上，在尺度空间上训练相关滤波器完

收稿日期：2016-08-29；修订日期：2016-09-01。

作者简介：李泽仁（1987-），男，安徽阜阳人，硕士研究生，研究方向为计算机图形图像处理。

基金项目：国家自然科学基金（61440044, 61102008）；北方民族大学科研项目（2014XYZ04）。

成目标尺度估计，提出了一种尺度自适应的目标跟踪方法。同时基于目标位置和尺度提出了一种新的时空上下文模型更新机制，减少了遮挡等外界干扰带来的模型更新错误，提高了目标跟踪的准确率。

1 时空上下文跟踪器（STC）

STC 跟踪器将跟踪问题看成是在每一帧的目标置信图 $c(x)$ 中的目标定位问题。基于贝叶斯框架， $c(x)$ 可以表示为：

$$\begin{aligned} c(x) &= P(x|o) = \sum_{c(z) \in X^c} P(x, c(z)|o) \\ &= \sum_{c(z) \in X^c} P(x, c(z)|o)P(c(z)|o) \end{aligned} \quad (1)$$

式中： o 代表场景中存在目标；假设 $\Omega_c(x^*)$ 代表当前目标的局部区域，为此上下文信息定义为 $X^c = \{c(z) = (I(z), z) | z \in \Omega_c(x^*)\}$ ， $I(z)$ 表示在位置 z 处的图像特征（比如 STC 使用的灰度值特征），条件概率 $P(x|c(z), o)$ 构建了目标与局部区域之间的空间

位置关系， $P(c(z)|o)$ 是表征局部上下文的先验概率，

其定义为：

$$P(c(z)|o) = I(z)w_\sigma(z - x^*) = I(z) \cdot a \cdot \exp\left[-\frac{(z - x^*)^2}{\sigma^2}\right] \quad (2)$$

式中： $w_\sigma(z - x^*)$ 是图像特征的权重函数； a 是归一化参数； σ 是尺度参数。

假设函数 $h^{sc}(x-z)$ 构建了目标与局部区域之间的位置关系，式(1)中条件概率 $P(x|c(z), o)$ 定义为：

$$P(x|c(z), o) = h^{sc}(x - z) \quad (3)$$

因为边缘包含背景信息的可能性较大，因此式(1)中目标置信图 $c(x)$ 定义为：

$$c(x) = b \cdot \exp\left(-\frac{|x - x^*|^\beta}{\alpha}\right) \quad (4)$$

式中： b 是归一化系数， α 、 β 分别是尺度参数和形状参数，文献[11]通过实验选取 $\beta=1$ 可以取得鲁棒的跟踪结果。综合式(2)、(3)、(4)，式(1)可以构建为：

$$\begin{aligned} c(x) &= b \cdot \exp\left(-\frac{|x - x^*|^\beta}{\alpha}\right) \\ &= \sum_{c(z) \in X^c} h^{sc}(x - z)I(z)w_\sigma(z - x^*) \\ &= h^{sc}(x) \otimes I(x)w_\sigma(x - x^*) \end{aligned} \quad (5)$$

式中： \otimes 代表卷积操作，经过一次傅里叶变换可将式(5)转换到频率域：

$$F(c(x)) = F(h^{sc}(x)) \odot F(I(x)w_\sigma(x - x^*)) \quad (6)$$

式中： \odot 代表点积操作； $F(\cdot)$ 代表快速傅里叶变换；再经过一次傅里叶反变换 $F^{-1}(\cdot)$ 可以得到空间域上的空间上下文模型 $h^{sc}(x)$ ：

$$\begin{aligned} h^{sc}(x) &= F^{-1}\left(\frac{F(c(x))}{F(I(x)w_\sigma(x - x^*))}\right) \\ &= F^{-1}\left(\frac{F(b \cdot \exp\left(-\frac{|x - x^*|^\beta}{\alpha}\right))}{F(I(x)w_\sigma(x - x^*))}\right) \end{aligned} \quad (7)$$

再考虑目标之间的时间信息，假设 H_t^{sc} 为第 t 帧的时空上下文模型，其更新方式为：

$$H_{t+1}^{sc} = (1 - \rho)H_t^{sc} + \rho h_t^{sc} \quad (8)$$

式中： ρ 为学习参数。得到第 t 帧时空上下文模型 H_t^{sc} 后，第 $t+1$ 帧目标位置置信图 $c_{t+1}(x)$ 可构建为：

$$c_{t+1}(x) = F^{-1}(F(H_{t+1}^{sc}) \odot F(I_{t+1}(x)w_{\sigma_{t+1}}(x - x_{t+1}^*))) \quad (9)$$

在目标位置置信图中寻找最大值，则最大值对应位置即为当前帧目标位置：

$$x_{t+1}^* = \arg \max_{x \in \Omega_c(x^*)} c_{t+1}(x) \quad (10)$$

从以上的分析可以发现，STC 跟踪器采用快速傅里叶变换将原本在空域中进行的特征提取等操作转换到频域中进行，大大提高了目标跟踪的实时性，但由于其采用全局信息更新目标在线模型，一旦目标出现尺度变化以及目标遮挡等情况，势必会引入更多的更新错误，进而最终引发跟踪漂移。

2 本文算法

针对以上存在的问题，本文提出如下改进策略：在 STC 模型完成目标定位的基础上，为了给出目标精确的尺度变化，在尺度空间上训练一个独立的相关滤波器。结合目标位置和尺度对时空上下文模型进行更新，避免引入模型更新错误。

2.1 相关滤波器尺度估计

现有的多尺度跟踪算法多采用目标尺度穷举搜索的方式，在获得尺度估计的同时，计算量也随着候选尺度的个数大大提高，难以满足实时应用的需求，本文基于相关滤波器的高效性，提出一种全新的快速尺度估计策略，大大提高了多尺度跟踪算法的实时性。

在 STC 模型完成目标定位后，第 $t+1$ 帧目标位置为 x_{t+1} ，为了给出目标精确的尺度变化，在尺度空间上训练一个独立的相关滤波器 $h^{scale}(x)$ 。假设 $A \times B$ 代表当前目标大小， N 代表尺度滤波器的大小，令

$n \in \left\{ \left[-\frac{N-1}{2} \right], \dots, \left[\frac{N-1}{2} \right] \right\}$, s 代表尺度因子, 则每一个基于当前目标中心位置的样本大小为 $s^n A \times s^n B$, 其图像特征可表示为 $I_n(x)$, 则尺度相关滤波器 $h^{\text{scale}}(x)$ 推导如下:

$$\begin{aligned} h^{\text{scale}}(x) &= F^{-1}\left(\frac{F(c(x))}{F(I_n(x)w_\sigma(x-x^*))}\right) \\ &= F^{-1}\left(\frac{F(b \cdot \exp(-\left|\frac{x-x^*}{\alpha}\right|^\beta))}{F(I_n(x)w_\sigma(x-x^*))}\right) \end{aligned} \quad (11)$$

同理按照式(8)得到第 $t+1$ 帧的尺度相关滤波器 $h_{t+1}^{\text{scale}}(x)$ 后, 第 $t+1$ 帧尺度置信图 $c_{t+1}^{\text{scale}}(s)$ 可构建为:

$$c_{t+1}^{\text{scale}}(s) = F^{-1}(F(h_{t+1}^{\text{scale}}) \odot F(I_{t+1}(x)w_{\sigma_t}(x-x_t^*))) \quad (12)$$

在尺度置信图中寻找最大值, 即可获得第 $t+1$ 帧的最佳尺度估计 s_{t+1}^* :

$$s_{t+1}^* = \arg \max_s c_{t+1}^{\text{scale}}(s) \quad (13)$$

至此即完成了目标定位和尺度估计, 基于此本文提出了一种自适应的时空上下文模型更新机制, 以适应目标的尺度变化等各种外观变化。

2.2 时空上下文模型自适应更新

式(2)表明: STC 跟踪器中采用固定的尺度参数 σ , 对目标外围的背景信息进行抑制, 当目标尺度发生变化时, 势必会引入错误的模型更新, 为此局部上下文的先验概率中的尺度参数应当随着目标尺度的变化而变化, 本文基于第 t 帧获得的尺度估计 s_t 来更新局部上下文的先验概率:

$$P(c(z)|o) = I(z) \cdot a \cdot \exp\left[-\frac{(z-x_t^*)^2}{(s_t \sigma)^2}\right] \quad (14)$$

式中: x_t^* 为第 t 帧的目标位置; s_t 为第 t 帧的目标尺度。

式(8)表明: STC 跟踪器采用固定的学习率对时空上下文模型进行更新, 一旦目标被遮挡或出现剧烈的姿态变化等, 容易将错误的目标信息更新到上下文模型中, 考虑到当前帧置信图中极大值反应了目标受到外界因素干扰的程度, 本文采用如下的模型更新机制:

$$H_{t+1}^{\text{sc}} = (1 - \rho T_t^{\text{conf}}) H_t^{\text{sc}} + \rho T_t^{\text{conf}} h_t^{\text{sc}} \quad (15)$$

$$T_t^{\text{conf}} = \begin{cases} 0 & \text{if } T_t^{\text{conf}} < 0 \\ \max c_t(x) & \text{else} \end{cases} \quad (16)$$

式中: T_t^{conf} 代表第 t 帧置信图中的最大值, 一旦 T_t^{conf} 低于 0, 表明跟踪结果不可靠, 目标有可能被完全遮挡或者出现剧烈的外观变化, 此时式(15)将第 t 帧空间

上下文模型 h_t^{sc} 权重置为 0, 此时模型不再更新, 避免了模型的更新错误。

2.3 算法描述

结合 2.1 的目标尺度估计和 2.2 的时空上下文模型更新, 本文的目标跟踪算法的流程图描述如图 1 所示。

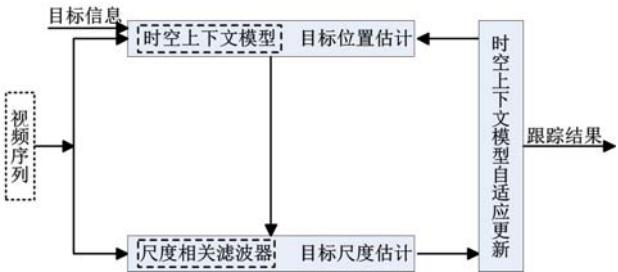


图 1 本文算法框架图

Fig.1 Algorithm framework diagram

输入图像 I_t 和目标位置 x_t^* 以及相应的尺度变化 s_t , 通过在线更新的时空上下文模型利用式(9)、(10)完成目标位置估计, 根据 2.1 节设计尺度相关滤波器, 利用式(12)、(13)完成目标尺度估计, 最后根据 2.2 节的自适应更新机制完成时空上下文模型更新, 循环执行上述步骤完成目标跟踪任务。

3 实验与结果分析

实验采用常见的 6 段具有诸如尺度变化、光照变化和姿态变化遮挡等不同挑战性的视频序列, 并与原始 STC 算法^[10]以及压缩感知跟踪算法 CT^[8]、多示例学习跟踪算法 MIL^[6]进行对比实验, 对比算法采用文献[10]的参数设置。实验视频帧数共计 4598 帧, 其不同的挑战性如表 1 所示。

3.1 评估标准

为全面地评估本文提出的多尺度时空上下文跟踪方法的有效性, 本文采用文献[11]提出的两种主流的评估标准: 正确率和成功率, 其中正确率曲线描述的是中心位置误差低于某一阈值的帧数比例, 中心位置误差定义为 $\sqrt{(x_T - x_G)^2 + (y_T - y_G)^2}$, 其中 (x_T, y_T)

表 1 测试视频的主要挑战性

Table 1 Main challenges of testing video

视频名称	帧数	主要挑战性
Car4	659	尺度变化, 光照变化
Dog1	1350	尺度变化, 姿态变化
Suv	945	局部遮挡, 全局遮挡
Faceocc2	812	局部遮挡, 姿态变化
David	761	光照变化, 姿态变化
Deer	71	运动模糊, 姿态变化

是跟踪结果的目标位置, (x_G, y_G) 是目标真实位置; 成功率曲线描述的是跟踪成功的帧数所占的比例, 定义 $s = \frac{r_T \cap r_G}{r_T \cup r_G}$ 代表每一帧跟踪结果的得分, 其中 r_T 和 r_G 分别代表跟踪结果区域和目标真实区域, 如果 s 大于一定阈值则认为跟踪成功。

3.2 结果分析

3.2.1 定量分析

图2展示了所有算法的整体性能, 红色(圆形线)代表本文算法的结果, 绿色(虚线)、黑色(实线)、和蓝色(菱形线)分别代表STC、CT和MIL的跟踪结果。本文的多尺度时空上下文跟踪方法(Proposed)在正确率曲线和成功率曲线中均排名第一。在成功率曲线中, 本文方法的AUC(area under curve)值比原STC方法提高了0.329, 在正确率曲线中, 本文方法较原STC方法提高了38.4%, 充分验证了本文的多尺度估计和模型自适应更新机制的有效性。表2给出了算法在各个视频序列中的平均中心位置误差对比, 其

中黑体代表最佳性能, 斜体代表次优性能, 可见本文方法在各类具有不同挑战性的视频中均有不错的表现, 所有测试视频的平均中心位置误差为10.3pixel, 优于其他算法。

由于本文算法采用了尺度相关滤波器进行尺度估计, 并添加了模型自适应更新机制, 算法计算量较原始STC算法有所提升。表3给出了所有评估算法的平均跟踪速度对比, 可见本文算法较STC的350FPS的平均跟踪速度有所降低, 但107FPS的跟踪速度依然可以满足实时应用的要求。

3.2.2 定性分析

图3给出了各种算法在所有测试视频上的部分跟踪结果(黑色矩形框代表本文算法的跟踪结果, 蓝色、绿色和红色分别代表STC、CT和MIL的跟踪结果)。(注: 该彩色原图详见《红外技术》网站的电子版)下面将结合各类视频的不同挑战性对算法的性能进行分析。

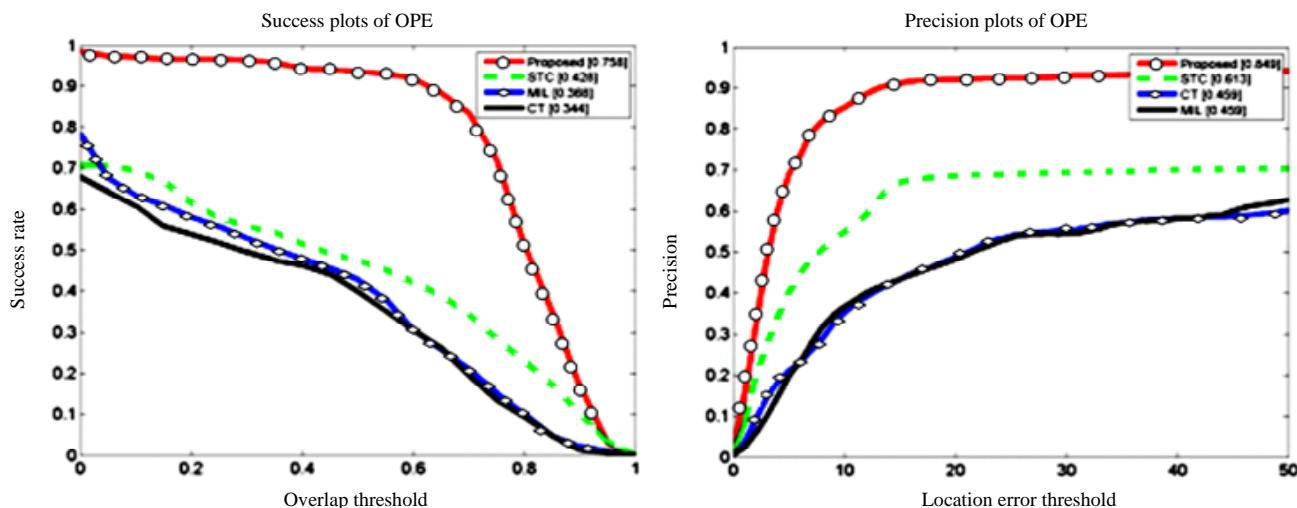


图2 整体性能的成功率曲线和正确率曲线

Fig.2 The success rate curves and the correct rate curves of the whole performance

表2 算法平均中心位置误差对比/pixel

Table 2 Mean center position error comparison/pixel

Sequence	Proposed	STC	CT	MIL
Car4	7.2	23.5	29.5	43.4
Dog1	5.7	4.6	14.8	19.1
Suv	11.2	33.1	75.4	63.6
Faceocc2	6.6	34.2	33.6	43.8
David	12.5	10.3	29.6	33.4
Deer	18.3	36.2	53.1	46.5
Average CLE	10.3	23.7	39.3	41.63

表3 算法平均跟踪速度对比

Table 3 Average tracking speeds of algorithm contrast

Methods	Proposed	STC	CT	MIL
Average FPS	107	350	80	38

图3 不同算法的部分跟踪结果

Fig.3 Partial tracking results of different algorithms

1) 尺度变化: Car4 序列和 David 序列展示了本文算法对目标尺度变化的适应性, 在 Car4 序列中随着摄像头与目标的距离不断变化, 目标尺度经历了一个由大到小再不断变大的过程, 可以发现本文算法可以很好地捕捉到目标的尺度变化, 而 STC 算法则在跟踪框变大后无法再捕捉到目标尺度变小的过程(如图 3(a)所示); 在 David 序列中, 目标在第 464 帧转身的过程中目标尺度变小, 只有本文算法捕捉到了这种变化, 跟踪框自适应地进行了调整, 充分验证了本文算法相关滤波器尺度估计的有效性。

2) 全局和局部遮挡: Suv 序列和 Faceocc2 序列展示了本文算法的抗遮挡能力, 在 Suv 序列中, 目标分别经历了 3 段不同程度的树木遮挡, 第 206 帧左右第一次遮挡后, CT 和 MIL 已经丢失目标, STC 的跟踪结果与目标中心位置距离越来越大, 而本文方法由于采用了自适应的模型更新机制, 能较好抑制背景信息, 因此在遮挡后依然能够定位到目标, 在 Faceocc2 序列中人脸受到了不同程度的遮挡, 第 730 帧左右人脸被书本和帽子遮挡后 STC 发生跟踪漂移, 只有本文方法和 CT 方法取得了较好的跟踪结果。

3) 光照变化和姿态变化: 目标的各种外观变化

会影响目标在线模型的更新, Car4 序列和 David 序列中出现了不同程度的光照变化, Dog1 序列和 David 序列中的目标经历了不同程度的姿态变化, 由于本文方法采用不同的尺度参数 σ 和自适应的学习率对目标外围的背景信息进行抑制, 减少了模型的错误更新, 因此取得了较好的跟踪效果。

4) 快速运动和运动模糊: Deer 序列中由于目标运动较快带来了不同程度的运动模糊, STC、CT 和 MIL 在第 10 帧左右即开始偏移目标, 第 18 帧左右则完全丢失目标, 只有本文方法成功完成了整个跟踪任务, 充分显示了时空上下文模型自适应地更新目标信息的高效性。

4 结论

针对现有尺度目标跟踪算法难以满足实时应用的问题, 本文提出了一种高效的尺度自适应目标跟踪算法。在时空上下文跟踪器的基础上, 基于相关滤波器的高效性, 提出了一种新的目标尺度估计方法。结合目标位置和尺度, 自适应更新时空上下文模型, 减少了模型错误更新, 较好地保留了目标信息。最后通过具有不同挑战性的测试视频进行实

验,表明:本文方法在尺度变化、局部遮挡、目标姿态变化等情况下均能完成鲁棒跟踪,跟踪正确率较原始时空上下文跟踪算法提高了38.4%。实验中我们发现当目标丢失并重新出现在场景中时本文算法的跟踪效果并不理想,下一步拟结合目标重检测机制对本文算法进行改进。

参考文献:

- [1] Emilio M. Video Tracking: Theory and Practice[M]. Queen Mary University of London, UK, 2012: 15-24
- [2] Ross D, Lim J, Lin R, et al. Incremental learning for robust visual tracking[C]//International Conference on Computer Vision, 2008: 125-141.
- [3] MEI X, LING H. Robust visual tracking and vehicle classification via sparse representation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, **33**(11): 2259-2272.
- [4] LI H, SHEN C, SHI Q. Real-time visual tracking using compressive sensing[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2011: 1305-1312.
- [5] Grabner H, Leistner C, Bischof H. Semi-supervised On-Line Boosting for Robust Tracking[C]//European Conference on Computer Vision, 2008, Part I: 234-247.
- [6] Babenko B, YANG M H, Belongie S. Robust object tracking with online multiple instance learning[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, **33**: 1619-1632.
- [7] Kalal Z, Matas J, Mikolajczyk K. Tracking-learning-detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, **34**(7): 1409-1422.
- [8] ZHANG K H, ZHANG L, YANG M H. Real-time compressive tracking[C]//European Conference on Computer Vision, 2012: 864-877.
- [9] David S Bolme, J Ross Beveridge, Bruce A Draper, et al. Visual object tracking using adaptive correlation filters[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2010: 2544-2550.
- [10] ZHANG K H, ZHANG L, LIU Q S, et al. Fast visual tracking via dense spatio-temporal context learning[C]//European Conference on Computer Vision, 2014: 127-141.
- [11] WU Y, Lim J, YANG M H. Online object tracking: a benchmark[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2013: 2411-2418.