

Research Article

A Robust Object Tracking Approach based on Mean Shift Algorithm

Zhang Xiaojing, Yajie Yue and Chenming Sha
Department of Software Engineering, School of Software,
Harbin University of Science and Technology, Harbin, China

Abstract: Object tracking has always been a hotspot in the field of computer vision, which has a range of applications in real world. The object tracking is a critical task in many vision applications. The main steps in video analysis are two: detection of interesting moving objects and tracking of such objects from frame to frame. Most of tracking algorithms use pre-defined methods to process. In this study, we introduce the Mean shift tracking algorithm, which is a kind of important no parameters estimation method, then we evaluate the tracking performance of Mean shift algorithm on different video sequences. Experimental results show that the Mean shift tracker is effective and robust tracking method.

Keywords: Computer vision, mean shift, no parameters estimation, object tracking

INTRODUCTION

Object tracking has always been a hotspot in the field of computer vision. Its application includes video monitoring, human computer interaction, vehicle tracking and so on (Xue and Haibin, 2009). Many of these applications which require reliable object tracking techniques should meet in real-time constraints. To track the object in a video sequence should be defined the dynamic entities which constantly change under few of several influence factors. The challenges in a robust tracking algorithm are caused by the presence of noise, occlusion, background clutter, varying viewpoints and illumination changes. Recently, a few of algorithms are proposed to overcome these difficulties (Yimaz *et al.*, 2006; Saravanakumar *et al.*, 2010; Meng *et al.*, 2009). In a tracking video sequence, an unknown target should be defined as anything which is interesting to analysis. For example, one people walking on a road, cars in the road, planes in the air, hand or a face in motion and so on. Recently, the form and appearance representations are classified into three families as the representation by points, representation by bounding boxes and the object silhouettes and contour (Gmez, 2009).

Recently, to select these right features, which play an important role in the tracking object in a video sequences (Collins *et al.*, 2005; Ying-Jia and Chiou-Ting, 2009). Feature selection is closely similarly to the object and target representation. We always analysis the object motion to focuses on simple characteristics such as color, texture, shape, geometry and so on. Generally, most of tracking algorithms always use a combination of these features to track an object. In general, these

features are selected manually by the user, which depend on the different fields and application domains. Yet, the problem of automatic feature selection has brought more and more attention in computer vision and pattern recognition, namely the detection of objects and then to track object in video sequences. As we all known, the object tracking is usual the first step in activities analysis system, including of interactions and relationships between objects of interest. Most of tracking algorithms have been proposed and improved. Object tracking is an estimation question or the trajectory analysis for an object when moving through a sequence of images or video. These usual tracking methods consists of block-matching, KLT, the Kalman filter (Sun *et al.*, 2010; Comaniciu *et al.*, 2003), Mean shift (Chen *et al.*, 2009; Ido *et al.*, 2010), Camshift (Ruiguo and Xinrong, 2009) and so on.

Object tracking algorithm based on color histogram for Mean shift which has obtained a wide range of applications, because of the method is simple and good real-time, it can deal with target deformation and some shelter. In these algorithms, a lot of things used statistical model for target tracking with neighbor pixel domain expression, often use a reference model through the nonlinear estimation for the parameters of the moving target model. But in actual problems, we can't get the current image related to probability density distribution of information and the lack of probability density distribution of prior knowledge, these parameters not easy to determine the limit and estimate methods in the application of target tracking. Therefore, no parameters estimation method is used in the estimation of target best position information, also is in

Corresponding Author: Zhang Xiaojing, Department of Software Engineering, School of Software, Harbin University of Science and Technology, Harbin, China

This work is licensed under a Creative Commons Attribution 4.0 International License (URL: <http://creativecommons.org/licenses/by/4.0/>).

the continuous point probability density value available near the point of observation samples probability estimation. Mean shift algorithm is a kind of important no parameters estimation method, which has been successfully applied in target tracking and related computer vision field (Chen *et al.*, 2009; Ido *et al.*, 2010).

The aim of this study is to research on object tracking algorithm and present a new object tracking algorithm. Object tracking has always been a hotspot in the field of computer vision, which has a range of applications in real word. In this study, we introduce the Mean shift tracking algorithm, which is based on color histogram and it is a kind of important no parameters estimation method, then we evaluate the tracking performance of Mean shift algorithm on different video sequences. Experimental results show that the Mean shift tracker is effective and robust tracking method.

MATERIALS

In the visual tracking, a method based on gradient matching algorithm is adopted to find the image pattern corresponding to the target image in the current frame efficiently. This kind of method is the key to obtain the similarity function in the current frame of probability density distribution once get the probability density distribution, we can accord to its gradient change direction matching search for the best path, in order to improve the efficiency of the pattern matching and meet the requirement of real-time tracking algorithm. Because we can't obtain probability density distribution information of the current frame and it is lack of the prior knowledge of the overall probability distribution. So we only use non-parameter estimation method to get the gradient of probability density. Non-parameter density estimation is based on the following thought that continuous point's probability density value can be estimated by using points of observation of the area nearby samples set.

Kernel density gradient estimation: Suppose x_1, x_2, \dots, x_N are N independent distribution of n dimension vector space. The kernel density estimate is proposed as following:

$$\hat{p}_N(x) = (Nh^n)^{-1} \sum_{i=1}^N K\left(\frac{x-x_i}{h}\right) \quad (1)$$

where $K(x)$ is a scale function, it must meet:

$$\sup_{y \in R^n} |K(y)| < \infty \quad (2)$$

$$\int_{R^n} |K(y)| dy < \infty \quad (3)$$

$$\lim_{\|y\| \rightarrow \infty} \|y\|^m K(y) = 0 \quad (4)$$

$$\int_{R^n} K(y) dy = 1 \quad (5)$$

where, $\|\cdot\|$ represents European distance. The parameter h is sample number function, it must meet:

$$\lim_{N \rightarrow \infty} h(N) = 0 \quad (6)$$

To ensure estimated progressive unbiasedness. Minimum-variance estimation continuity by type guarantee:

$$\lim_{N \rightarrow \infty} Nh^n(N) = \infty \quad (7)$$

Probability meaning of uniform continuity by the following equation:

$$\lim_{N \rightarrow \infty} Nh^{2n}(N) = \infty \quad (8)$$

We use density gradient estimates as a probability density estimation gradient:

$$\hat{\nabla}_x p_N(x) \equiv (Nh^n)^{-1} \sum_{i=1}^N \nabla_x K\left(\frac{x-x_i}{h}\right) \quad (9)$$

$$\begin{aligned} \hat{\nabla}_x p_N(x) &\equiv (Nh^n)^{-1} \sum_{i=1}^N \nabla_x K\left(\frac{x-x_i}{h}\right) \\ &= (Nh^{n+1})^{-1} \sum_{i=1}^N \nabla K\left(\frac{x-x_i}{h}\right) \end{aligned} \quad (10)$$

where,

$$\nabla K(y) \equiv \left(\frac{\partial K(y)}{\partial y^1}, \frac{\partial K(y)}{\partial y^2}, \dots, \frac{\partial K(y)}{\partial y^n}\right)^T \quad (11)$$

Mean shift theory: To give a point set $\{x_i\}_{i=1..n}$, using the kernel function $K(x)$ and kernel function bandwidth h probability density estimation of the point x is represented as following:

$$\hat{f}(x) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) \quad (12)$$

Generally, Epanechnikov kernel is a common kernel function:

$$K_E(x) = \begin{cases} \frac{1}{2} c_d^{-1} (d+2) (1-\|x\|^2), & \text{if } \|x\| < 1 \\ 0, & \text{if } \|x\| \geq 1 \end{cases} \quad (13)$$

where cd is the sphere volume. Another is often used kernel function is Gaussian kernel:

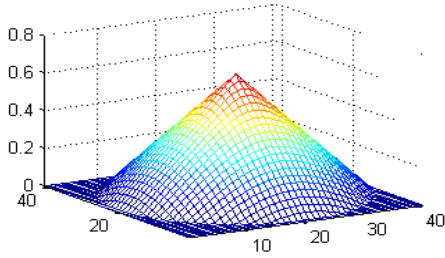


Fig. 1: The profile of epanechnikov

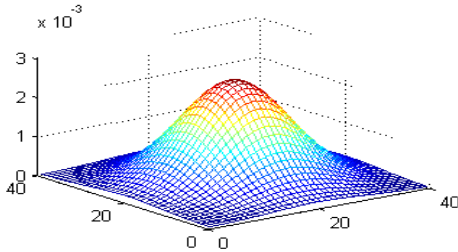


Fig. 2: The profile of Gaussian

$$K_N(x) = (2\pi)^{-d/2} \exp\left(-\frac{1}{2}\|x\|^2\right) \quad (14)$$

To define profile function of kernel function K to meet as following $k: [0, \infty) \rightarrow \mathbb{R}$. So the corresponding $k(x)$ for Epanechnikov kernel function is described as:

$$k_E(x) = \begin{cases} \frac{1}{2} c_d^{-1} (d+2)(1-x), & \text{if } x < 1 \\ 0, & \text{if } x \geq 1 \end{cases} \quad (15)$$

where Fig. 1 shows the corresponding $k(x)$ profile shape of Epanechnikov kernel function. The Eq. (3) corresponding $k(x)$ is:

$$k_N(x) = (2\pi)^{-d/2} \exp\left(-\frac{1}{2}x\right) \quad (16)$$

where, Fig. 2 shows the corresponding $k(x)$ profile shape of Gaussian kernel function.

Using the above definition, density estimation formula can rewrite as following:

$$\hat{f}(x) = \frac{1}{nh^d} \sum_{i=1}^n k\left(\left\|\frac{x-x_i}{h}\right\|^2\right) \quad (17)$$

To define:

$$g(x) = -k'(x) \quad (18)$$

where the first order derivative of $k(x)$ is existed in the interval except few points by hypothesis. Define the kernel function:

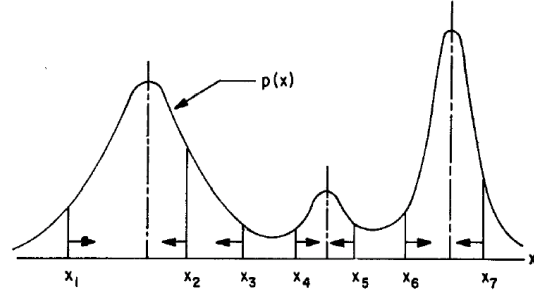


Fig. 3: Mean-shift based mode seeking

$$G(x) = Cg(\|x\|^2) \quad (19)$$

where, C is the normalized constant. Then we use the density gradient estimates said gradient density estimation:

$$\begin{aligned} \hat{\nabla}f_K(x) &\equiv \nabla\hat{f}(x) = \frac{2}{nh^{d+2}} \sum_{i=1}^n (x-x_k) k'\left(\left\|\frac{x-x_i}{h}\right\|^2\right) \\ &= \frac{2}{nh^{d+2}} \sum_{i=1}^n (x-x_k) g\left(\left\|\frac{x-x_i}{h}\right\|^2\right) \\ &= \frac{2}{nh^{d+2}} \left[g\left(\left\|\frac{x-x_i}{h}\right\|^2\right) \left[\frac{\sum_{i=1}^n x_i g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} - x \right] \right) \end{aligned} \quad (20)$$

where, we assume $\sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)$ to be no-zero. Sample mean shift vector is described as:

$$M_{h,G}(x) = \frac{\sum_{i=1}^n x_i g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} - x \quad (21)$$

The density estimate at x computed with the kernel G:

$$\hat{f}_G(x) = \frac{C}{nh^d} \sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|^2\right) \quad (22)$$

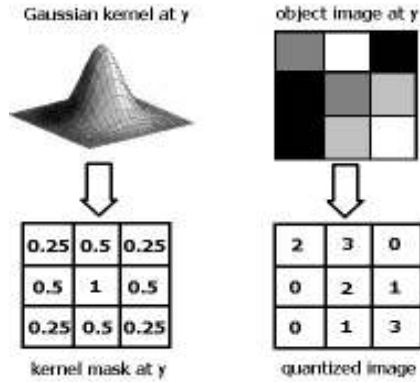
We can obtain by using the above equations:

$$\hat{\nabla}f_K(x) = \hat{f}_G(x) \frac{2/C}{h^2} M_{h,G}(x) \quad (23)$$

Then to obtain:

$$M_{h,G}(x) = \frac{h^2}{2/C} \frac{\hat{\nabla}f_K(x)}{\hat{f}_G(x)} \quad (24)$$

Figure 3 is mean-shift iterative convergence for pattern matching schematic diagram, including $p(x)$ for



(a)

$$C = 0.25 \times 4 + 0.5 \times 4 + 1 = 4$$

$$\hat{q}_0 = (0.25 + 0.5 + 0.25) / C$$

$$\hat{q}_1 = (0.5 + 0.5) / C$$

$$\hat{q}_2 = (0.25 + 1) / C$$

$$\hat{q}_3 = (0.5 + 0.25) / C$$

$$\hat{q}_0 + \hat{q}_1 + \hat{q}_2 + \hat{q}_3 = 1$$

(b)

Fig. 4: An example of kernel-histogram construction

sample density distribution function. It can be found that, in a certain kernel function window wide range Mean shift iterative exist local maxima problems, namely convergence point just local maxima gradient is zero point and is not necessarily global gradient is zero point.

METHODOLOGY

Target representation: Suppose an image $\{x_i^*\}$, $i = 1 \dots n$ consists of n points, the intensity level of each pixel is m . The center of the image coordinates is y . The kernel function histogram of target image is described as following:

$$\hat{q}_b = C \sum_{i=1}^n k\left(\left\|\frac{y-x_i^*}{h}\right\|^2\right) \delta[B(x_i^*)-b], \quad b = 1 \dots m \quad (25)$$

where,

\hat{q}_b : The target model

$B(x_i^*)$: Associates the pixel x_i to the histogram bin

$k(x)$: An isotropic kernel profile and the constant C is restrained as following:

$$C = \frac{1}{\sum_{i=1}^n k\left(\left\|\frac{y-x_i^*}{h}\right\|^2\right)} \quad (26)$$

Actually, kernel function histogram is a kind weighted representation about image color distribution. Weighted value is determined by kernel function, the

closer to the kernel center position and weights are much larger. Figure 4 indicates that the kernel function of the center of the gray-scale image histogram creation process (quantitative level 4).

Similarly, the target candidate model from the target candidate region centered at position y is given by:

$$\hat{p}_b(y) = C_h \sum_{i=1}^{nh} k\left(\left\|\frac{y-x_i}{h}\right\|^2\right) \delta[B(x_i)-b], \quad b = 1 \dots m \quad (27)$$

where,

$$C_h = \frac{1}{\sum_{i=1}^{nh} k\left(\left\|\frac{y-x_i}{h}\right\|^2\right)} \quad (28)$$

C_h is dependent on the position y . When given the kernel function and kernel function bandwidth, C_h can be computed in advance. The scale of candidate region is determined by bandwidth of kernel function and the number of pixels should be used to calculate the probability distribution.

Similarity measure: We compute the target model and the candidate target model used Eq. (25) and (27), respectively. Generally, Bhattacharyya coefficient is used to measure the similarity between the target model and the candidate model. Similarity measurement is computed as following:

$$d(y) = \sqrt{1 - \rho[\hat{p}(y), \hat{q}]} \quad (29)$$

where,

$$\hat{\rho}(y) \equiv \rho[\hat{p}(y), \hat{q}] = \sum_{b=1}^m \hat{p}_b(y) \hat{q}_b \quad (30)$$

Object tracking algorithm: To minimize the Eq. (29) is equivalent to maximize the Bhattacharyya coefficient $\hat{\rho}(y)$. The matching position in the previous frame will be the original position in the current frame. To use Taylor series expansion for Eq. (30), we can obtain:

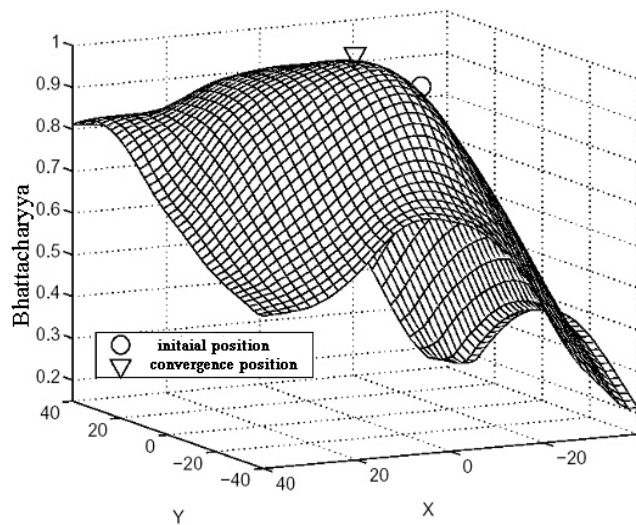
$$d[\hat{p}(y), \hat{q}] \approx \frac{1}{2} \sum_{b=1}^m \sqrt{\hat{p}_b(\hat{y}_0) \hat{q}_b} + \frac{1}{2} \sum_{b=1}^m \hat{p}_b(y) \sqrt{\frac{\hat{q}_b}{\hat{p}_b(\hat{y}_0)}} \quad (31)$$

$$d[\hat{p}(y), \hat{q}] \approx \frac{1}{2} \sum_{b=1}^m \sqrt{\hat{p}_b(\hat{y}_0) \hat{q}_b} + \frac{C_h}{2} \sum_{b=1}^{nh} w_b k\left(\left\|\frac{y-x_i}{h}\right\|^2\right) \quad (32)$$

where,



(a) Target model



(b) Plot of the bhattacharyya surface within the block region in (a)

Fig. 5: Similarity measure

$$w_i = \sum_{\mu=1}^m \sqrt{\frac{\hat{q}_b}{\hat{p}_b(\hat{y}_0)}} \delta[B(x_i) - b] \quad (33)$$

$$k(x) = \begin{cases} \frac{1}{2} c_d^{-1} (d+2)(1-x), & \text{if } x < 1 \\ 0, & \text{if } x \geq 1 \end{cases} \quad (35)$$

In the matching process, the kernel function center is moved from \hat{y}_0 in the current frame to the new position \hat{y}_1 by the following equation:

Due to the first order derivative $g(x)$ is constant, then the iteration formula can be transformed as following:

$$\hat{y}_1 = \frac{\sum_{b=1}^{nh} x_i w_i g\left(\left\|\frac{\hat{y}_0 - x_i}{h}\right\|^2\right)}{\sum_{b=1}^{nh} w_i g\left(\left\|\frac{\hat{y}_0 - x_i}{h}\right\|^2\right)} \quad (34)$$

$$\hat{y}_1 = \frac{\sum_{\mu=1}^{nh} x_i w_i}{\sum_{\mu=1}^{nh} w_i} \quad (36)$$

where,

$$g(x) = -k'(x)$$

If in the iteration process is used Epanechnikov kernel function:

Figure 5 shows the tracking result of one frame. Figure 5a is a frame image, including elliptical area said target (target model has been given in advance). Figure 5b is the inside of the rectangular area in (a) Bhattacharyya surface. From (b) graph can be obvious



Fig. 6: Results of mean shift tracking

surface maximal positions shown in triangle symbol. The position corresponding is (a) is the center of the ellipse area. Circle symbols shown in (b) position of convergence of the former a frame position and the frame of the starting position.

RESULTS

In order to evaluate the performance of Mean shift tracker, we implement many experiments on available public image sequences. The mean shift iterations were computed with the uniform profile. The target histogram has been derived in the RGB space with $16 \times 16 \times 16$ bins. Figure 6, 7 and 8 show the tracking results on different image sequences by Mean shift tracker. From Fig. 6 we can see that the mean shift based tracer proved to be robust tracking results.

From Fig. 7 we can obtain that the mean shift tracking algorithm will be robust under these conditions with varying viewpoints and illumination changes. In the David image sequence, the

illumination is varying from weak to be strong, as well as the head has a great angle rotation. So under such conditions, the tracker still has a good tracking performance and it is very robust.

To demonstrate the efficiency of the mean shift tracking algorithm, Fig. 8 presents the tracking results under the object in a fast motion condition. On the ball sequence, the ball has a greater speed and do back and forth movement up and down. The Mean shift



Fig. 7: Results of mean shift tracking on david image sequence

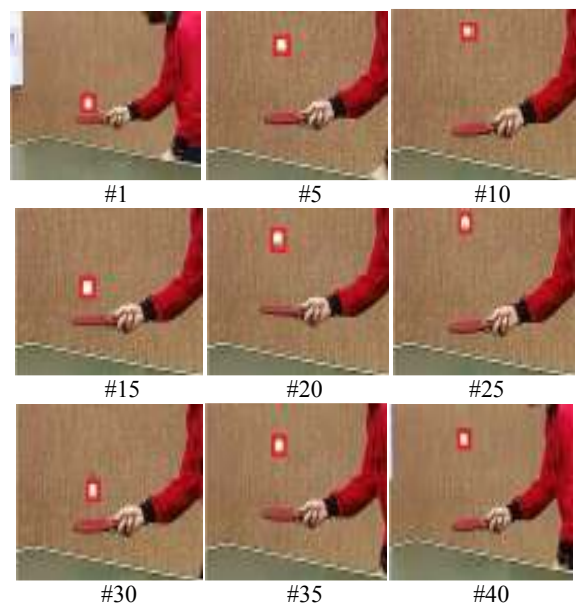


Fig. 8: Results of mean shift tracking on the ball sequence

tracking algorithm can obtain a better tracking performance and it can be robust to track the ball with a fast movement.

CONCLUSION

Object tracking has always been a hot issue in computer vision research, its application area include video surveillance, human-machine, virtual reality and so on. It requires accurate and efficient track in different background, light changes, object block and so on. Track the object for stable is a hot research issue in the field of object tracking. The Mean shift tracking algorithm basing on the Mean shift use the color histogram to track object, because of its simpleness perfect real-time performance, the ability to process the target deformation and the block situation and other different situation and also the accuracy, so the Mean shift video tracking algorithm has been widely used. Experimental results show that the Mean shift tracking algorithm is effective, robust and can be used for tracking in different scenes.

REFERENCES

- Chen, D., F. Bai, B. Li and T. Wang, 2009. BEST: A real-time tracking method for scout robot. Proceeding of the IEEE/RSJ International Conference on Intelligent Robots and Systems, pp: 1001-1006.
- Collins, R.T., L. Yanxi and M. Leordeanu, 2005. Online selection of discriminative tracking features. IEEE T. Pattern Anal., 27(10): 1631-1643.
- Comaniciu, D., V. Ramesh and P. Meer, 2003. Kernel-based object tracking. IEEE T. Pattern Anal., 25(14): 564-577.
- Gmez, N.A., 2009. A probabilistic integrated object recognition and tracking framework for video sequences. Ph.D. Thesis. Universitat Rovira I Virgili, Espagne.
- Ido, L., L. Michael and R. Ehud, 2010. Mean shift tracking with multiple reference color histograms. Comput. Vis. Image Und., 114(9): 400-408.
- Meng, G., Z. Jiang, D. Zhao and K. Ye, 2009. Real-time illumination robust maneuvering target tracking based on color invariance. Proceedings of the 2nd International Congress on Image and Signal, pp: 1-5.
- Ruiguo, Y. and Z. Xinrong, 2009. The design and implementation of face tracking in real time multimedia recording system. Proceeding of the 2nd International Congress on Image and Signal Processing, pp: 1-3.
- Saravanakumar, S., A. Vadivel and C.G. Saneem Ahmed, 2010. Multiple human object tracking using background subtraction and shadow removal techniques. Proceeding of the International Conference on Signal and Image Processing (ICSIP), pp: 15-17.
- Sun, X., H. Yao and S. Zhang, 2010. A refined particle filter method for contour tracking. Vis. Commun. Image Process., 7744(8): 1-8.
- Xue, M. and L. Haibin, 2009. Robust visual tracking using L1 minimization. Proceeding of the IEEE 12th International Conference on Computer Vision, pp: 1436-1443.
- Yimaz A., O. Javed and M. Shah, 2006. Object tracking: A survey. ACM Comput. Sur., 38(4): 13-20.
- Ying-Jia, Y. and H. Chiou-Ting, 2009. Online selection of tracking features using adaboost. IEEE T. Circuits Syst Video Tech., 19(3): 442-446.