IVA FOR ABANDONED OBJECT DETECTION: EXPLOITING DEPENDENCE ACROSS COLOR CHANNELS

Suchita Bhinge[1]*, Zois Boukouvalas*[2]*, Yuri Levin-Schwartz*[1] *and Tülay Adalı*[1]

[1]University of Maryland, Baltimore County, Dept. of CSEE, Baltimore, MD 21250
[2]University of Maryland, Baltimore County, Dept. of Mathematics and Statistics, Baltimore, MD 21250

ABSTRACT

Automated detection of abandoned object (AO) is an important application in video surveillance for security purposes. Because of its importance, a number of techniques have been proposed to automatically detect abandoned objects in the past years. However, these techniques require prior knowledge on the properties of the object such as its shape and color, in order to classify foreground objects as abandoned object. In contrast, independent component analysis (ICA) does not require such prior knowledge. However, it can only model one dataset at a time, thus limiting its usage to monochrome frames. In this paper, we propose to use independent vector analysis (IVA), a recent extension of ICA to multivariate data that takes the dependence across multiple datasets into account while retaining the independence within each dataset. We present a new framework for AO detection using IVA and show that it provides successful performance in complicated scenarios, such as for videos with crowd, illumination change, and occlusion.

Index Terms— Abandoned objects, Background subtraction, Independent vector analysis, Object detection, Video surveillance

1. INTRODUCTION

The need for automated video surveillance for detection of AOs, has dramatically increased recently due to increased security concerns, since manual surveillance is still the primary security measure for AO detection. Previously proposed techniques for AO detection implement a foreground object detection scheme and then apply a classifier to classify the foreground objects as an AO. In [1], the authors feed the shape, intensity and motion cues of static regions into a support vector machine (SVM) model to classify the static regions as true or false positives. In [2], a k-nearest neighbors (KNN) classifier identifies the foreground object as bag or non-bag, based on the shape and size of the regions. However, one drawback with background subtraction techniques is that the foreground objects that remain stationary for a sufficient period of time become a part of the background. Different approaches are proposed to avoid this issue. In [3] and [4], a dual background concept is implemented, which includes long term and short term background, while in [5], the algorithm updates a region mask to avoid losing the AO in the background image. Thus, these methods require additional post-processing for updating the background. The methods described in [2], [3], [5] and [6], use an empirical threshold based on some prior knowledge for the classification of AO. A technique based on ICA, which does not require prior knowledge of the nature of the object, was implemented to detect AOs, [7]. However, it makes use of monochrome images and does not take advantage of the dependence across color channels.

In this paper, we present a technique for detection of AO based on IVA that exploits dependence across multiple datasets. The proposed technique does not require any prior knowledge of the properties of the object and does not make use of any user defined parameters. The proposed technique provides desirable performance in complicated scenarios, such as crowd, occlusion, and illumination change. The rest of the paper is organized as follows. Section 2 describes the IVA algorithms implemented in this technique along with a brief description of the order selection scheme. Section 3 describes the detection technique implemented for AO detection and the results of the detection technique on real world videos are shown in Section 4. Section 5 concludes the paper.

2. BACKGROUND

2.1. Independent vector analysis

IVA is a generalization of ICA that achieves source separation by taking independence across latent sources, in each dataset, into account in addition to dependence across multiple datasets. The general form of IVA model is given as

$$\mathbf{x}^{[k]} = \mathbf{A}^{[k]}\mathbf{s}^{[k]}, \quad k = 1, ..., K, \tag{1}$$

where $\mathbf{A}^{[k]} \in \mathbb{R}^{N\times N}, k = 1, ..., K$ are the mixing matrices, and $\mathbf{s}^{[k]} = [s_1^{[k]}, ..., s_N^{[k]}]^\top$ are the latent sources for the kth dataset. For each dataset k, the observation matrix $\mathbf{x}^{[k]}, k = 1, ..., K$, is formed from a linear mixture of N source components in dataset k. The nth source component

vector (SCV) $\mathbf{s}_n = [s_n^{[1]}, ..., s_n^{[k]}]^\top$, is defined by concatenating the nth source from each of the K datasets. The goal in IVA is to estimate K demixing matrices in order to estimate the source components, $\mathbf{y}^{[k]}$, using $\mathbf{y}^{[k]} = \mathbf{W}^{[k]}\mathbf{x}^{[k]}$, such that each SCV is statistically independent of all other SCVs. This independence is achieved by minimizing the mutual information cost function,

$$\mathcal{I}_{\text{IVA}} = \sum_{n=1}^{N} \mathcal{H}[\mathbf{y}_n] - \sum_{k=1}^{K} \log \left| \det \left(\mathbf{W}^{[k]} \right) \right| - C, \quad (2)$$

where $\mathcal{H}[\mathbf{y}_n]$ denotes the entropy of the nth SCV and C is the constant term $\mathcal{H}[\mathbf{x}^{[1]}, ..., \mathbf{x}^{[K]}]$. The gradient of the cost function in (2) is given by

$$\frac{\partial \mathcal{I}_{\text{IVA}}}{\partial \mathbf{W}^{[k]}} = -\sum_{n=1}^{N} E\left\{ \frac{\partial \log p(\mathbf{y}_n)}{\partial y_n^{[k]}} \frac{\partial y_n^{[k]}}{\partial \mathbf{W}^{[k]}} \right\} - \left(\mathbf{W}^{[k]} \right)^{-\top}.$$

A number of algorithms have been proposed that take different types of statistical properties, such as, second order statistics (SOS) and higher order statistics (HOS), into account. IVA-Gaussian (IVA-G) [8], makes full use of SOS while IVA-Laplacian (IVA-L) [9], which assumes a Laplacian distribution as source prior, makes use of only statistics higher than two. IVA for multivariate generalized Gaussian distribution (IVA-GGD) [10] assumes a multivariate generalized Gaussian distribution (MGGD) as the source prior, which includes a wide range of unimodal distributions, such as sub-Gaussian, super-Gaussian and normal distribution and thus exploits SOS and HOS between and within datasets. IVA-GGD assumes the samples to be, independent and identically distributed (i.i.d.) and uses a fixed set of shape parameter values while estimating the scatter matrix.

2.2. Order selection

Estimating the signal subspace and performing ICA within this reduced space enables more accurate detection of the components. A number of techniques have been proposed to estimate the number of informative components, see *e.g.*, [11, 12, 13, 14]. However, most order selection methods assume that the samples are i.i.d., which is not the case for videos, [7]. Hence, we estimate the order based on the technique described in [14], which downsamples the original samples in order to get the i.i.d. samples and implements the formulation described in [14].

For this application, the number of informative components are estimated for each dataset separately. Hence, the order estimated for the kth dataset is denoted by $M^{[k]}$. In order to include more variability, the final order, $\hat{M}$, is selected to be the maximum of all the orders estimated across datasets, $\max_{k=1,...,K} \{M^{[k]}\}$.

3. IVA FOR AO DETECTION

ICA has previously been implemented on video sequences for object detection in an indoor environment, see *e.g.*, [15, 16].

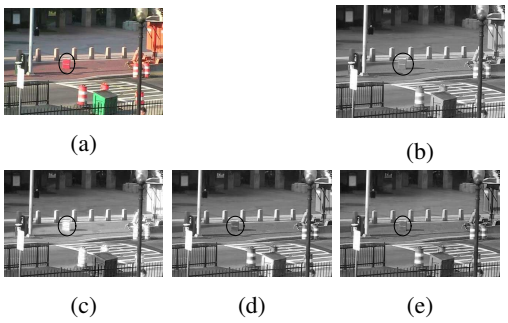

Fig. 1: Abandoned object in (a) Original color frame. (b) Gray-scale frame. (c) Red channel. (d) Green channel. (e) Blue channel.

However, the model used in these techniques require the number of frames or background and foreground frames to be specified by the user. Thus, these techniques are incapable of dealing with complicated environments. When order selection, as described in Section 2.2, is used to determine the signal subspace, ICA estimates independent components that consist of a background component and several time independent objects, since the background exhibits pixel-wise dependence across the frames, while the foreground objects exhibit an independent relationship with the background and other foreground objects. This independence is based on the pixel intensities of the background versus the foreground. Hence, if the foreground object has a similar pixel intensity as the background, the foreground object would not be extracted as an independent component. Thus, for monochrome images it is difficult to distinguish the foreground objects from the background. However, it is easy to distinguish the background from the foreground in the color space. As seen in the Figure 1, the object can be more clearly distinguished from the background in the R-channel than in the gray scaled image. Since ICA is limited to univariate data and thus, is limited to monochrome images, we assume the IVA model given in (1) that can incorporate multivariate data and make use of dependence across multiple datasets while still maintaining independence within each dataset.

3.1. IVA model for videos

In the proposed application of IVA to video processing, each of the RGB color channels is represented by an $\mathbb{R}^{N \times P}$ matrix, where N is the number of frames and P is the number of pixels. The rows of each observation matrix, $\mathbf{x}^{[k]} \in \mathbb{R}^{N \times P}, k = 1, 2, 3$, are formed by scanning the frame column-wise to form a vector of length equal to P. The dimension of the observation matrix is reduced from N to $\hat{M}$ using principal component analysis (PCA), $\hat{M} < N$, where $\hat{M}$ is estimated using the technique described in Section 2.2. The signal subspace is denoted as, $\hat{\mathbf{x}}^{[k]} \in \mathbb{R}^{\hat{M} \times P}$, which is related to the observation matrix, $\mathbf{x}^{[k]}$ by the data reduction matrix, $\mathbf{F}^{[k]} \in \mathbb{R}^{\hat{M} \times N}$ that is formed by the eigenvectors with the first $\hat{M}$ highest eigenvalues of $\mathbf{x}^{[k]}$. The signal subspace thus contains the com-

ponents that have high variance. The IVA-GGD algorithm is implemented on $\hat{\mathbf{x}}^{[k]}$, to estimate the demixing matrices, $\mathbf{W}^{[k]} \in \mathbb{R}^{\hat{M}\times\hat{M}}$. IVA-GGD makes use of fixed shape parameters that covers a wide range of unimodal distributions, such as sub-Gaussian($\beta > 1$), super-Gaussian ($\beta < 1$) and normal distribution ($\beta = 1$) The choice of parameters, β, used for this application is described in Section 3.2.

3.2. Parameter selection for IVA-GGD

IVA-GGD makes use of fixed shape parameters from a finite list of values and the choice of shape parameter, β, highly affects the identifiability of the IVA model. The (non)identifiability condition for IVA, with i.i.d. assumption, states that the IVA model cannot be identified if there are two or more α-Gaussian SCV's that satisfy the condition, $\mathbf{R}_{m,\alpha} = \mathbf{D}\mathbf{R}_{n,\alpha}\mathbf{D} \in \mathbb{R}^{K_\alpha\times K_\alpha}$, where $\mathbf{R}_{n,\alpha}$ is the covariance matrix of α-Gaussian components in the nth SCV and $\mathbf{D} \in \mathbb{R}^{K_\alpha\times K_\alpha}$ is any diagonal matrix, [17]. The α-Gaussian components refer to the group of the sources that are independent of other sources within a SCV and come from a multivariate Gaussian distribution. Here α is the index that denotes the subset of α-Gaussian components. Since, for videos, the RGB channels and the frames are highly correlated, *i.e.*, $\mathbf{R}_{m,\alpha}$ and $\mathbf{R}_{n,\alpha}$ are nearly identical, the IVA model is likely to be non-identifiable. Hence, taking this issue into account, we consider the shape parameter, β to be a positive real number not equal to 1, *i.e.*, the SCV's are chosen to be non-Gaussian.

3.3. Significance of the mixing matrix

Once the demixing matrices, $\mathbf{W}^{[k]}$, are estimated using IVA-GGD, the source components are estimated using, $\hat{\mathbf{s}}^{[k]} = \mathbf{W}^{[k]}\hat{\mathbf{x}}^{[k]}$. The mixing matrix is computed by performing back-reconstruction, *i.e.*, $\tilde{\mathbf{A}}^{[k]} = \mathbf{F}^{\dagger[k]}\hat{\mathbf{A}}^{[k]}$, where $\hat{\mathbf{A}}^{[k]} = \mathbf{W}^{[k]^{-1}}$ and $\mathbf{F}^{\dagger[k]}$ is the pseudo-inverse of the data reduction matrix $\mathbf{F}^{[k]}$.

The columns of the estimated mixing matrices represent the time courses for the source components estimated for the dataset. Thus, for a source component, $\mathbf{s}_i^{[k]}$, the ith column of the kth mixing matrix , holds a relatively larger value at the jth time point, where j denotes the frame index in which the source component, $\mathbf{s}_i^{[k]}$, appears. Hence, for a source component that represents an AO, its corresponding column in the mixing matrix would exhibit a step response, where the step increase would occur at the time point when the object is abandoned. Figure 2 shows the IVA model used for videos with the SCV that represents the AO and its corresponding time course. The next section describes the technique implemented to detect a step response in the time course, *i.e.*, an AO.

3.4. AO detection

The technique for the detection of AOs consists of, first, locating the time point at which a potential step change occurs and second, a two sample t-test on the time course that decides if the step is present or not.

In order to locate the point where the step change occurred, each column of each mixing matrix is correlated with an ideal step function and an area of interest is obtained that specifies the time points surrounding the step change. The length of the ideal step function is L time points, with $L/2$ time points before the step and $L/2$ time points after the step. The time points, at which the correlation coefficient is greater than a certain threshold, c_1, are labeled to be in the area of interest. This step eliminates most of the true negatives, *i.e.*, the time courses that do not have a step response. Next, we perform a two-sample t-test at every time point in the area of interest in order to locate the exact time when the potential step occurred.

After locating the index of the step change in the time course, a two sample t-test is performed on each point within the regions of interest, with one group containing the time points before the step change and the second group containing the time points after the step change. A higher value of the t-statistic denotes a significant difference in the intensities of the two groups, that further implies the presence of a step response, or an AO. The sign of the t-statistic obtained in this step is also used for the reconstruction of the AO component, that is described in the next section.

3.5. Reconstruction of the AO component

Once the AO component is detected in all the channels, a reconstruction step is implemented to estimate the color of the AO component. Due to the sign ambiguity inherent in IVA, the original color of the AO cannot be obtained by simple fusion of the AO source components estimated in each channel. Thus sign correction is implemented on the AO components by making use of the time course of the AO. As mentioned in Section 3.3, the time course of the AO component is likely to have a increasing step function. Using this information and assuming the object is an AO, we can correct the sign of the AO component by flipping the components that have a decreasing step function. The sign of the *t*-statistic is used to automatically detect the decreasing step function, *i.e.*, if the sign of the two-sample *t*-statistic is negative, the time course is a decreasing step function. This approach is applied to flip the component in each channel, such that the time course is an increasing step function across all the channels.

4. EXPERIMENTAL RESULTS

The proposed method is tested on the AVSS2007 dataset [18] and the CDW2014 dataset [19]. The AVSS 2007 dataset consists of a parking scenario (Easy (E), Medium (M), Hard (H)

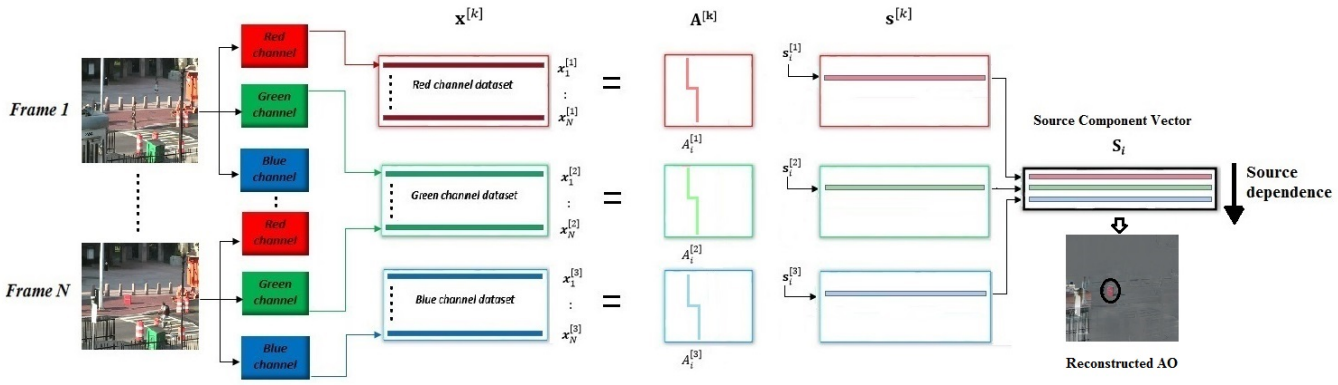

Fig. 2: IVA model for videos. Each color channel is vectorized to form a row in the respective dataset of $\mathbf{x}^{[k]}$. The SCV, $\mathbf{s}_i$, represents the source components of an AO across all channels and the corresponding column in the mixing matrix, $A_i^{[k]}$, represents the time course of the AO component, *i.e.*, a step response.

and Night (N)), each of which has a abandoned vehicle, respectively. Since the frame rate of the videos, 25fps, is quite high, the frames are subsampled by considering every fifth frame. We then perform order selection using the method described in 2.2. The steps, IVA-GGD [10], back-reconstruction as explained in Section 3.3 and detection of the step response as explained in Section 3.4, are performed five times and the best run is selected based on the t-statistic, *i.e.*, the run with a higher t-statistic. The list for the shape parameter, which are selected based on the inference described in Section 3.2, are $\beta = [0.4, 0.7, 4]$. The parameters, L and c_1 are set to 200 and 0.9, respectively. The AO component is reconstructed as described in Section 3.5. Our method fails to detect the AO only in the case when the camera is shaking (PV-Medium), however, it provides a desirable performance in the cases when the camera is still, which is the general case for video surveillance. In the PV-Medium case, since the camera is shaking, pixel values change constantly over time causing the variance related to the AO component to decrease in the dimension reduction stage, hence it is not captured as part of the signal subspace.

In order to demonstrate the improved performance due to the consideration of an additional diversity— dependence across multiple datasets— we compare IVA with ICA, since ICA is limited to univariate data. The performance of the algorithms is measured in terms of t-statistics computed on the time course of the AO component, that represents a step response. Thus, the superiority of the algorithm is based on its ability to estimate a less noisy step response, giving a higher value for the t-statistic. The ICA algorithm used for comparison is the entropy rate minimization using a MGGD model (ERM-MG) [20]. ERM-MG is referred to as ICA-GGD, when the dimension of the sources set to 1, *i.e.*, equivalent to GGD and hence making it an ICA equivalent of IVA-GGD. Table 1 demonstrates the t-statistic computed on the time course of the AO, for both ICA-GGD and IVA-GGD on the different videos.

Table 1: Comparison of ICA-GGD and IVA-GGD

Video	$\hat{M}$	ICA-GGD	IVA-GGD		
			R	G	B
Abandoned Box	22	123.13	108.49	136.30	134.44
Tramstop	35	99.95	124.63	119.55	117.91
PV-Easy	23	287.15	92.42	90.13	91.27
PV-Medium	29	-	-	-	-
PV-Hard	21	55.01	77.81	71.98	73.78
PV-Night	18	46.84	58.23	59.89	59.23

The results in Table 1 show that IVA performs better than ICA for all cases except for the video PV-Easy. This might be due to the improper estimation of the scatter matrix for the SCV representing the AO, since the SCV in this case is highly correlated. IVA-GGD implements the method of moments technique to estimate the scatter matrix of the MGGD distribution and if the condition number of the estimated scatter matrix of the SCV representing the AO is high, the inversion of the scatter matrix is inaccurate. This affects the IVA score function that would lead to the sub-optimal performance of IVA, in this case.

5. DISCUSSION

In this paper, we implemented a technique based on IVA to detect AOs and demonstrated its superior performance to ICA in complex environments, such as: crowd, occlusion and illumination changes. The performance is measured using the t-statistic computed on the time course of the AO component, since the higher value of the t-statistic allows for easier detection. Thus, using the t-statistic, we demonstrated that the performance increases for IVA since it takes an additional type of diversity- dependence across color channels- into account.

The success of the proposed method raises several interesting questions that can be explored in future work. The IVA-GGD algorithm used in this paper exploits SOS and HOS. However, the performance can be compared with algorithms that exploit different types of diversity.

6. REFERENCES

[1] J. Kim and D. Kim, "Accurate static region classification using multiple cues for aro detection," *IEEE Signal Processing Letters,*, vol. 21, no. 8, pp. 937–941, Aug 2014.

[2] M. Bhargava, C.-C. Chen, M. Ryoo, and J. Aggarwal, "Detection of abandoned objects in crowded environments," in *2007 Proc. Advanced Video and Signal Based Surveillance (AVSS).*, Sept 2007, pp. 271–276.

[3] F. Porikli, "Detection of temporarily static regions by processing video at different frame rates," in *2007 Proc. Advanced Video and Signal Based Surveillance (AVSS)*, Sept 2007, pp. 236–241.

[4] A. Singh, S. Sawan, M. Hanmandlu, V. Madasu, and B. Lovell, "An abandoned object detection system based on dual background segmentation," in *2009 Proc. Advanced Video and Signal Based Surveillance (AVSS).*, Sept 2009, pp. 352–357.

[5] N. Bird, S. Atev, N. Caramelli, R. Martin, O. Masoud, and N. Papanikolopoulos, "Real time, online detection of abandoned objects in public areas," in *2006 Proceedings Robotics and Automation (ICRA).*, May 2006, pp. 3775–3780.

[6] Y. Tian, R. Feris, H. Liu, A. Hampapur, and M.-T. Sun, "Robust detection of abandoned and removed objects in complex surveillance videos," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews.*, vol. 41, no. 5, pp. 565–576, Sept 2011.

[7] S. Bhinge, Y. Levin-Shwartz, G.-S. Fu, B. Pesquet-Popescu, and T. Adalı, "A data-driven solution for abandoned object detection: Advantages of multiple types of diversity," in *2015 Global Conference on Signal and Information Processing (GlobalSIP), (in Press)*, 2015.

[8] M. Anderson, T. Adalı, and X.-L. Li, "Joint blind source separation with multivariate gaussian model: algorithms and performance analysis," *2012 Signal Processing*, vol. 60, no. 4, pp. 1672–1683, 2012.

[9] T. Kim, T. Eltoft, and T.-W. Lee, "Independent vector analysis: An extension of ICA to multivariate components," in *Independent Component Analysis and Blind Signal Separation*. Springer, 2006, pp. 165–172.

[10] M. Anderson, G.-S. Fu, R. Phlypo, and T. Adalı, "Independent vector analysis, the kotz distribution, and performance bounds," in *2013 Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 3243–3247.

[11] M. S. Bartlett, "A note on the multiplying factors for various $\chi 2$ approximations," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 296–298, 1954.

[12] D. Lawley, "Tests of significance for the latent roots of covariance and correlation matrices," *Biometrika*, pp. 128–136, 1956.

[13] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *Acoustics, Speech and Signal Processing.*, vol. 33, no. 2, pp. 387–392, 1985.

[14] Y.-O. Li, T. Adalı, and V. D. Calhoun, "Estimating the number of independent components for functional magnetic resonance imaging data," *Human brain mapping*, vol. 28, no. 11, pp. 1251–1266, 2007.

[15] D.-M. Tsai and S.-C. Lai, "Independent component analysis-based background subtraction for indoor surveillance," *IEEE Transactions on Image Processing.*, vol. 18, no. 1, pp. 158–167, Jan 2009.

[16] X.-P. Zhang and Z. Chen, "An automated video object extraction system based on spatiotemporal independent component analysis and multiscale segmentation," *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. 184–184, 2006.

[17] T. Adalı, M. Anderson, and G.-S. Fu, "Diversity in independent component and vector analyses: Identifiability, algorithms, and applications in medical imaging," *Signal Processing Magazine, IEEE*, vol. 31, no. 3, pp. 18–33, 2014.

[18] "i-lids dataset for AVSS 2007," http://www.eecs.qmul.ac.uk/ andrea/avss2007_d.html.

[19] Y. Wang, P.-M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar, "CDnet 2014: An expanded change detection benchmark dataset," in *2014 Computer Vision and Pattern Recognition Workshops (CVPRW).*, June 2014, pp. 393–400.

[20] G.-S. Fu, R. Phlypo, M. Anderson, X.-L. Li, and T. Adalı, "Blind source separation by entropy rate minimization," *Transactions on Signal Processing*, vol. 62, no. 16, pp. 4245–4255, Aug 2014.