Multimed Tools Appl
DOI 10.1007/s11042-015-2839-3

Online sequential extreme learning machine-based co-training for dynamic moving cast shadow detection

Deepak Ghimire[1] **· Joonwhoan Lee**[2]

Received: 11 January 2015 / Revised: 11 June 2015 / Accepted: 21 July 2015

Abstract Cast shadow detection and removal is one of the key problems in vision-based systems for accurate and robust segmentation of moving objects. This paper proposes a co-training-based adaptive method for detecting moving shadows in video sequences. Shadow detection based on static methods cannot adapt to changing environment such as gradual illumination changes. In order to solve this problem, we have proposed an online sequential extreme learning machine (OS-ELM)-based semi-supervised technique for moving cast shadow detection. Online learning of OS-ELM is much faster and provides better generalization performance compared to other popular online learning algorithms. First, we extracted color, texture, gradient, and image patch similarity features using a background model and input video frame, which are useful for discriminating moving shadows and objects. Co-training scheme is used for online updating of the OS-ELM classifier in order to adapt to the dynamic environment. Experimental results on different benchmark video sequences shows that the proposed method performs better shadow detection and discrimination compared with other methods.

Keywords Foreground segmentation · Cast shadow detection · Co-training · Online sequential extreme learning machine

1 Introduction

Moving object detection is one of the most important topics in computer vision, with many applications such as surveillance systems, vehicle tracking, pedestrian tracking,

✉ Joonwhoan Lee
chlee@jbnu.ac.kr

Deepak Ghimire
deepak@keti.re.kr

1 Korea Electronics Technology Institute, Jeonju-si, Jeollabuk-do 561-844, Republic of Korea

2 Division of Computer Engineering, Chonbuk National University, Jeonju-si, Jeollabuk-do 561-756, Republic of Korea

Published online: 01 August 2015

and video conferencing. Background subtraction is the most common approach to detect foreground objects in a video sequence. However, background subtraction has many problems associated with shadows, sudden illumination changes, etc. Among them, the separation of cast shadow pixels from object pixels is the most challenging problem for robust moving object detection. For an indoor scene, there could be multiple light sources and reflective surfaces, whereas for an outdoor scene, shadow orientation, size, and intensity usually vary with cloud conditions, the time of day, and camera position. These inherent issues make shadow detection a challenging task. This paper addresses an interesting and challenging problem for robust moving object detection. The key idea of this work is to focus on detecting moving shadow and ghost pixels across changing scene conditions with minimal human intervention. This paper solves the problem of cast shadow removal with gradual illumination change mainly in outdoor scenes. Due to its dynamic nature to adapt in changing environment it works better as compared to shadow detection methods which work in static settings.

In semi-supervised learning-based shadow detection in a video sequence, online training time is the main concern. The shadow detection, the collection of new training data, and updating the classification model according to the newly collected training data should be performed in real time. As such, a fast sequential learning scheme is needed. In this paper, we use a co-training scheme with OS-ELM for the detection of moving cast shadows in a dynamic environment. OS-ELM proposed by Liang et al. [18] is a fast and accurate online sequential learning algorithm for single hidden layer feedforward networks (SLFNs) with additive and radial basis function (RBF) hidden nodes. The development of OS-ELM was based on extreme learning machine (ELM) [9], which is used for batch learning and has been shown to be extremely fast with good generalization performance. OS-ELM can learn data one by one or chunk by chunk with fixed or varying chunk size. It has been shown that OS-ELM is faster than other sequential algorithms and produces better generalization performance on many benchmark problems [18]. In the co-training loop the classifier can be updated frequently, even in every frame. This will have advantage in fast adaptability of the algorithm for changing environment. Therefore combination of OS-ELM and co-training scheme for adaptive shadow detection problem is a better choice.

The main contributions of this paper are summarized as follows:

1. We use a slightly different co-training scheme as compared to the literature, especially in the case of new training data collection. Mostly, new data in the co-training loop is collected based on the highest confidence level on previously trained classifiers. But, here we include a post-processing step inside the co-training loop, and the new training data for both shadow and object pixels are collected randomly from the shadow and object regions, respectively. This scheme helps by preventing the collection of misclassified training samples. As long as a chunk of the new training sample is collected, the OS-ELMs are learned for those new samples to adapt the classifier for a changing environment.
2. Since OS-ELM is used as a classifier, the online learning is extremely fast. It is even possible to update the OS-ELM in co-training loop for every frame in the video sequence.

3. An extended feature set extracted from both grayscale and color images are used. The color, image gradient, image patch similarity-based features are used; which helps to significantly improve the shadow detection and discrimination accuracy.

The rest of the paper is organized as follows. Brief description of the related work in the field of shadow detection is presented in Section 2. In Section 3, we describe the method to extract useful features for moving shadow detection. Section 4 presents the proposed technique for shadow detection and discrimination which includes brief review of OS-ELM, proposed co-training scheme and post-processing of the shadow detection result. Section 5 describes the experimental results including quantitative, qualitative, and comparative evaluations. Finally, a summary of the work is presented in Section 6.

2 Related work

An early review of shadow detection and removal techniques was done in 2003 by Prati et al. [20]. They categorized shadow detection methods in an algorithm-based taxonomy. They also presented qualitative and quantitative comparisons of most of the shadow detection methods published before 2003. A recent survey and comparative evaluations of shadow detection and removal methods was presented by Sanin et al. [23]. They categorized recently published shadow detection methods into four feature-based categories: chromaticity-based methods, physical methods, geometry-based methods, and texture-based methods. Chromaticity-based methods use color information in different color spaces for distinguishing shadows and object pixels [2, 4]. The main assumption is that the region under the shadow becomes darker but retains its chromaticity. These methods are sensitive to strong illumination changes and fail with strong cast shadows. Physical methods make assumptions about the physical properties of the shadow in order to model the shadow pixels [8]. Usually, they fail when the spectral properties of the object are similar to the background. Geometry-based methods use knowledge of the illumination source, object shape, object size, and ground plane to detect the orientation, size, and location of the shadow region [3, 7]. But those assumptions are very strict and do not work in most cases, such as with indoor scenes in which there could be multiple light sources, resulting in imprecise shadow orientation. Texture-based methods classify candidate pixels or regions into shadows or object pixels based on the texture correlation [16, 22]. If the texture in the input frame and background frame is similar, the region is classified as a shadow; otherwise, it is classified as an object. Texture-based methods are not dependent on colors, and are robust to illumination changes.

There has been significant work that deals with the problem of shadow detection. Most of the proposed techniques for shadow detection use features such as color, texture, and gradient image features. Recently, Li et al. [17] presented an adaptive moving cast shadow detection method based on the cast shadow model. This method combines ratio edge and ratio brightness between a shadow region and the background. Zhigang et al. [29] combined texture, chromaticity, brightness, and spatial-temporal context for detecting object and shadow pixels in different environments. Moving cast shadows were used for the detection of vehicles using the HSI and c1c2c3 color model [21]. Firstly, the ratio of hue over intensity in the HSI color model is used to detect the bright object pixels, and then the color invariant in the c1c2c3 color model is used to detect the dark and colorful object pixels. Information of color, shading, texture, neighborhood and temporal consistency are used to detect shadows efficiently and adaptively in [26]. Jiang et al. [13] uses three components in YUV color space to identify shadow pixels from the candidate foreground. An adaptive threshold estimator is designed using edges to improve shadow detection accuracy and adaptive capacity in various lighting conditions. Shadow suppression scheme based on directional coefficient of Daubechies complex wavelet transform is presented in [10] in combination with background subtraction. Normalized cross correlation (NCC) between an input frame and a background model is also widely used to detect moving shadows [6, 27, 28]. Edge-based algorithms were also developed for removing shadows and ghost pixels [14].

Most of these methods require significant human input, and they work in static settings with known scene conditions. Recently, Joshi et al. [15] presented a semi-supervised learning-based shadow detection method for dynamic environments. Color and edge features are extracted based on the input frame and background model, which are useful for discriminating shadows and object pixels. Then, a co-training scheme with support vector machines (SVM) is used to learn the shadow and object pixels. Extensive experiments were performed to show that the co-training scheme works much better and adapts according to changing environments. The strengths of this method are the small quantity of human-labeled data required and the ability to adapt automatically to changing scene conditions via co-training. The problem associated with the method is the speed of online learning. When new data is presented, the SVMs should be updated online to adapt to new training data. As time passes the size of the training data will be increased, which will require more time for retraining. Another semi-supervised learning-based method for adaptive shadow detection uses chromaticity, saturation, texture, and edge features for the ensemble-driven self-training of multiple SVMs classifiers [5]. Jarraya et al. [11] proposed a variant of the co-training technique for shadow detection and removal in uncontrolled scenes. They focused on reducing training time by periodically running the co-training process according to a temporal framework.

3 Extracting features

The features extracted for our system also utilize information from background frame of the video sequence. The reason for including background model to extract features is that there are some properties between background model and shadow in the corresponding position such as texture in shadow region is similar to the one in the background. The other features are extracted based on single pixel as well as shadow regions. The background model is first obtained using a Gaussian mixture model (GMM) approach [25]. The background model separates the moving foreground objects from stationary backgrounds. The moving cast shadow is also included with foreground region pixels. The color, edge, and image patch similarity features are extracted for each pixel in the foreground mask obtained using the GMM model. We give a brief description of all the features used in our system.

1. *RGB Component Ratios*: It is assumed that the brightness of the shadow pixel is less than that of the same pixel in the background. When a shadow covers the background, the brightness decline proportion of each color component is independent, and the brightness ratio between a shadow and the background is constant [24]. This feature is also used for the cast shadow detection of vehicles in traffic scenes [12, 17].
2. *Blue and Red Ratio*: A point covered by a shadow will change the RGB components as the blue component increases and the red component decreases compared with its

appearance [12]. Therefore, the B/R ratio of each pixel in a shadow area is different from those of a background or object.

3. *Normalized RGB Component*: Normalized RGB values are widely used in many applications, including moving cast shadow detection [2]. The normalized-RGB color space is chosen because of its fast computation, since it does not require any color space transformation. Since we are using the OS-ELM classifier for learning the shadow and object pixels, the necessary transformation will be performed within the network itself.
4. *Normalized Cross Correlation* (*NCC*): NCC can identify scaled versions of the same signal. The closer the NCC value of the two signals is to 1, the more similar the two signals are. The texture in the shadow region is similar to the one in the background region. The difference between them is only luminosity. Therefore, the shadow region is a scaled version of the corresponding background region. NCC has been used for shadow detection by many researchers [6, 27, 28].
5. *Normalized Square Difference* (*NSD*): Similar to NCC, NSD also calculates the difference between background and input image patch caused by shadow.
6. *Edge Magnitude Distortion* (*EMD*): Sobel edges are extracted from both the background and the input image frame. The difference in average edge magnitude around the candidate pixel in the current frame and the background model is now extracted.
7. *Edge Orientation Distortion* (*EOD*): The difference in Sobel edge orientation at the candidate pixel position in the current frame and the background model is extracted. EMD and EOD will provide the feature for shadow discrimination regarding change in edge of the pixel neighborhood caused by shadow.

The features such as RGB component ratio, blue and red ratio, and normalized RGB component are first extracted only from candidate pixel position. The remaining features are extracted from the four-neighborhood regions of the candidate pixel position. Each feature contributes some information for shadow detection and discrimination. The dimension of the total features used in our system is higher than the features used in a previous study [15]. In that system, intensity ratio, color distortion, edge magnitude distortion, and edge gradient distortion extracted from the four-neighborhood of the candidate pixel are used. In our system, we use image patch similarity-based features (NCC and NSD) as additional features, which helped in improving shadow detection and discrimination. The response time and training time of the OS-ELM are faster, which also encourage us for adding more features to model shadow and object pixels.

4 Classification of shadow and object pixels

Enough labeled training data might not be available to learn from because of the high cost of manual labeling. On the other hand, training the classifier in one video scene might not work for another scene. Collecting the new training sample manually every time for each different scene is not practical. The solution for these problems is to use a semi-supervised learning technique, in which only a small set of labeled training data can be used to train the system at the beginning. Now, using a co-training scheme, the new training data can be collected online to update the classification system sequentially. In a previous study [15], several experiments were performed to prove the applicability of the co-training scheme for moving cast shadow detection. In our system, the OS-ELM classifier is used along with a co-training scheme for fast learning and updating the classification model.

At first, offline labeled training samples are collected and the features are extracted as explained in Section 3. We then split features into two groups; image patch similarity based features, and edge based features. Now the OS-ELMs in co-training scheme are trained separately for each group of feature set. In the shadow detection stage, in each frame, the detection result is first subjected to post-processing for result refinement. Once the refined result of shadow detection is obtained the shadow and object pixels are randomly selected and used for updating the OS-ELMs in co-training scheme. The process repeats for every frame. The online learning of OS-ELMs can be performed as soon as enough training samples are collected. In the following subsection, we will explain OS-ELM, co-training and post-processing scheme used in our system.

4.1 Brief review of OS-ELM

The development of OS-ELM was based on ELM for SLFNs with additive and RBF hidden nodes. Assume, we have N distinct samples $(x_i,t_i)\in R^n\times R^m$. If an SLFN with L hidden nodes can approximate these N samples with zero error, it then implies that there exist β_i, a_i, and b_i such that:

$$f_L(x_j)=\sum_{i=1}^{L}\beta_i G(a_i,b_i,x_j)=t_j,\quad j=1,\dots,N. \tag{1}$$

where a_i and b_i are the learning parameters of the hidden nodes, β_i is the output weight, and $G(a_i,b_i,x_j)$ denotes the output of the ith hidden node with respect to the input x_j. When using an additive hidden node, $G(a_i,b_i,x_j)=g(a_i.x_j+b_i),b_i\in R$, where a_i is the weight vector connecting the input layer to the ith hidden node, and b_i is the bias of the ith hidden node. When using an RBF hidden node, $G(a_i,b_i,x_j)=g(b_i\|x_j-a_i\|),b_i\in R^+$, where a_i and b_i are the center and impact width of the ith RBF node, respectively, and R^+ indicates the set of all positive real values. In this paper, we use additive hidden nodes with sigmoid activation functions.

Assume the network has L hidden nodes, and the data $\xi=\{(x_i,t_i)|x_i\in R^n,t_i,R^m,\ i=1,\dots,N\}$ is presented to the network sequentially (one by one or chunk by chunk). There are two phases in the OS-ELM algorithm: an initialization phase and a sequential phase. To insure the same learning performance of the OS-ELM as ELM, the number of training data required in the initialization phase N_0 has to be equal to or greater than L.

Step 1 (Initialization phase) Given a chunk of the initial training set, $\xi_0=\{(x_i,t_i)\}_{i=1}^{N_0}$ with $N_0\geq L$:

1. Randomly assign the input weight and bias within the range of [−1, 1].
2. Calculate the initial hidden layer output matrix H_0.

$$H_0=\begin{bmatrix} g(a_1.x_1+b_1) & \cdots & g(a_L.x_1+b_L)\\ \vdots & \cdots & \vdots \\ g(a_1.x_{N_0}+b_1) & \cdots & g(a_L.x_{N_0}+b_L)\end{bmatrix}_{N_0\times L} \tag{2}$$

3. Estimate the initial output weight $\beta^{(0)}=P_0H_0^TT_0$, where $P_0=(H_0^TT_0)^{-1}$ and $T_0=[t_1,\dots,t_{N_0}]^T$
4. Set k=0 (k indicates the number of chunks of data that is presented to the network).

Step 2 (Sequential learning phase) Present the $(k+1)$th chunk of new observations, $\xi_0 = \{(x_i, t_i)\}_{i=\left(\sum_{j=0}^{K} N_j\right)+1}^{\sum_{j=0}^{K+1} N_j}$, and N_{k+1} denotes the number of observations in the $(k+1)$th chunk.

1. Compute the partial hidden layer output matrix H_{k+1}

$$H_{k+1} = \begin{bmatrix} g(a_1.x_1 + b_1) & \cdots & g(a_L.x_1 + b_L) \\ \vdots & \cdots & \vdots \\ g\left(a_1.x_{N_{k+1}} + b_1\right) & \cdots & g\left(a_L.x_{N_{k+1}} + b_L\right) \end{bmatrix}_{N_{k+1} \times L} \tag{3}$$

2. Calculate the output weight matrix $\beta^{(k+1)}$.

$$P_{k+1} = P_k - P_k H_{k+1}^T \left(I + H_{k+1} P_k H_{k+1}^T\right)^{-1} H_{k+1} P_k \tag{4}$$

$$\beta^{(k+1)} = \beta^{(k)} + P_{k+1} H_{k+1}^T \left(T_{k+1} - H_{k+1} \beta^{(k)}\right) \tag{5}$$

3. Set $k=k+1$. Go to step 2, a sequential learning phase.

4.2 Co-training based shadow detection

Co-training is a common technique in semi-supervised learning, which was proposed by Blum and Michell in 1998 [1]. The idea comes from the fact that labeled data are usually scarce, whereas unlabeled data are plentiful and easy to obtain. The assumption is that, features can be split into two sets; and each sub-feature set is sufficient to train a good classifier. The two classifiers are trained using two different feature sets on the initial labeled data. For each classifier, the unlabeled data classified with the highest confidence are added to the labeled dataset, and the new dataset thus obtained is used to retrain the classifiers.

The process of choosing new data and labeling according to the classification results is an important step in the co-training technique. If the data is labeled incorrectly, after learning the classifier, it will have negative impact. In moving cast shadow detection, the classification of shadows and objects is a pixel-level operation. Generally, in the co-training scheme, the unlabeled data is chosen according to the highest confidence score. The problem is that sometimes a pixel inside the object region could be classified as a shadow pixel, and vice-versa. So, just choosing the new data according to the highest confidence score could result in choosing shadow data from an object region and object data from a shadow region. Therefore, instead of choosing new unlabeled data according to the highest confidence score, we have collected new training data from each video frame after post-processing of the shadow detection result. The post-processing step involves correction of the misclassified pixels according to the region label reasoning. Once the post-processing is performed with the result of shadow and object pixel classification, new training data is collected randomly from the refined shadow and object regions. This will force the classifiers to learn object pixels inside the object region as object pixels, even if they look like shadow pixels, and vice-versa. In each frame of a video sequence, there could be thousands of shadow and object pixels, and is not possible to update the classifier using all of the newly labeled data. Therefore, we have to set the temporal behavior of the co-training scheme. In OS-ELM, the classifier can be learned sequentially, and the data could be learned one by one or chunk by chunk. The sizes of chunks do not necessarily need to be equal. Therefore, we set a chunk size limit, and as long as the newly collected labeled data reach the chunk size, the OS-ELMs classifiers are updated to adapt to the new environment. In our experiment, we set the chunk size up to 100. Figure 1 shows the overall block diagram of the shadow detection based on co-training scheme used in our system.

As there are two independent OS-ELMs in the co-training scheme, the available features should be split into two sets F_1 and F_2. It is assumed that the two feature sets F_1 and F_2 should be conditionally independent. However, this assumption is not satisfied in most real-world data sets. Nigam et al. [19] claimed that even in such case, there is an advantage to be gained using co-training. We have extracted seven different types of features for each candidate pixel to be classified. The first three features (RGB component ratio, blue and red ratio, and normalized RGB components) are color features extracted only from the candidate pixel, whereas remaining features (NCC, NSD, EMD, and EOD) are extracted from the region around the candidate pixels. Basically, there are 3 types of features: color features, edge-based features, and image similarity-based features. We did several experiments by splitting features with different combinations. The best result was obtained by splitting edge-based features and image similarity-based features. The color features are combined with both feature sets. Finally, the features for OS-ELM classifier 1 are the RGB component ratio between the input frame and the background model, the blue and red component ratio within the input and background model, RMD, and EOD. In contrast, the features for OS-ELM classifier 2 are normalized RGB components within the input frame and background model, NCC, and NSD. Therefore, there are a total of 15 dimensional features for OS-ELM classifier 1 and 16 dimensional features for OS-ELM classifier 2.

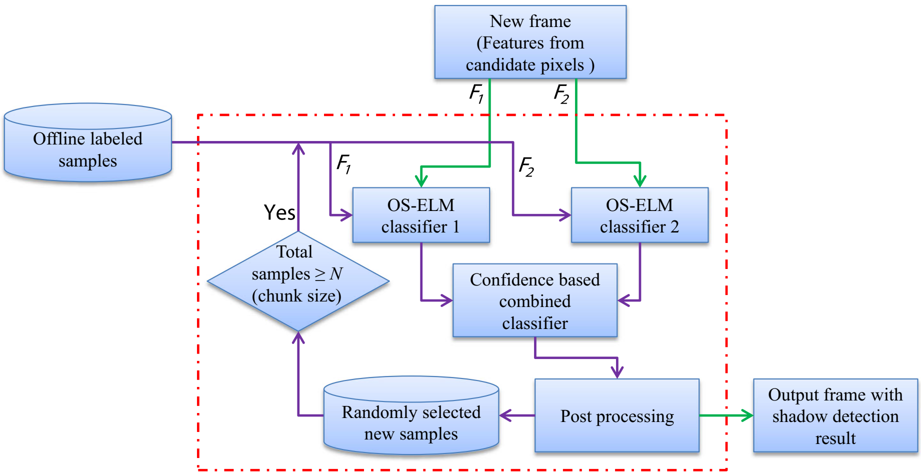


Fig. 1 Flow chart of the shadow detection with co-training scheme. The post-processing step is included inside the loop, and new training sample data from each input frame are collected based on the post-processing result of shadow detection

4.3 Post-processing

In the proposed OS-ELM co-training-based moving cast shadow detection method, the post-processing step is included inside the co-training loop, as shown in Fig. 1. The shadow detection using OS-ELMs is based on pixel-level operation, though the features for each candidate pixel are extracted by considering the small neighborhood around that pixel. There are many regions of misclassification because of self-shadows, background-like object regions, etc., which result in erroneous object shape estimates. Thus, we need some technique to recover shapes from frames obtained from the OS-ELM output. We used a post-processing approach similar to that proposed in a previous study [15] in order to refine the shadow detection result. Blob labeling is done along with other computations like blob area, blob boundary, and connectivity relation with the neighboring blobs. Based on some heuristic information about shadow and object regions, the object and shadow regions are flipped if necessary in order to obtain better shadow and object detection results. Figure 2 shows the results of shadow detection using OS-ELMs, the post-processing result, and the automatic collection of new training samples from shadow and object regions.

5 Results and discussion

5.1 Data sets

We performed experiments on a dataset that contains a set of standard indoor and outdoor video sequences. The sequences used in our experiments are (a) Campus, (b) Hallway, (c) Highway, (d) Lab, (e) Intelligent room, (f) Crossroad1, and (g) Crossroad2. The first five sequences were introduced in a previous study [20], and have been widely used for testing shadow detection performance. The remaining two sequences Crossroad1 and Crossroad2 are our own recorded outdoor sequences. There is a variation in object type, object size, background texture, background noise, shadow strength, and shadow direction from one sequence to another sequence. Therefore, each sequence presents a different challenge for the shadow detection methods to test their robustness. The shadow detection performance of the proposed method is compared with different methods in the literature in all the video sequences mentioned above.

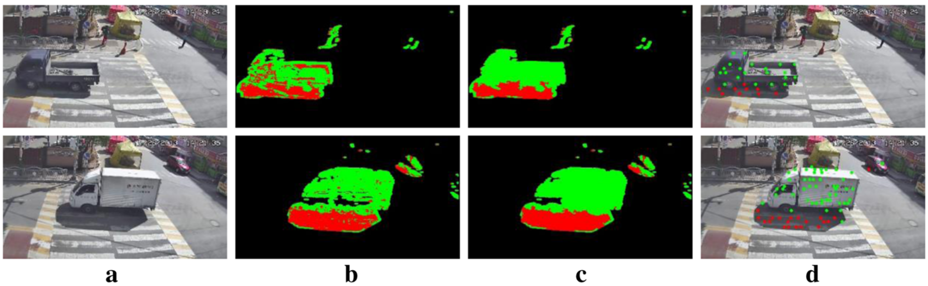


Fig. 2 **a** Input video frame, **b** result of background subtraction and shadow detection using OS-ELM co-training scheme, **c** result after post-processing using blob-level reasoning, and **d** sample training data collection (*green dots* show the pixel position from which object features are extracted, whereas *red dots* show the pixel position from which shadow features are extracted for online learning of the OS-ELM in the co-training process)

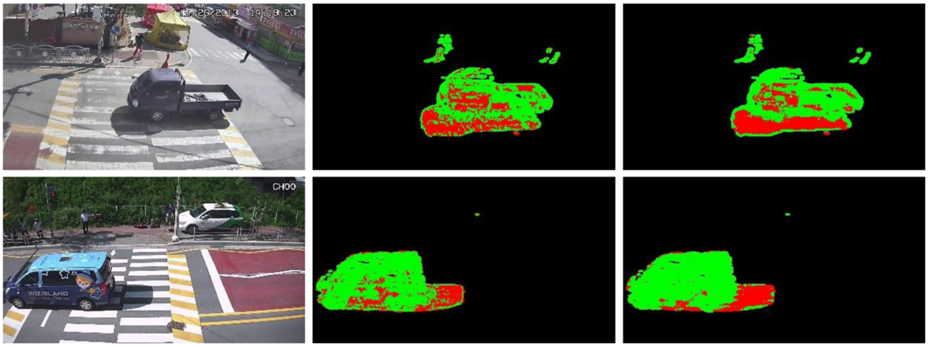
5.2 Shadow detection performance evaluation

Sanin et al. [23] presented a comparative evaluation of recent techniques for moving cast shadow detection. In this paper, we also use their implementation for evaluating the proposed method in comparison with different methods in the literature. We compared our method with a chromacity-based method [4], geometry-based method [7], texture-based methods [16, 22], and an SVM co-training-based method [15]. The chromacity-based method uses HSV color space to detect shadows. The rules are created in HSV space based on chromacity and luminosity information to separate shadow and object pixels. The geometric-based method [7] separates the blobs into individual objects and does geometric analysis to separate shadow and object regions. The small-region texture-based method [16] correlates texture around a small neighborhood of the candidate pixel using Gabor functions. The method first creates a mask with the potential shadow pixels in the foreground. Then, if the textures of the small region created at each pixel are correlated to the background reference, the pixels are classified as shadows. The large-region texture-based method [22] uses color features first to create large candidate shadow regions, which are then discriminated from objects using gradient-based texture correlation. Candidate shadow pixels are found using intensity and chromacity features, as in a previous study [4]. The texture for each candidate region is then correlated between the frame and the background reference to discriminate shadow and object regions. Finally, the SVM co-training-based method [15] uses color and edge features to learn shadow and object pixels. It is a semi-supervised learning technique, which adapts the classifiers using an automatically collected online training data set.

In the proposed method, we first analyzed the features that are proper for OS-ELM co-training-based moving cast shadow detection. There are two steps in the feature extraction module. The first step is background model generation and foreground pixel detection based on the background model. Next, features based on illumination, color, edges, and texture are extracted for each pixel in the foreground mask. The gradient and intensity-based features have been used for shadow and object pixel classification in many approaches. The proposed set of features performs better shadow detection and discrimination with OS-ELM compared with the features used previously [15]. Although the dimensionality of the proposed set of features is higher than that of the features used previously [15], the response time of the OS-ELM is much faster than that of the SVMs; therefore, OS-ELM can operate in real time even if the feature dimensionality is increased. Figure 3 shows the comparative results of the shadow detection using the proposed feature set and the features used previously [15]. From the figure we can see that the amount of misclassification of object and shadow pixel using [15] is larger than that of our proposed method.

5.2.1 Qualitative results

The qualitative results are shown in Fig. 4. The proposed method is compared with five other methods, and it is tested with seven different video samples. The first row in Fig. 4 shows an example frame from each video sample. The second row shows the ground truth of the moving shadow and object pixels in that frame. The remaining row shows the result of shadow detection and discrimination by different methods. Some methods are developed only to adapt for particular object detection. For example, the geometry-based method will work fine with separation of the shadows associated with pedestrians. In most of the video scenes, the shadow detection performance of the SVM co-training-based method is similar to that of the proposed



a b c

Fig. 3 Comparative results of the shadow detection using the proposed method with different sets of features without any post-processing operation: **a** input video frames, **b** shadow and object pixel classification using the features used in a previous study [6], and **c** shadow and object pixel classification using the proposed set of features

method, because both of them are co-training-based methods. Almost all methods works fine if the intensity of the shadow is not so strong. Almost all methods struggle to detect the shadow in the highway video scene, because the intensity of the shadow is strong in this sample video. In hallway video sample, almost all methods perform well, because the strength of the shadow is very low in that sequence. The shadow detection performance of the proposed method is stable in all test sequences, which proves that the proposed method works well with dynamic environments.

There are some advantages and disadvantages of the proposed dynamic shadow detection method. The method works perfectly if the intensity of the shadow is low such as hallway, campus, lab, room video sequences. On the other hand there is some misclassification if the intensity of shadow is very strong such as highway, crossroad1 and crossroad2 video sequences. In fact this is the common problem of all the shadow detection methods. From Fig. 4 we can see that in highway sequence because of very strong intensity shadow, some of the shadow pixels are misclassified as object pixel using proposed method. The shadow intensity in crossroad1 and crossroad2 video sequence is comparatively less than in highway sequence. Therefore the result of shadow pixel discrimination in the crossroa1 and crossroad2 sequence is better than the result in highway sequence. The proposed method also performs well if the illumination changes gradually. On the other hand, if the illumination changes abruptly, such as turning light source on and off, the proposed method may not work properly, because in such case there may not be enough time to update the classifier with changing environment,.

5.2.2 Quantitative results

We measured the shadow detection performance of each method in every test sequences. Two metrics were proposed by Prati et al. [20] to evaluate the quantitative shadow detection performance: the shadow detection rate (η) and the shadow discrimination rate (ξ).

$$\eta = \frac{TP_{Shadow}}{TP_{Shadow} + FN_{Shadow}} \tag{6}$$

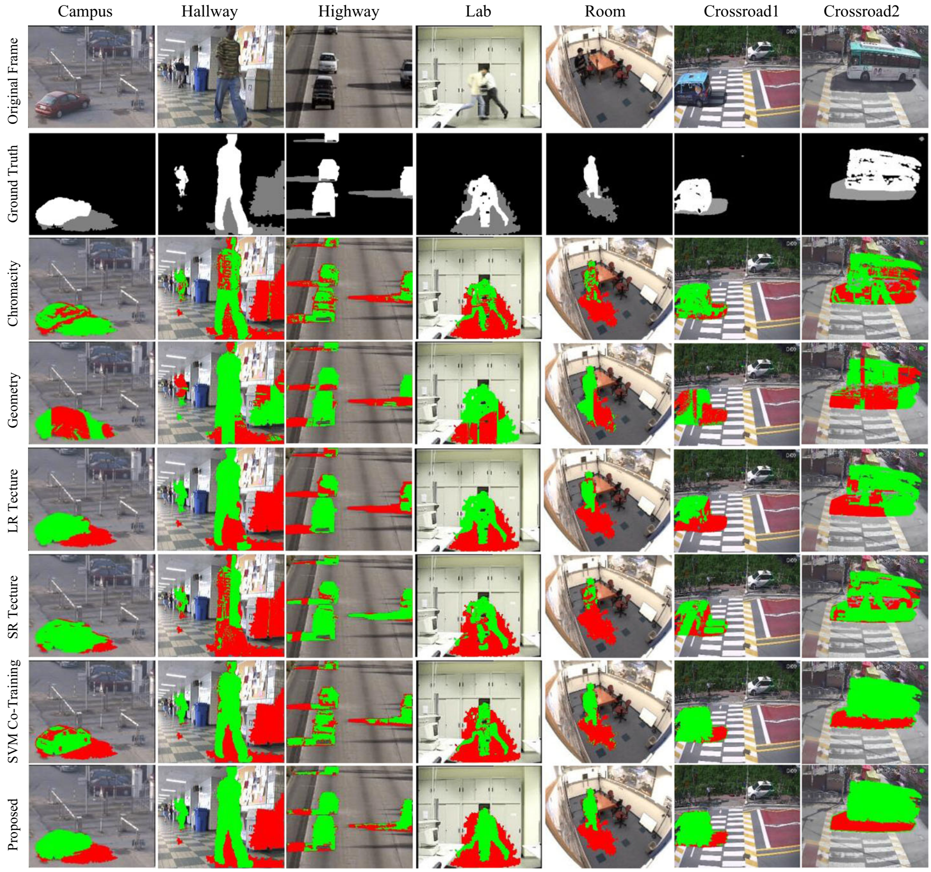


Fig. 4 Qualitative shadow detection results. The first row shows an example frame from each video sequence. The second row shows the ground truth of the moving object and shadow pixels. The remaining rows show the shadow detection results for each method. The shadow detection result of the proposed method is given in the last row. *Red* pixels denote the detected shadow pixels, and *green* pixels denote the detected object pixels

$$\xi = \frac{TP_{Object}}{TP_{Object} + FN_{Object}} \tag{7}$$

where *TP* and *FN* stand for the true positive and false negative pixels with respect to either shadows or foreground objects. These measures are calculated based on the ground truth shadow and object pixels obtained by manually labeling the foreground pixels as shadow and object pixels. The shadow detection rate is concerned with classifying the maximum number of shadow pixels as shadow pixels, and the shadow discrimination rate is concerned with classifying moving object pixels as object pixels.

Figure 5a, b show the shadow detection and discrimination rates, respectively, and Fig. 5c shows the average of the shadow detection and discrimination rates on each test sequence using different methods. For better shadow detection performance, the balance between shadow detection and the discrimination rate is very important. If all the moving pixels in the image are labeled as shadow pixels, the detection rate will be the maximum while

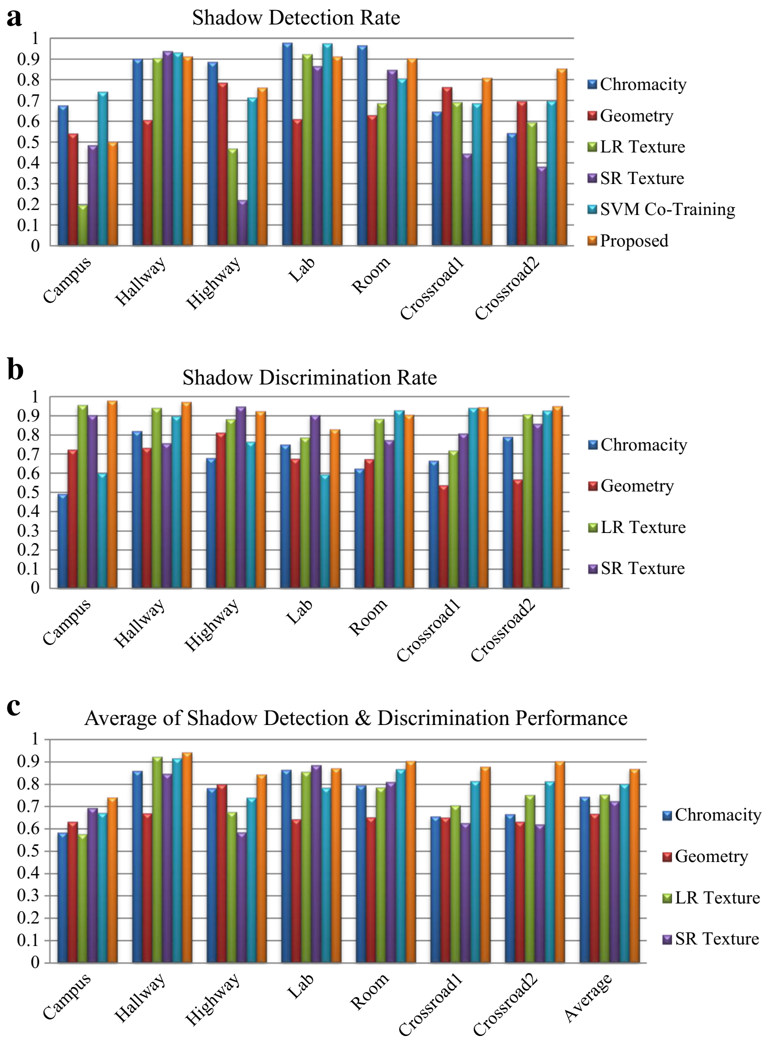


Fig. 5 Quantitative evaluation of shadow detection results, **a** shadow detection rate, **b** shadow discrimination rates, and **c** average between shadow detection and discrimination rates

the shadow discrimination rate will be very low, and vice versa. Therefore, less the difference between the shadow detection and discrimination rate improves the performance with obviously high detection and discrimination rates. The shadow detection rate of the proposed method seems to be a little less than the rest of the methods, but the average of the shadow detection and discrimination rate of the proposed method is better than the rest of the methods with little difference between the detection and discrimination rates, and the result are stable in all video sequences. The average shadow detection and discrimination rates of the co-training-based methods are better than the rest of the methods. On the other hand, co-training-based methods are stable across all the test video sequences. The performance of the proposed method is better than the other methods in almost all test sequences.

Figure 6 shows the frame-wise shadow detection and discrimination performance in the Crossroad2 video sequence. There is little fluctuation in the shadow detection rate from one frame to another frame. This is due to the nature of the shadow in each frame. The frames are recorded at certain intervals of time (every 100th frame). Therefore, in each frame, the object could vary in size, color, position, occlusion, etc., which affects the nature of the shadow in each case. Due to the the different nature of the shadow, the detection rate is also different in each case, but the shadow discrimination rate follows the curve of the shadow detection rate. It is concluded that in some frames, shadow detection is easy, whereas in some frames, shadow detection is a little difficult. Additionally, the post-processing operation will also affect the shadow detection and discrimination rate in each frame. Overall, in Fig. 6, it can be seen that the proposed method is quite stable. The average of the shadow detection and discrimination performance curve is smooth and ranges from 0.81 to 0.97.

The average processing time of proposed dynamic shadow detection method ranges from 9 to 16 frames per second. We are using 64-bit Intel CPU running at 3.50 GHz. It depends on the amount of candidate pixels in a scene to be classified as object or shadow. If the scene is congested and consists of large portion of candidate pixels computational complexity also increases accordingly. The frame processing rate can be further improved by realizing co-training and shadow pixel classification stage into parallel manner because we do not need to wait for the co-training result to classify pixels in every frame. In general, the computational complexity of non-adaptive shadow detection method is quite higher than that of adaptive methods.

5.3 Adaptability with dynamic environment

One of the main advantages of the proposed shadow detection method is the adaptability of the OS-ELM classifier according to the changing environment. In the Crossroad2 video sequence, the static shadow and illumination changes over time. Figure 7 shows the shadow detection result of the proposed method in three different frames captured at different time intervals. There is a little change in the illumination and static shadow due to surrounding buildings in all three sample frames, but the shadow is detected well regardless of the illumination conditions and changing environment. Although the manually labeled training data has been collected from the first few frames of the video sequence, the co-training technique adapted the classifier for the changing environment. As we are using OS-ELM for shadow and object pixel

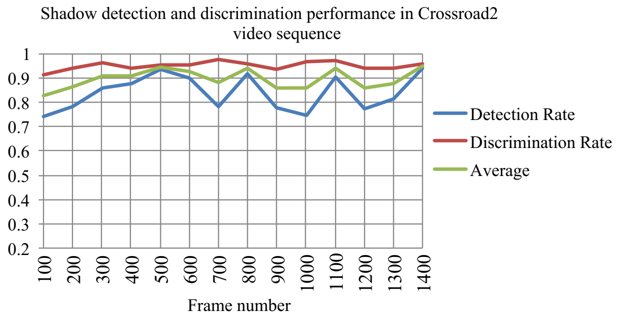


Fig. 6 Shadow detection and discrimination performance of the proposed method for Crossroad2 video sequence in fixed frame intervals

Fig. 7 Input sample video frames captured at different time intervals from the Crossroad1 sequence and corresponding shadow detection results. In all three frames, the static shadow position is different, and there is little change in the illumination as well

classification, we can update the classifier in the co-training loop for every frame if the illumination of the scene is changing frequently.

6 Conclusions

We presented a semi-supervised approach to differentiate between moving shadows and object pixels in a video sequence using OS-ELM and co-training algorithms. We proposed modified co-training technique, which improved the shadow detection performance significantly. The collection of new training samples after the post-processing step minimizes the possibility of wrong sample collection for online training of the OS-ELM classifiers. As the classifiers are updated in real time with new training samples, the system adapts automatically to the changing environmental conditions such as illumination changes, which happens quite often in outdoor scenes. The proposed scheme can run in real time because the online learning of OS-ELM is much faster than other sequential learning algorithms. Several experiments have been performed to show the validity of the proposed shadow detection technique. Our method was tested with several video samples containing both indoor and outdoor scenes with shadows of different nature. Both qualitative and quantitative results show that the proposed method can perform better shadow detection and discrimination compared with other methods. The strong points of the proposed technique are, it can operate in real time, small set of human-labeled data required at the beginning, and it can adapt automatically to changing scene conditions with the help of co-training.

References

1. Blum A, Mitchell T (1998) Combining labeled and unlabeled data with co-training. 11th Annual Conference on Computation Learning Theory, pp 92–100
2. Cavallaro A, Salvador E, Ebrahimi T (2005) Shadow-aware object-based video processing. IEEE Proc Vision Image Signal Process 152(4):398–406

3. Chen CC, Aggarwal JK (2010) Human shadow removal with unknown light source. International Conference on Pattern Recognition, pp 2407–2410
4. Cucchiara R, Grana C, Piccardi M, Prati A (2003) Detecting moving objects, ghosts, and shadows in video streams. IEEE Trans Pattern Anal Mach Intell 25(10):1337–1342
5. El-Zahhar MM, Karali A, ElHelw M (2011) A semi supervised learning-based method for adaptive shadow detection. IEEE International Conference on Signal and Image Processing Applications, pp 348–353
6. Grest D, Frahm JM, Koch R (2003) A color similarity measure for robust shadow removal in real-time. Vision Modeling and Visualization, pp 253–260
7. Hsieh JW, Hu WF, Chang CJ, Chen YS (2003) Shadow elimination for effective moving object detection by Gaussian shadow modeling. Image Vis Comput 21:505–516
8. Huang JB, Chen CS (2009) Moving cast shadow detection using physics-based features. IEEE Conference on Computer Vision and Pattern Recognition, pp 2310–2317
9. Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: theory and applications. Neurocomputing 70:489–501
10. Jalal AS, Singh V (2014) A framework for background modeling and shadow suppression for moving object detection in complex wavelet domain. Multimedia Tools Appl 73:779–801
11. Jarraya SK, Boukhriss RR, Hammami M, Ben-Abdallah H (2012) Cast shadow detection based on semi supervised learning. LNCS 7324:19–26
12. Jia Y, Yu X, Dai J, Hu W (2012) A novel moving cast shadow detection of vehicles in traffic scene. Proceedings of the Third Sino-foreign-Interchange Conference on Intelligent Science and Intelligent Data Engineering, pp 115–124
13. Jiang K, Li A, Cui Z, Wang T, Su Y (2013) Adaptive shadow detection using global texture and sampling detection. IET Comput Vis 7(2):115–122
14. Joshi AJ, Atev S, Masoud O, Papanikolopoulos N (2007) Moving shadow detection with low and mid level reasoning. IEEE International Conference on Robotics and Automation, pp 4827–4832
15. Joshi AJ, Papanikolopoulos NP (2008) Learning to detect moving shadows in dynamic environments. IEEE Trans Pattern Anal Mach Intell 30(11):2055–2063
16. Leone A, Distance C (2007) Shadow detection for moving objects based on texture analysis. Pattern Recogn 40:1222–1233
17. Li G, Qin L, Huang Q (2013) Adaptive moving cast shadow detection. The Era of Interactive Media, pp 399–409
18. Liang NY, Huang GB, Saratchandran P, Sundararajan N (2006) A fast and accurate online sequential learning algorithms for feedforwards networks. IEEE Trans Neural Netw 17:1411–1423
19. Nigam K, Ghani R (2000) Analyzing the effectiveness and applicability of co-training. 9th International Conference on Information and Knowledge Management, pp 86–93
20. Prati A, Mikic I, Trividi M, Cucchira R (2003) Detecting moving shadows: algorithms and evaluation. IEEE Trans Pattern Anal Mach Intell 25(7):918–923
21. Salvador E, Cavallaro A, Ebrahimi T (2004) Cast shadow segmentation using invariant color features. Comput Vis Image Underst 95:238–259
22. Sanin A, Sanderson C, Lovell BC (2010) Improved shadow removal for robust person tracking in surveillance scenarios. International Conference on Pattern Recognition, pp 141–144
23. Sanin A, Sanderson C, Lovell BC (2012) Shadow detection: a comparative evaluation of recent methods. Pattern Recogn 45:1684–1695
24. Song KT, Tai JC (2007) Image-based traffic monitoring with shadow suppression. Proc IEEE 95(2):413–426
25. Staufer C, Grimson WEL (1999) Adaptive background mixture model for real-time tracking. IEEE International Conference on Computer Vision and Pattern Recognition, pp 246–252
26. Yang MT, Lo KH, Chiang CC, Tai WK (2007) Moving cast shadow detection by exploiting multiple cues. IET Image Process 2(2):95–104
27. Yang Y, Ming Y, Yongchao M (2009) A Strategy to detect the moving vehicle shadows based on gray-scale information. Second International Conference on Intelligent Networks and Intelligent Systems, pp 358–361
28. Yuan C, Yang C, Xu Z (2010) Simple vehicle detection with shadow removal at intersection. Second International Conference on Multimedia and Information Technology, pp 188–191
29. Zhigang L, Xiao L, Yaonan W, Xi H (2013) Adaptive moving cast shadow detection by integrating multiple cues. Chin J Electron 22(4):757–762

Deepak Ghimire received undergraduate degrees in Computer Engineering from Pokhara University, Nepal in 2007. He received M.S. degree and Ph.D. degree in Computer Science and Engineering from Chonbuk National University, South Korea in 2011 and 2014 respectively. He is currently a Researcher in IT Application Research Center of KETI (Korea Electronics Technology Institute), South Korea. His main research interests include image processing, computer vision, machine learning, and biometric information processing.

Joonwhoan Lee received undergraduate degrees in Electronic Engineering from the Hanyan University in 1980. He received M.S. degree in Electrical and Electronic Engineering from KAIST (Korea Advanced Institute of Science and Technology) University in 1982 and Ph.D. degree in Electrical and Computer Engineering from University of Missouri, USA in 1990. He is currently a Professor in the Department of Computer Engineering at Chonbuk National University, South Korea. His research interests include image processing, computer vision, emotion engineering.