Journal of information and communication convergence engineering
http://jicce.org

J. Inf. Commun. Converg. Eng. 13(1): 42-49, Mar. 2015 **Regular paper**

Laser Spot Detection Using Robust Dictionary Construction and Update

Zhihua Wang, Yongri Piao, and Minglu Jin*, *Member, KIICE*

School of Information and Communication Engineering, Dalian University of Technology, Dalian, Liaoning 116024, China

Abstract

In laser pointer interaction systems, laser spot detection is one of the most important technologies, and most of the challenges in this area are related to the varying backgrounds, and the real-time performance of the interaction system. In this paper, we present a robust dictionary construction and update algorithm based on a sparse model of background subtraction. In order to control dynamic backgrounds, first, we determine whether there is a change in the backgrounds; if this is true, the new background can be directly added to the dictionary configurations; otherwise, we run an online cumulative average on the backgrounds to update the dictionary. The proposed dictionary construction and update algorithm for laser spot detection, is robust to the varying backgrounds and noises, and can be implemented in real time. A large number of experimental results have confirmed the superior performance of the proposed method in terms of the detection error and real-time implementation.

Index Terms: Background subtraction, Laser spot detection, Dictionary construction and update, Compressive sensing

I. INTRODUCTION

Recently, we have witnessed a growing interest in laser pointer interaction (LPI), which allows users to interact directly from a distance through a laser pointer. In laser pointer-based interaction systems, the captured laser spot is recognized and used for interactions by using various image processing techniques. The advantage of ensuring movement flexibility for users has led to the widespread use of this method for multimedia presentations [1-4], robot navigation [5-7], medical purposes [8], virtual reality systems [9, 10], and smart houses [11].

Recently, Kim et al. [2] summarized three fundamental problems with LPI: laser spot detection, interaction function, and coordinate mapping. In [11-13], the researchers focused on the development of a laser spot detection algorithm that directly influences the performance of LPI systems. The most difficult challenges of laser spot detection are strong light environments, real-time implementation, and dynamic backgrounds. For example, the background information always changes when the speaker turns the slides in practical presentation cases.

To overcome the above mentioned problems, two types of algorithms, namely target search (TS) and background subtraction (BGS), have been developed to detect a laser spot. The TS method directly searches the laser spot without considering the background. Shin et al. [12] simply searches for pixels with maximum intensities to detect the location of the laser spot. Chávez et al. [11] used a combination of template matching and fuzzy rule-based systems to improve the success rate of laser spot detection. Geys and Van Gool [13] determined the laser spot by using clusters along with the fact that a group effect is caused on laser spots by hand jitters. However, the TS method fails because of the strong

Received 29 November 2014, **Revised** 24 December 2014, **Accepted** 20 January 2015
***Corresponding Author** Minglu Jin (E-mail: mljin@dlut.edu.cn, Tel: +86-411-8470-7719)
School of Information and Communication Engineering, Dalian University of Technology, Dalian, Liaoning116024, China.

Open Access **http://dx.doi.org/10.6109/jicce.2015.13.1.042** print ISSN: 2234-8255 online ISSN: 2234-8883

light environment and the appearance change of the moving laser spot. On the other hand, BGS covers a set of methods that aim to distinguish between the foreground and the background areas by utilizing a background model. The traditional models used to represent background include statistical models, neural networks, estimation models, and some recent models including fuzzy models, subspace models, transform domain models, and sparse models [14]. Among them, sparse models have been successfully applied in compressive sensing [15]. Cevher et al. [16] considered background subtraction as a sparse approximation problem and provided different solutions based on convex optimization. Hence, the background is learned and adapted in a low-dimensional compressed representation, which is sufficient to determine spatial innovations. Huang et al. [17] proposed a new learning algorithm called dynamic group sparsity (DGS). The idea is that the nonzero coefficients in the sparse data are often not random but tend to be a cluster such as those in the case of foreground detection. However, the dictionary of backgrounds is constructed simply by using video frames that make this model sensitive to noise and background changes. In order to solve the problem of background changes and outliers in training samples, Zhao et al. [18] formulated background modeling as a dictionary learning problem. However, the learning process is time consuming and needs all the background information, which makes it difficult to apply in practice. Therefore, to solve the problem discussed in [18], we propose a novel robust algorithm for the construction and update of a dictionary for laser spot detection. Subsequently, the proposed model can control the varying backgrounds and the real-time performance.

The remainder of this paper is organized as follows: Section II briefly explains the proposed method of background modeling and foreground detection. In Section III, we show the experimental results in comparison with those of the existing methods, and some conclusions of the proposed method are presented in Section IV.

II. THE PROPOSED SYSTEM MODEL

Suppose that we have an image Y of size $n_1 \times n_2$ and we vectorize it into a column vector y of size $n \times 1$ ($n = n_1 \times n_2$) by concatenating the individual column of Y in the order from first to last. We formulate the background subtraction as a linear decomposition problem, i.e., to find a background component y_B and a foreground component y_F that together constitute a given frame y:

$$y = y_B + y_F , \tag{1}$$

where y_B and y_F denote the column vectors of background a nd foreground, respectively.

A. Sparse Representation

Suppose that we have K different backgrounds $y_{B1}, y_{B2}, ..., y_{BK} \in R^n$; then, we can build K configurations for dynamic backgrounds with each configuration standing for one background. Therefore, at a specific frame, the background y_B can choose from one of these configurations. We define a new matrix $D = [d_1, d_2, ..., d_K]$ as the concatenation of all the configurations; here, d_i denotes the i^{th} configuration. Then, we say that background y_B has the linear representation $y_B = d_i x_i$, where x_i denotes a coefficient representing the relationship between y_B and d_i. Thus, the background can be modeled as a sparse linear combination of atoms from a dictionary D, each atom of which characterizes one of the configurations. Next, we rewrite y_B in terms of D as follows:

$$y_B = Dx , \tag{2}$$

where $x = [0, ..., 0, x_i, 0, ..., 0]^T$ denotes a sparse coefficient vector whose entries are ideally zeros except at positions associated with x_i.

Zhao et al. [18] summarized two assumptions for this sparse model:

Assumption 1. Background y_B of a specific frame y has a sparse representation over a dictionary D.

Assumption 2. The candidate foreground y_F of a frame is sparse after background subtraction.

On the basis of these two assumptions, the BGS problem can be interpreted as follows: given a frame y, find a decomposition that has the sparse coded background $y_B = Dx$ and the sparse foreground $y_F = y - Dx$:

$$x = \arg\min_x \| y - Dx \|_0 + \lambda \| x \|_0 , \tag{3}$$

where $\| x \|_0$ denotes the ℓ_0-norm counting the number of nonzero elements of x, D indicates the dictionary capturing of all the background configurations, and λ represents the weighting parameter balancing between the two terms.

Since Eq. (3) is an NP-hard problem because of the non-convexity of ℓ_0-norm, Zhao et al. [18] replaced ℓ_0-norm with ℓ_1-norm and obtained the ℓ_1-measured and ℓ_1-regularized convex optimization problem:

$$x = \arg\min_x \| y - Dx \|_1 + \lambda \| x \|_1 . \tag{4}$$

Considering the LPI application, the foreground (laser spot) generally occupies a far smaller spatial area than the background. Therefore, we can simply treat the foreground as noises and obtain a Lasso problem:

$$x = \arg\min_x \| y - Dx \|_2^2 + \lambda \| x \|_1 . \tag{5}$$

This problem can be easily and rapidly solved using least

angle regression (LARS) [19], and then, we can obtain the foreground using

$$y_F = y - Dx . \quad (6)$$

B. Dictionary Construction

To make the sparse model robust against dynamic backgrounds, the dictionary must be able to represent all the backgrounds. Huang et al. [17] assumed that background subtraction has already been performed on the first K frames of the video sequences and let $D = [\mathrm{y}_1, \mathrm{y}_2, ..., \mathrm{y}_\mathrm{K}] \in \mathrm{R}^{\mathrm{n}\times\mathrm{K}}$. It is noteworthy that this method is sensitive to noise and cannot be used in practice. Zhao et al. [18] collected all background training samples and developed a robust dictionary learning approach to construct the dictionary:

$$D = \arg\min_{D, x_m} \sum_{m=1}^{M} \| y_m - Dx_m \|_1 + \lambda \| x_m \|_1 . \quad (7)$$

However, in LPI applications, we are unable to collect a sufficient number of training samples. For example, we are unable to capture a large number of backgrounds in a presentation application since we do not know the information of the next slide until the user gives the 'PageDown' or 'PageUp' command. Besides, solving this optimization problem is time consuming and the solution is difficult to implement in real-time.

Since the use of video sequences as a dictionary is sensitive to noise, we use information from multiple frames for ensuring robustness. Therefore, the strategy is to apply an exponentially decaying weight to run an online cumulative average on the backgrounds:

$$\begin{cases} d_j = y_j & \text{for } j = 1 \\ d_j = \alpha d_{j-1} + (1-\alpha)\,\mathrm{y}_j & \text{for } 1 < j \le K \end{cases}, \quad (8)$$

where α denotes the decay rate often chosen as a tradeoff between stability and quick update and K represents the parameter that controls the number of backgrounds. The advantage of this approach apart from its simplicity is that it can suppress noise and solve the problem low-frequency background changes to some extent. We assume that the background changes at a high frequency at the dictionary update stage but not the dictionary construction stage, which is often true in an LPI application.

C. Dictionary Update

The dictionary needs to update quickly in order to handle the occurrence of a new background. Huang et al. [17] set a time window to update the dictionary. For frame t, the dictionary is updated by $D = [\mathrm{y}_{\mathrm{t}-K}, ..., \mathrm{y}_{\mathrm{t}-2}, \mathrm{y}_{\mathrm{t}-1}]$. However, this method is still sensitive to noise, which makes the model unstable. Zhao et al. [18] updated the dictionary D by solving the following optimization problem with the coefficients being updated and considered constant:

$$D = \arg\min_{D} \sum_{m=1}^{M} \| \mathrm{y}_\mathrm{m} - \mathrm{Dx}_\mathrm{m} \|_1 . \quad (9)$$

Zhao et al. [18] assumed that the atoms in D are independent of each other and thus, updated each of them separately. However, solving this optimization problem is still time consuming.

Considering that when a new background occurs, the foreground y_F solved by Eqs. (5) and (6) will not be a sparse result, we can figure out whether a new background occurs by setting a threshold for the ℓ_0-norm of y_F. Whenever a new background occurs, we add the new background configuration into the dictionary; otherwise, we directly update the dictionary by using Eq. (8). This method can be formulated as follows:

$$\begin{cases} d_j = y_i & \text{if } \| y_F \|_0 > Th \\ d_j = \alpha d_{j-1} + (1-\alpha)\,\mathrm{y}_\mathrm{i} & \text{otherwise} \end{cases}, \quad (10)$$

Where Th can be set as the size of the laser spot.

Table 1. Description of the proposed dictionary construction and update algorithm

Algorithm: Dictionary Construction and Update

Input:

The continuously captured frames $y_1, \mathrm{y}_2, ..., \mathrm{y}_\mathrm{N}$; number of dictionary configurations K ; weighting parameter λ ; decay rate α ; threshold Th for x_F

Dictionary Construction:

$d_1 = y_1$

for $i = 2, ..., K$ do

$d_i = \alpha d_{i\text{-}1} + (1\text{-}\alpha) y_i$

end for

Dictionary Update:

for all $i = K+1, ..., \mathrm{N}$ do

$j = \mathrm{mod}(\mathrm{i}, \mathrm{K}) + 1$

$x = \arg\min_x \| \mathrm{y_i} - \mathrm{Dx} \|_2^2 + \lambda \| \mathrm{x} \|_1$

$y_F = y_i - Dx$

if $\| y_F \|_0 > Th$ then

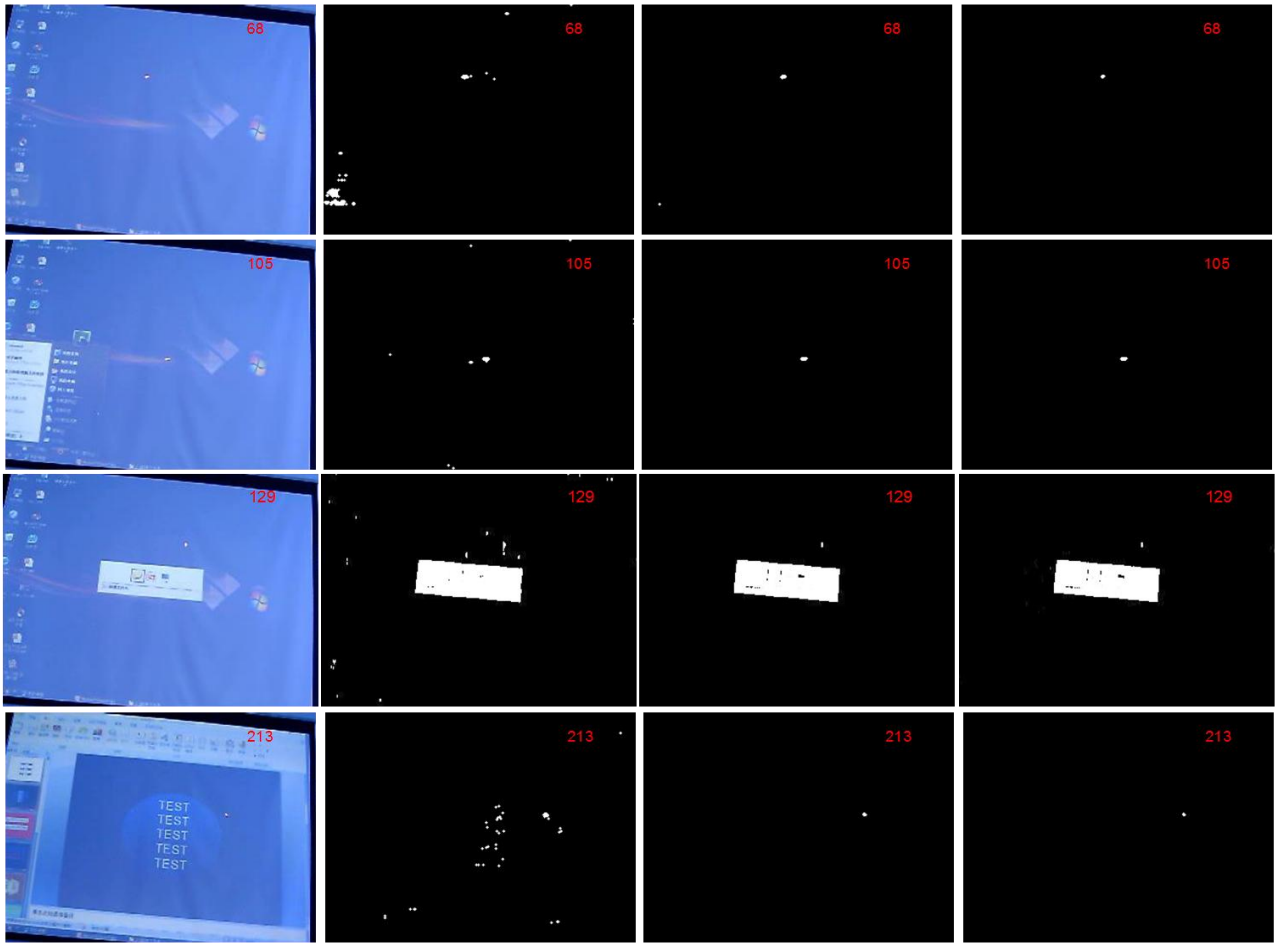
$d_j = y_i$

else then

$\begin{cases} d_j = \alpha d_K + (1-\alpha)\,\mathrm{y_i} & \text{for j=1} \\ d_j = \alpha d_{j-1} + (1-\alpha)\,\mathrm{y_i} & \text{for j} \neq 1 \end{cases}$

end if

end for



(a) (b) (c) (d)

Fig. 1. Results on laser pointer-operated Windows. (a) Original image (size: 320×240). (b) Using video images as dictionary [17]. (c) Dictionary learning method [18]. (d) Proposed method.

The proposed strategy is made sensitive to changing backgrounds by adding new background configurations, and robust against noise by using the online cumulative average of the backgrounds. The proposed dictionary construction and update algorithm is summarized in Table 1.

III. EXPERIMENTS AND RESULTS

To validate the ability of the proposed algorithm to handle the above mentioned high-frequency background changes and evaluate the algorithm's real-time performance, in this section, we discuss two experiments of LPI. Through these experiments, we evaluated the performance of the proposed algorithm with the different parameters used in this algorithm, measured the detection error under dynamic backgrounds, and compared it with the running times of different algorithms as well.

A. Laser Pointer-Operated Windows

A typical example of LPI in practice is the interactive demonstration of software with a computer whose screen content is sent to a video beamer by using a common laser pointer tracked by a video camera as an input device. Algorithms use the behavior of the laser spot to realize the functions of Button Press, Button Release, and Mouse Move. When Button Press is recognized, the corresponding file or dialog may show up, which leads to a background change immediately. We record three videos of the size 160×120, 320×240, and 640×480, respectively, to simulate this process on such a system.

In LPI, the laser spot cannot be static because of the hand jitter, thus instead of measuring the detection error compared with the ground truth, we validate it using the possibility of false detected frames as follows:

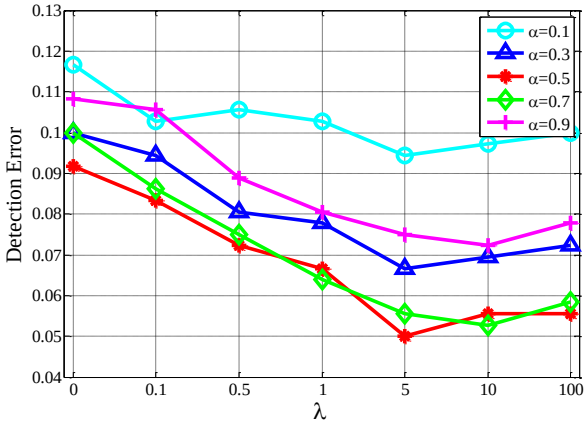


Fig. 2. Detection error with different parameters λ and α.

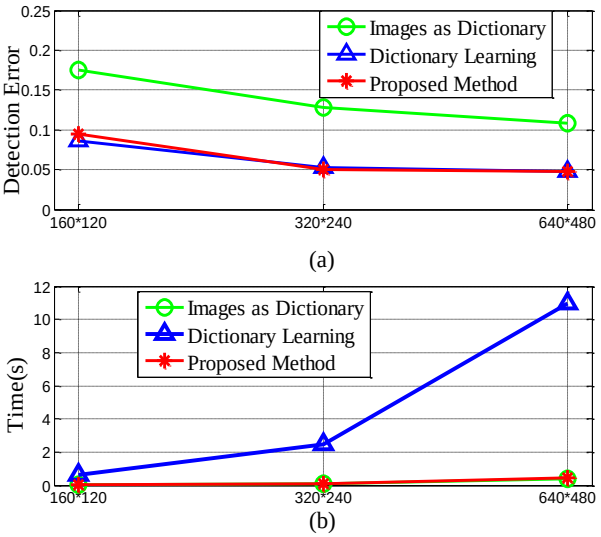


Fig. 3. Results on laser pointer-operated Windows. (a) Detection errors. (b) Running time.

$$Error = \frac{Number\ of\ false\ detected\ frames}{Number\ of\ total\ frames}. \qquad (11)$$

The performance of the proposed algorithm is compared with that of two algorithms representing state-of-the-art sparse model approaches [17, 18]. Notice that we use LARS [19] to solve Eq. (5) for all these methods in order to evaluate the dictionary construction and update approach. Fig. 1 illustrates some results of the abovementioned algorithms.

Image sequences having a size of 320×240 are used to test how the parameters λ and α determine the detection performance. The detection errors of different parameter values are shown in Fig. 2. As we can see from Fig. 2, a larger weighting parameter λ is helpful for the detection since the sparsity of the background is the key assumption of the proposed algorithm. However, a considerably large λ value increases the reconstruction error, which leads to relatively low performance. Thus, the value of λ can be chosen from 5 to 10 in order to obtain good performance. The decay rate α is used against noises; a small α value is sensitive to noises, and a large one cannot adapt to a low frequency of background changes.

As can be observed in Fig. 2, a moderate α value of 0.5 can lead to better performance. In our experiments, the weighting parameter was set at $\lambda = 5$ and the decay rate at $\alpha = 0.5$.

As the other parameter values used in these tests, we select $K = 20$ to build the dictionary and $Th = 50$ to control the sparsity of the laser spot. A standard PC with a 2.0-GHz Intel CPU processor and 3 GB of memory is used in our experiments. As can be seen from Fig. 1, our algorithm can handle a situation that has dynamic backgrounds and is robust against noise. The final results of the detection error defined by Eq. (11) and the running time per frame are illustrated in Fig. 3. As can be observed, our algorithm achieves detection errors that are as low as those of the dictionary learning approach and consumes as little time as the using video images as dictionary method. Notice that the detection error of the using video images as dictionary method [17] is considerably higher than that of our algorithm, and that dictionary learning [18] consumes a considerably large amount of time and thus, cannot be implemented in real time.

B. Multimedia Presentation

In a presentation application, we can use the laser pointer to change slides and draw lines. It should be noted that high-frequency changes are caused when the user changes the slides. Further, each slide may be totally different from the others. For this application, we manually change the slides to obtain dynamic backgrounds and use the above mentioned algorithms for the detection of the laser spot. The final results are shown in Figs. 4 and 5.

From Figs. 4 and 5, we can see that the proposed algorithm can achieve a lower detection error with a low time cost, which is similar to the results of the laser pointer-operated windows method. Thus, the proposed algorithm is robust against different scenarios with dynamic backgrounds. From Table 2, we can see that the detection error when the image resolution 160×120 is the highest, while similarly low detection errors are obtained when the resolutions of 320×240 and 640×480 are used. However, the time cost of using the resolution of 640×480 is considerably higher than that of using the resolution of 320×240. Thus, we recommend the use of the 320×240 resolution in practice.

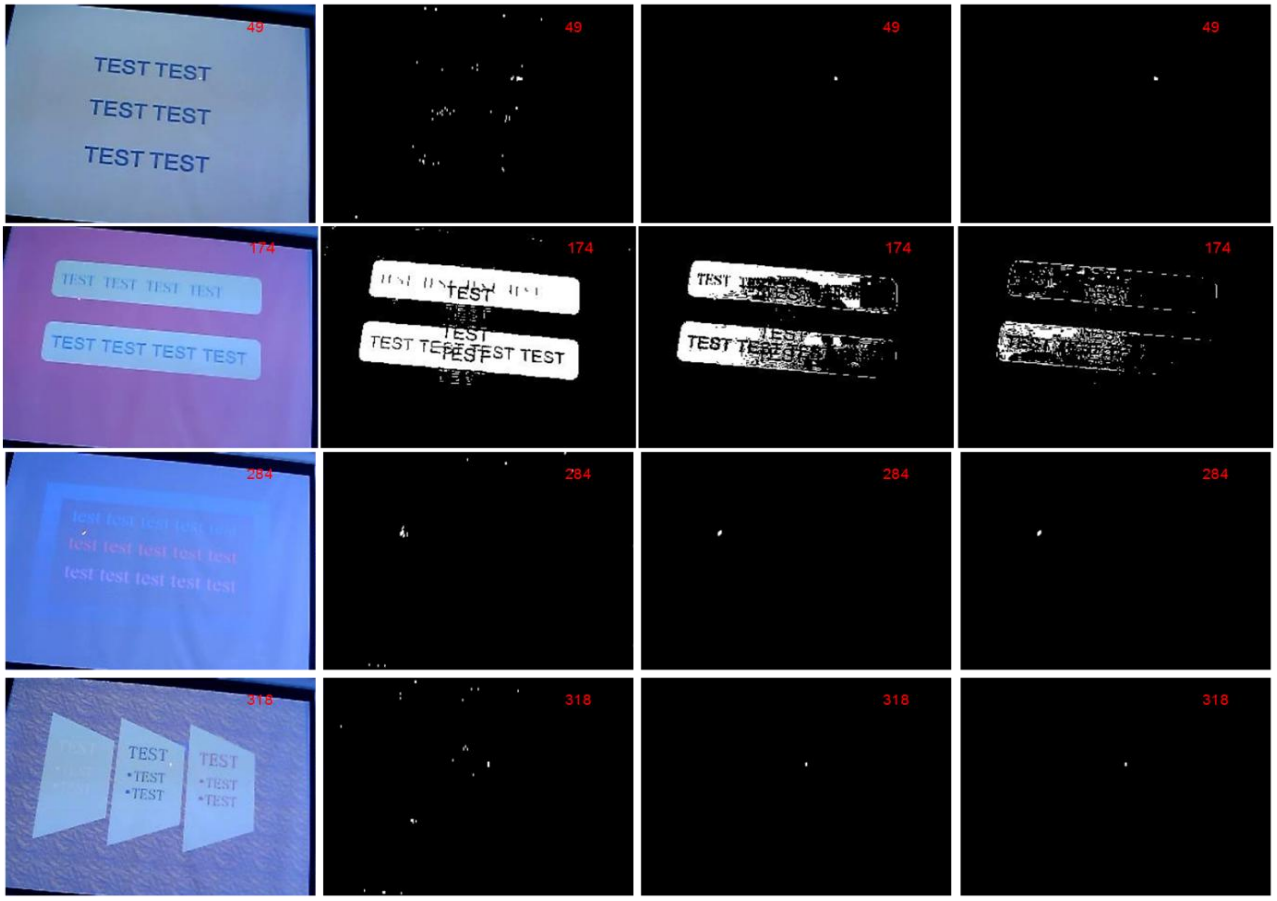


Fig. 4. Results of multimedia presentation. (a) Original image (size: 320×240). (b) Using video images as dictionary [17]. (c) Dictionary learning method [18]. (d) Proposed method.

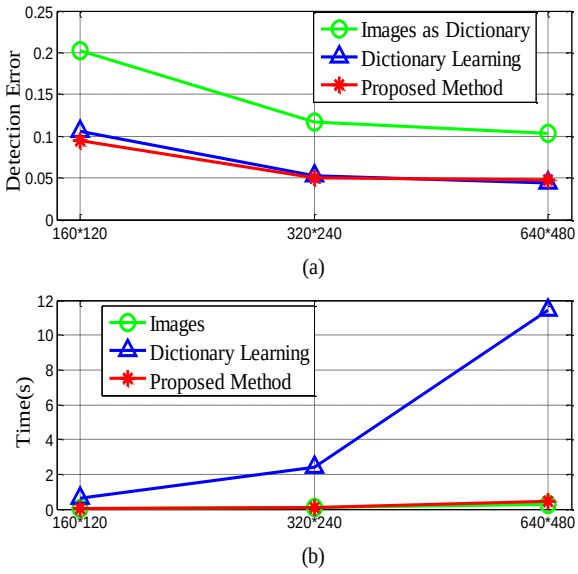


Fig. 5. Results of multimedia presentation. (a) Detection errors. (b) Running time.

Table 2. Performance comparison of different image resolutions

Resolution	160×120	320×240	640×480
Detection error	0.0944	0.0500	0.0472
Time cost (s)	0.0231	0.1060	0.4661

IV. CONCLUSION

In this paper, we focus on the laser spot detection algorithm and model it as a background subtraction problem. Further, we propose a robust dictionary construction and update algorithm based on the sparse model for laser spot detection. To test the performance of the proposed method, a large number of experiments are conducted from the perspectives of detection error and real-time performance. The experimental results confirm that the proposed method outperforms the existing methods with a lower detection error and better real-time performance when the background exhibits a high frequency of changes.

Finally, the proposed robust algorithm can also be applied to solve other practical problems, such as traffic monitoring [18] where the background switches among several configurations controlled by the status of traffic lights.

ACKNOWLEGDMENTS

This work was supported by the Natural Science Foundation of China under Grant 61405022.

REFERENCES

[1] C. Kirstein and H. Muller, "Interaction with a projection screen using a camera-tracked laser pointer," in *Proceedings of Multimedia Modeling (MMM'98)*, Lausanne, Switzerland, pp. 191-192, 1998.

[2] N. W. Kim, S. J. Lee, B. G. Lee, and J. J. Lee, "Vision based laser pointer interaction for flexible screens," in *Proceedings of the 12th International Conference on Human-Computer Interaction*, Beijing, China, pp. 845-853, 2007.

[3] L. Zhang, Y. Shi, and B. Chen, "NALP: navigating assistant for large display presentation using laser pointer," in *Proceedings of the 1st International Conference on Advances in Computer-Human Interaction*, Sainte Luce, Martinique, pp. 39-44, 2008.

[4] R. B. Widodo, W. Chen, and T. Matsumaru, "Interaction using the projector screen and spot-light from a laser pointer: handling some fundamentals requirements," in *Proceedings of SICE Annual Conference (SICE)*, Akita, Japan, pp. 1392-1397, 2012.

[5] S. Shojaeipour, S. M. Haris, A. Shojaeipour, R. K. Shirvan, and M. K. Zakaria, "Robot path obstacle locator using webcam and laser emitter," *Physics Procedia*, vol. 5, pp. 187-192, 2010.

[6] Y. Minato, T. Tsujimura, and K. Izumi, "Sign-at-ease: robot navigation system operated by connoted shapes drawn with laser beam," in *Proceedings of SICE Annual Conference (SICE)*, Tokyo, Japan, pp. 2158-2163, 2011.

[7] S. Shibata, T. Yamamoto, and M. Jindai, "Human-robot interface with instruction of neck movement using laser pointer," in *Proceedings of IEEE/SICE International Symposium on System Integration (SII)*, Kyoto, Japan, pp. 1226-1231, 2011.

[8] Y. Fukuda, Y. Kurihara, K. Kobayashi, and K. Watanabe, "Development of electric wheelchair interface based on laser pointer," in *ICCAS-SICE International Joint Conference*, Fukuoka, Japan, pp. 1148-1151, 2009.

[9] N. W. Kim and H. Lee, "Developing of vision-based virtual combat simulator," in *Proceedings of International Conference on IT Convergence and Security (ICITCS)*, Macao, China, pp. 1-4, 2013.

[10] S. J. Kim, M. S. Jang, and T. Y. Kuc, "An interactive user interface for computer-based education: the laser shot system," in *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, Lugano, Switzerland, pp. 4174-4178, 2004.

[11] F. Chávez, F. Fernández, R. Alcalá, J. Alcalá-Fdez, G. Olague, and F. Herrera, "Hybrid laser pointer detection algorithm based on template matching and fuzzy rule-based systems for domotic control in real home environments," *Applied Intelligence*, vol. 36, no. 2, pp. 407-423, 2012.

[12] J. Shin, S. Kim, and S. Yi, "Development of multi-functional laser pointer mouse through image processing," in *Proceedings of International Conference on Multimedia, Computer Graphics and Broadcasting (MulGraB)*, Jeju, Korea, pp. 290-298, 2011.

[13] I. Geys and L. Van Gool, "Virtual post-its: visual label extraction, attachment, and tracking for teleconferencing," in *Proceedings of the 3rd International Conference on Computer Vision Systems (ICVS)*, Graz, Austria, pp. 121-130, 2003.

[14] T. Bouwmans, "Traditional and recent approaches in background modeling for foreground detection: an overview," *Computer Science Review*, vol. 11-12, pp. 31-66, 2014.

[15] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21-30, 2008.

[16] V. Cevher, A. Sankaranarayanan, M. F. Duarte, D. Reddy, R. G. Baraniuk, and R. Chellappa, "Compressive sensing for background subtraction," in *Proceedings of the 10th European Conference on Computer Vision (ECCV)*, Marseille, France, pp. 155-168, 2008.

[17] J. Huang, X. Huang, and D. Metaxas, "Learning with dynamic group sparsity," in *Proceedings of IEEE 12th International Conference on Computer Vision*, Kyoto, Japan, pp. 64-71, 2009.

[18] C. Zhao, X. Wang, and W. K. Cham, "Background subtraction via robust dictionary learning," *EURASIP Journal on Image and Video Processing*, vol. 2011, pp. 1-12, 2011.

[19] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *The Annals of Statistics*, vol. 32, no. 2, pp. 407-499, 2004.

Zhihua Wang
received his B.S. in Electronic and Information Engineering from Dalian Maritime University, Dalian, China, in 2012. He is currently working toward his M.Sc. in Communication and Information System at Dalian University of Technology (DUT), Dalian, China. His research interests include computer vision and human–computer interactions.

Yongri Piao

Received his B.S. in Automation Engineering from Jilin University, China, in 2003, and his M.S. and Ph.D. in Information and Communication Engineering from Pukyong National University, Republic of Korea, in 2005 and 2008, respectively. From September 2008 to December 2011, he was Research Professor at the 3D Display Research Center of Kwangwoon University. Since March 2012, he has been an Associate Professor at the School of Information and Communication Engineering, Dalian University of Technology, Dalian, China. His research interests include optical imaging and 3D display, optical and digital encryptions, 3D pattern recognition and tracking, and 2D/3D image processing. He has more than 40 publications, including 20+ peer reviewed journal articles and 20+ conference proceedings.

Minglu Jin

is a professor at the School of Information and Communication Engineering, Dalian University of Technology, Dalian, China. He received his Ph.D. and M.Sc. degrees from Beihang University, Beijing, China, his B.Eng. degree from University of Science & Technology, Hefei, China. He was a visiting scholar at the Arimoto Lab, Osaka University, Osaka, Japan from 1987 to 1988. He was Research Fellow at the Radio & Broadcasting Research Lab, Electronics Telecommunications Research Institute (ETRI), Korea from 2001 to 2004. Professor Jin's research interests are in the general areas of signal processing and communications systems. His specific current interests are cognitive radio, multiple-input and multiple-output (MIMO) radio antenna design, and wireless sensor networks.