Underwater Moving Target Detection Based on Image Enhancement

Yan Zhou[1,2](✉), Qingwu Li[1,2], and Guanying Huo[1,2]

[1] College of Internet of Things Engineering, Hohai University, Changzhou 213022, China
strangeryan@163.com, huoguanying@163.com, liqw@hhuc.edu.cn

[2] Key Laboratory of Sensor Networks and Environmental Sensing, Changzhou 213022, China

Abstract. Motion detection in underwater video scenes is very important for many underwater computer vision tasks, such as target location, recognition and tracking. However, due to the strong optical attenuation and light scattering in water, underwater images are essentially characterized by their poor visibility, especially the low contrast and distorted information. To solve these situations, underwater moving target detection algorithm based on image enhancement is presented. The algorithm improves the contrast and clarity of the target by an adaptive underwater color image enhancement, and then extracts the moving targets by using ViBe background model. Experimental results show that the proposed algorithm can effectively extract the complete moving target by overcoming the impact of underwater environment.

Keywords: Underwater moving target detection · Adaptive image enhancement · ViBe model · Background subtraction

1 Introduction

Underwater moving target detection is an important means for underwater vehicles to acquire underwater target information. Effective underwater moving target detection contributes to many scientific researches and engineering applications, such as marine biology, seabed topography, marine environment monitoring and marine exploration [1]. At present, the common detection methods can be divided into three categories: inter-frame difference, optical flow and background subtraction [2]. Inter-frame difference method [3], which extracts the moving targets by the difference of several adjacent frames, is real-time and simple. However, generally, moving targets extracted are not complete and have the void phenomenon. Optical flow method, which detects moving targets by using its optical flow characteristics over time, has high computational complexity and is sensitive to illumination variation [4].

Background subtraction methods construct a model for the background and compare the background model with the current frame so as to detect the regions where a significant difference occurs. The Gaussian Mixture Model is one of the most popular parametric background subtraction methods [5]. It can handle the multi-modal

F. Cong et al. (Eds.): ISNN 2017, Part II, LNCS 10262, pp. 427–436, 2017.
DOI: 10.1007/978-3-319-59081-3_50

appearance of the background under dynamic environments. However, the parameter estimation of the model may become difficult for noisy images. ViBe (for 'Visual Background Extractor') is a samples-based background subtraction method. Due to the use of memoryless update strategy, spatial information propagation method, and instantaneous initialization technique, it shows an outstanding detection rate and robustness to noise [6, 7].

However, due to the serious underwater interference and dynamic change of underwater scenes, it becomes very difficult to accurately extract underwater moving targets. The strong optical attenuation and light scattering caused by the water medium and suspended particles will obviously reduce the contrast between the target and the background. These low quality video image data seriously hamper the underwater computer vision tasks. In order to solve the problems in underwater moving target detection, this paper proposes an underwater moving target detection algorithm based on image enhancement [8]. An adaptive underwater color video image enhancement algorithm is presented to improve the contrast and clarity of the target and inhibit inhomogeneous illumination. Then, the moving targets are extracted by using ViBe background model. The proposed method has high dynamical adaptability in the underwater target extraction task and strong robustness to the underwater environment. The experiment results prove its efficiency in target detection under the complex underwater optical environments.

Section 2 describes our new underwater moving target detection algorithm. Experimental results are detailed in Sect. 3. Section 4 concludes the paper.

2 Underwater Moving Target Detection Based on Image Enhancement

Underwater environment is complex and dynamic, in which there are plenty of disturbances, e.g. wave, illumination changes, light absorption and scattering. Underwater video scenes are notorious for poor visibility, low contrast, edge-blurring and being full of noise. Therefore, motion detection in underwater video scenes is more difficult than that in air. It is necessary to improve the visual quality of underwater images for subsequent accurate motion detection. So we propose an underwater moving target detection algorithm based on image enhancement. Firstly, an adaptive underwater video image enhancement algorithm inspired by the human visual system (HVS) is used to suppress noise and improve the edge sharpness. Then, underwater moving targets are detected from background model by using ViBe background subtraction algorithm. Experimental results show that the proposed algorithm can extract underwater moving targets accurately and completely.

2.1 Algorithm Flow

The algorithm flow of underwater moving target detection based on image enhancement is shown in Fig. 1.

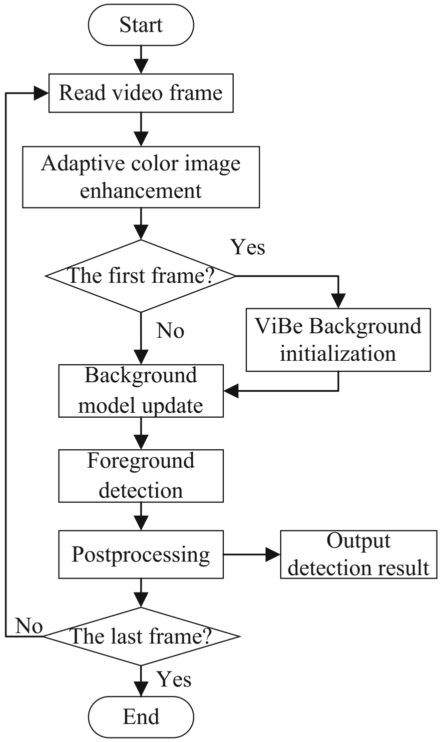


Fig. 1. Flow chart of underwater moving target detection

2.2 Adaptive Underwater Video Image Enhancement

In this paper, an adaptive underwater video image enhancement algorithm is proposed to improve the contrast and clarity of the target. Firstly, the color video image is converted from RGB to HSV color space. Secondly, the multiscale retinex (MSR) approach is used in nonsubsampled contourlet transform (NSCT) domain of V channel in order to eliminate non-uniform illumination, and threshold denoising method is adopted to suppress noise. Thirdly, the luminance masking (LM) and contrast masking (CM) characteristics of the HVS are integrated into NSCT to yield the HVS-based NSCT contrast. Subsequently, a nonlinear mapping function and a nonlinear gain function are designed to manipulate the HVS-based NSCT contrast coefficients and the NSCT lowpass subband coefficients respectively and automatically. Lastly, the enhanced V channel image can be reconstructed from NSCT coefficients, and the enhanced color image is obtained by the conversion from HSV to RGB color space. This adaptive color image enhancement algorithm, which is free of parameters adjusting, can effectively emphasize weak edges, suppress noise, remove uneven illumination and increase the identifiable characteristic information of the target. It will help to improve the accuracy of the subsequent target detection.

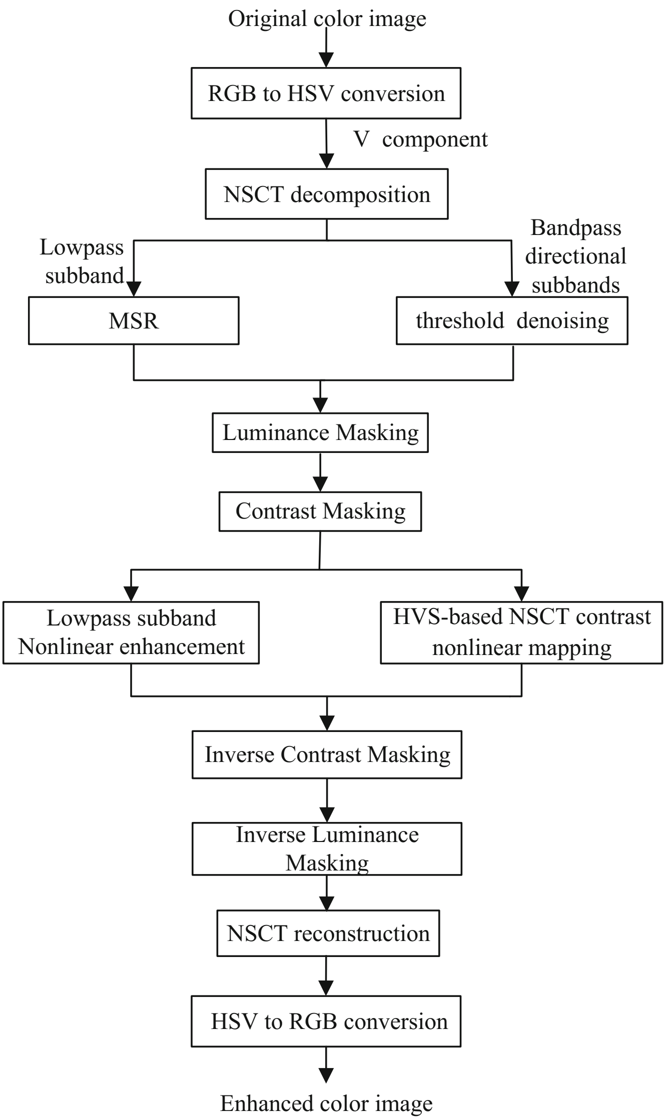


Fig. 2. Block diagram of adaptive underwater video image enhancement algorithm

The block diagram of the adaptive underwater video image enhancement algorithm is shown in Fig. 2. The NSCT is one kind of multiscale and multidirectional geometric transform, so it is able to effectively capture geometry and directional information of images. Furthermore, since each pixel of the NSCT subbands corresponds to that of the original image in spatial domain, we can collect the geometrical information pixel by pixel from the NSCT coefficients. We enhance the color video image in HSV color space by RGB to HSV conversion. The H and S component are kept unchanged, and only V component is handled. After NSCT decomposition on V channel, the lowpass subband, which is nearly noiseless, includes overall contrast information. While bandpass directional subbands contain not only edges but also noise.

Taking into account of the presence of non-uniform illumination in the underwater image, the NSCT lowpass subband is manipulated with MSR algorithm. In this stage, dynamic range is properly compressed to eliminate shadows and uneven illumination.

Because edges correspond to the large NSCT coefficients and noise corresponds to the small NSCT coefficients in bandpass directional subbands, noise can be effectively suppressed by thresholding. Thresholds for each bandpass directional subband can be chosen according to:

$$T_{s,d} = k\sigma\sqrt{\tilde{\sigma}_{s,d}} \tag{1}$$

The noise standard deviation σ of the original image is estimated by using the robust median operator, i.e.,

$$\sigma = median(abs(C))/0.6745 \tag{2}$$

where C refers to the NSCT coefficients in the finest scale. $\tilde{\sigma}^2_{s,d}$ refers to the approximate value of the individual variances at the directional subband indexed by scale s and direction d, which is calculated by using Monte-Carlo simulations.

Subsequently, HVS-based masking model in NSCT domain is constructed to yield the HVS-based NSCT contrast. To obtain HVS-based NSCT contrast, two steps are to be conducted. Firstly, the LM contrast in NSCT domain is measured by

$$C_{LM(s,d)} = \frac{y_{(s,d)}}{\left|y_{(s,0)}\right| + c} \tag{3}$$

where $y_{(s,d)}$ is the original NSCT bandpass directional subband indexed by scale s and direction d. $y_{(s,0)}$ is the original NSCT lowpass subband at the sth scale. In an N level NSCT decomposition of an image, $s = N$ denotes the scale after performing NSCT decomposition procedure one time (i.e., the finest scale), and $s = 1$ denotes the scale after performing NSCT decomposition procedure N times (i.e., the coarsest scale). c is a small constant to avoid dividing by 0. $C_{LM(s,d)}$ is the output of LM contrast indexed by scale s and direction d. Secondly, the LM contrast is masked with Contrast Masking to yield the LCM contrast, which is the HVS-based NSCT contrast. The multiscale LCM contrast, which is a function of the LM contrast, is defined as

$$C_{LCM(s,d)} = \frac{C_{LM(s,d)}}{\left|C_{LM(s-1,d)}\right|^{0.62} + c} \tag{4}$$

where c is a small constant to avoid dividing by 0.

We also propose a nonlinear mapping function to modify the HVS-based NSCT contrast coefficients at each scale and direction independently and automatically so as to achieve multiscale contrast enhancement. The proposed nonlinear mapping function is given by:

$$\hat{C}_{LCM(s,d)} = s \cdot max(\left|C_{LCM(s,d)}\right|) \cdot sign(C_{LCM(s,d)}) \cdot \left[\sin(\frac{\pi}{2} \cdot \frac{\left|C_{LCM(s,d)}\right|}{max(\left|C_{LCM(s,d)}\right|)})\right]^{\sqrt{p}} \tag{5}$$

where

$$p = \frac{\log(\frac{mean(\left|C_{LCM(s,d)}\right|)}{max(\left|C_{LCM(s,d)}\right|)})}{\log[\sin(\frac{\pi}{2}\frac{mean(\left|C_{LCM(s,d)}\right|)}{max(\left|C_{LCM(s,d)}\right|)})]} \tag{6}$$

$C_{LCM(s,d)}$ is the original HVS-based NSCT contrast coefficient at the subband indexed by scale s and direction d. $\hat{C}_{LCM(s,d)}$ is the modified HVS-based NSCT contrast coefficient. $max\left(\left|C_{LCM(s,d)}\right|\right)$ denotes the maximum absolute contrast coefficient amplitude at the subband indexed by scale s and direction d. $mean\left(\left|C_{LCM(s,d)}\right|\right)$ denotes the mean value of absolute contrast coefficient amplitude at the subband indexed by scale s and direction d. Given an N scale NSCT decomposition of an image, $C_{LCM(s,d)}$ includes N-1 scales (i.e., $1 < s \leq N$), according to Eq. (4).

This nonlinear mapping function can well enhance the low-contrast areas, and also avoid over-enhancement of the high-contrast areas simultaneously [8].

Furthermore, the global dynamic range of the image can be adjusted by using one nonlinear gain function in the lowpass subband at the coarsest scale of NSCT decomposition. The nonlinear gain function is defined as follows:

$$\hat{y}_{(1,0)} = max(\left|y_{(1,0)}\right|) \cdot sign(y_{(1,0)}) \cdot \left[\sin(\frac{\pi}{2} \cdot \frac{\left|y_{(1,0)}\right|}{max(\left|y_{(1,0)}\right|)})\right]^{q} \tag{7}$$

where

$$q = \frac{\log(\frac{mean(\left|y_{(1,0)}\right|)}{max(\left|y_{(1,0)}\right|)})}{\log[\sin(\frac{\pi}{2}\frac{mean(\left|y_{(1,0)}\right|)}{max(\left|y_{(1,0)}\right|)})]} \tag{8}$$

$y_{(1,0)}$ is the NSCT lowpass subband coefficient at the first scale (i.e., the coarsest scale). $\hat{y}_{(1,0)}$ is the modified NSCT lowpass subband coefficient at the first scale.

The modified LM contrast can be calculated from the modified LCM contrast by Inverse Contrast Masking:

$$\hat{C}_{LM(s,d)} = \hat{C}_{LCM(s,d)} * (\left|\hat{C}_{LM(s-1,d)}\right|^{0.62} + c) \tag{9}$$

The modified bandpass directional subband coefficients of NSCT are calculated from the modified LM contrast by Inverse Luminance Masking:

$$\hat{y}_{(s,d)} = \hat{C}_{LM(s,d)} * \left(\left|\hat{y}_{(s,0)}\right| + c\right) \tag{10}$$

Finally, the enhanced V channel image can be reconstructed from NSCT coefficients, and the enhanced color image is obtained by the conversion from HSV to RGB color space.

2.3 ViBe Model

ViBe is one kind of powerful samples-based background subtraction method. The background model is initialized from only single frame by using a random selection policy. Background model is updated dynamically by using a memoryless update strategy and a neighborhood propagation mechanism. It mainly includes three parts: background modeling, foreground target detection and model updating.

Background Modeling and Initialization. ViBe builds a model for each background pixel with a set of samples instead of with an explicit pixel model. Denote $f(x)$ by the pixel value located at x in the image in a given Euclidean color space, and f_i by a background sample value with an index i. The background pixel located at x is modeled by a collection of N background sample values as follows:

$$M(x) = \{f_1, f_2, \ldots, f_N\} \tag{11}$$

Considering that adjacent pixels share a similar temporal distribution, ViBe initializes the background model from a single frame. Each pixel model contains N values randomly extracted from the spatial neighborhood of each pixel in the first frame. Assume that $t = 0$ denotes the first frame and $N_G(x)$ is a spatial neighborhood of a pixel located at x, thus the model $M^0(x)$ is as follows:

$$M^0(x) = \left\{f^0(y|y \in N_G(x))\right\} \tag{12}$$

Foreground Target Detection. If we consider the background subtraction as a classification problem, we want to classify a new pixel as a background or foreground pixel with respect to its neighborhood. To classify a pixel value $f(x)$ according to its corresponding model $M(x)$, we compare it to the closest values within the set of samples by defining a sphere $S_R(f(x))$ of radius R centered on $f(x)$. The pixel value $f(x)$ is then classified as background if the cardinality, denoted $\#T(x)$, of the set intersection between the sphere and the model sample set is greater than or equal to a given threshold $\#\min$.

$$\#T(x) = \#\{S_R(f(x)) \cap (f_1, f_2, \ldots, f_N)\} \tag{13}$$

Assume that $t = k$ denotes the k frame and that $f^k(x)$ is the value of the pixel located at x in the k frame, with its corresponding model $M^{k-1}(x)$. To classify a pixel value as foreground or not as follows:

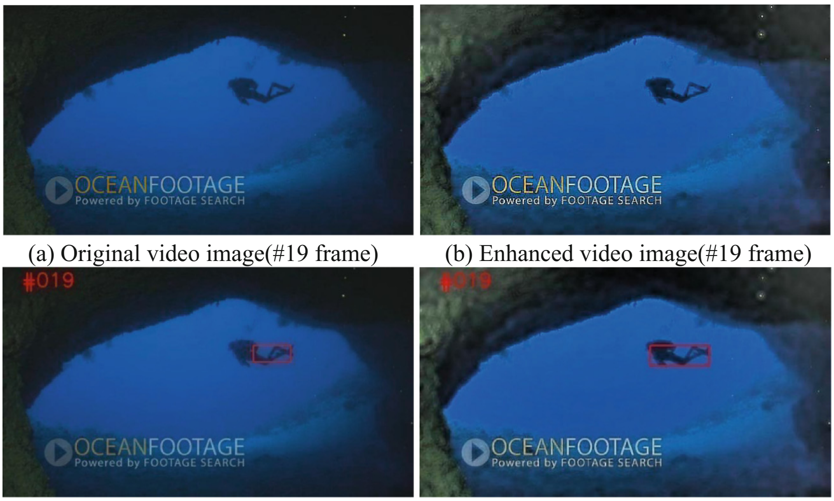
$$f^k(x) = \begin{cases} background, & \#T(x) \geq \#\min \text{ and } \\ & \{|f^k(x) - f_i^{k-1}(x)| \leq R, i = 1, \ldots, N\}, \\ foreground, & \text{else} \end{cases} \tag{14}$$

Update Background Model. ViBe can ensure a smooth decaying lifespan for the samples in the background pixel models by using a memoryless update strategy. The random time subsampling method is used to expand the time windows covered by the background pixel models. A time subsampling weighting factor φ is designed to control the probability of background pixel model updating. A background pixel value has one chance in φ to be selected to update its pixel model and also has the same chance to be selected to update its neighboring pixel model. By using spatial consistency in background samples propagation, a spatial diffusion of information about the background evolution is achieved. Therefore, this background model can adapt to structural evolutions and varying illumination.

3 Experimental Results and Analysis

All the tests are implemented at MATLAB R2011b platform on a PC with 2.4-GHz Intel(R) Xeon(R) CPU and 8-GB RAM. ViBe parameters are set as follows: $N = 20$, $R = 20$, $\#\min = 2$, $\varphi = 16$.

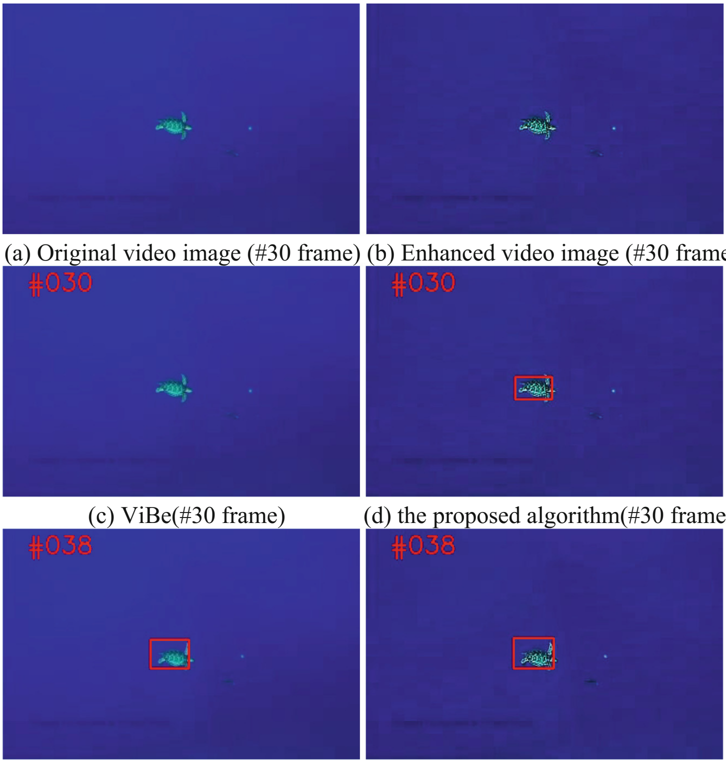
Simulation platform detects moving targets from video scenes by the underwater still camera. Figure 3 is the experimental results of video 'Diver'. The 19th frame in the original video is shown in Fig. 3(a), and its corresponding enhanced video image is shown in Fig. 3(b). It can be seen that the enhanced image is clearer and the contrast



(a) Original video image(#19 frame) (b) Enhanced video image(#19 frame)

(c) ViBe (d) the proposed algorithm

Fig. 3. Moving target detection results of video 'Diver'



(a) Original video image (#30 frame) (b) Enhanced video image (#30 frame)

(c) ViBe(#30 frame) (d) the proposed algorithm(#30 frame)

(e) ViBe(#38 frame) (f) the proposed algorithm (#38 frame)

Fig. 4. Moving target detection results of video 'Turtle'

between the target and the background is higher. Figure 3(c) and (d) are detection results of ViBe and the proposed algorithm, respectively. The visual results show that the proposed algorithm can extract the complete moving diver accurately.

Figure 4 is the experimental results of video 'Turtle'. The 30th frame in the original video is shown in Fig. 4(a), and its corresponding enhanced video image is shown in Fig. 4(b). The turtle in the enhanced image is clearer than that in the original image. Figure 4(c) and (e) are the detection results by using ViBe, and Fig. 4(d) and (f) are the results by using the proposed algorithm. The proposed algorithm can detect the target accurately in 30th frame, but ViBe cannot. ViBe algorithm finds the target until 38th frame, and the detection window is slightly offset.

4 Conclusions

In this paper, we present an underwater moving target detection algorithm based on image enhancement. The proposed algorithm achieves the adaptive underwater color video image enhancement so as to improve the contrast and clarity of the target and inhibit inhomogeneous illumination, and then efficiently extracts underwater moving targets by ViBe model. The experiments show that the proposed algorithm can accurately detect the moving targets under the complex underwater environment.

Acknowledgment. This work was supported by the National Natural Science Foundation of China (No. 41306089), the Natural Science Foundation of Jiangsu Province (No. BK20130240) and Changzhou Science and Technology Program (No. CJ20160055).

References

1. Shen, J., Fan, T., Tang, M., Zhang, Q., Sun, Z., Huang, F.: A biological hierarchical model based underwater moving object detection. Comput. Math. Methods Med. **2014**, 1–8 (2014)
2. Lei, F., Huang, W., Zhang, Z.: Underwater moving object detection based on codebook model. In: Fifth International Conference on Computational and Information Sciences, pp. 1319–1322 (2013)
3. Moscheni, F., Bhattacharjee, S., Kunt, M.: Spatialtemporal segmentation based on region merging. IEEE Trans. Pattern Anal. Mach. Intell. **20**(9), 897–915 (1998)
4. Negahdaripour, S.: Revised definition of optical flow: integration of radiometric and geometric cues for dynamic scene analysis. IEEE Trans. Pattern Anal. Mach. Intell. **20**(9), 961–979 (1998)
5. Stauffer, C., Grimson, W.E.L.: Learning patterns of activity using real-time tracking. IEEE Trans. Pattern Anal. Mach. Intell. **22**(8), 747–757 (2000)
6. Barnich, O., Droogenbroeck, M.V.: ViBE: a powerful random technique to estimate the background in video sequences. In: IEEE International Conference on Acoustics, Speech and Signal Processing -Proceedings, pp. 945–948. Taipei, Taiwan (2009)
7. Barnich, O., Droogenbroeck, M.V.: ViBe: a universal background subtraction algorithm for video sequences. IEEE Trans. Image Process. **20**(6), 1709–1724 (2011)
8. Zhou, Y., Li, Q., Huo, G.: Human visual system based automatic underwater image enhancement in NSCT domain. KSII Trans. Internet Inf. Syst. **10**(2), 837–856 (2016)