




Article

Element-Weighted Neutrosophic Correlation Coefficient and Its Application in Improving CAMShift Tracker in RGBD Video

Keli Hu ^{1,*} , En Fan ¹, Jun Ye ² , Jiatian Pi ³, Liping Zhao ¹ and Shigen Shen ¹ 

¹ Department of Computer Science and Engineering, Shaoxing University, Shaoxing 312000, China; efan@usx.edu.cn (E.F.); zhaoliping_jian@126.com (L.Z.); shigens@126.com (S.S.)

² Department of Electrical and Information Engineering, Shaoxing University, Shaoxing 312000, China; yejun@usx.edu.cn

³ College of Computer and Information Science, Chongqing Normal University, Chongqing 400700, China; pijiatian@126.com

* Correspondence: ancimoon@gmail.com

Received: 11 April 2018; Accepted: 15 May 2018; Published: 18 May 2018



Abstract: Neutrosophic set (NS) is a new branch of philosophy to deal with the origin, nature, and scope of neutralities. Many kinds of correlation coefficients and similarity measures have been proposed in neutrosophic domain. In this work, by considering that there may exist different contributions for the neutrosophic elements of T (Truth), I (Indeterminacy), and F (Falsity), a method of element-weighted neutrosophic correlation coefficient is proposed, and it is applied for improving the CAMShift tracker in RGBD (RGB-Depth) video. The concept of object seeds is proposed, and it is employed for extracting object region and calculating the depth back-projection. Each candidate seed is represented in the single-valued neutrosophic set (SVNS) domain via three membership functions, T , I , and F . Then the element-weighted neutrosophic correlation coefficient is applied for selecting robust object seeds by fusing three kinds of criteria. Moreover, the proposed correlation coefficient is applied for estimating a robust back-projection by fusing the information in both color and depth domains. Finally, for the scale adaption problem, two alternatives in the neutrosophic domain are proposed, and the corresponding correlation coefficient between the proposed alternative and the ideal one is employed for the identification of the scale. When considering challenging factors like fast motion, blur, illumination variation, deformation, and camera jitter, the experimental results revealed that the improved CAMShift tracker performs well.

Keywords: tracking; CAMShift; neutrosophic set; single-valued neutrosophic set; correlation coefficient; element-weighted neutrosophic correlation coefficient

1. Introduction

A neutrosophic set (NS) [1] is suitable for dealing with problems with indeterminate information. A neutrosophic set is always characterized independently via three membership functions, T , I , and F . These functions in the universal set X are real standard or nonstandard subsets of $[-0, 1^+]$. It is difficult to introduce the NS theory into science and engineering areas when considering the non-standard unit interval. The single-valued neutrosophic set (SVNS) [2] was proposed to handle such a problem. The membership functions are restrained to the normal standard real unit interval $[0, 1]$.

Until now, NS has been successfully applied in a lot of fields [3] such as medical diagnosis [4], image segmentation [5–9], skeleton extraction [10], and object tracking [11,12]. Several similarity measurements or multicriteria decision-making methods [13–20] were employed to handle the neutrosophic problems. Decision-making can be regarded as a problem-solving activity terminating

in a solution deemed to be satisfactory. It has been applied for residential house garage location selection [18], element and material selection [19], and sustainable market valuation of buildings [20]. For the application of image segmentation, several criteria in the NS domain were usually proposed for calculating a specific neutrosophic image [5–9]. The correlation coefficient between SVN_Ss [17] was applied for calculating a neutrosophic score-based image [9], and a robust threshold was estimated by employing the OTSU's method [9]. In [11], two criteria were proposed in both color and depth domain. The information fusion problem was converted into a multicriteria decision-making issue, and the single-valued neutrosophic cross-entropy was employed to tackle this problem [11]. For the neutrosophic theory-based MeanShift tracker [12], by taking the consideration of the background information and appearance changes between frames, two kinds of criteria were considered, the object feature similarity and the background feature similarity. The SVN_S correlation coefficient [17] was applied for calculating the weighted histogram, and then the histogram was finally used to enhance the traditional MeanShift tracker. Besides the fields mentioned above, the NS theory was also introduced into clustering algorithms such as *c*-means [21]. While NS-based correlation coefficients have been widely used for solving some engineering issues, the weights of the three membership functions are always treated equally. However, for the elements of *T*, *I*, and *F*, different contributions should be taken into consideration of the results of decision-making. In this work, a method of element-weighted neutrosophic correlation coefficient is proposed, and we try to utilize it for tracking a visual object in RGBD (RGB-Depth) video.

Visual object tracking is still an open issue [22–24], and its robustness still lacks the performance of vision applications like surveillance, traffic monitoring, video indexing, and auto-driving. For a tracking task, challenges like background clutter, fast motion, illumination variation, scale variation, motion blur, and camera jitter may happen during the tracking procedure. Two methods are considered to handle these challenging problems. One is utilizing robust features. The color feature is employed by the MeanShift [25] and CAMShift [26] tracker, due to its robustness when there exist challenges like deformation, blur, and rotation. Both trackers are of high efficiency [27]. However, when the surroundings have a similar color, they may easily drift from the target. To deal with such a problem, Cross-Bin metric [28], scale-invariant feature transform (SIFT) [29], and texture feature [30] were introduced into the mean shift-based tracker, and better performance was achieved. The other way is to train robust models, such as the multiple instance learning [31,32] and compressive sensing [33]-based tracking. Besides the kinds of trackers mentioned above, the local-global tracker (LGT) [34], incremental visual tracker (IVT) [35] and tracking-learning-detection (TLD) [36] also perform well.

Recently, due to the fact that the depth information can provide another dimension for tackling the object tracking problem, some RGBD-based trackers [11,37–39] have been proposed. Most algorithms are oriented to specific targets [37–39]. A few category-free RGBD trackers have been proposed. An improved CAMShift tracker was proposed by using the neutrosophic decision-making method in [11], but the indeterminate information was only considered in the information fusion phase. As mentioned above, the CAMShift tracker will perform well, with high efficiency, if a discriminative feature can be employed. In this work, based on the CAMShift framework, we focus on tackling the visual tracking problem when challenges like fast motion, blur, illumination variation, deformation, and camera jitter exist, but without serious occlusion. It is difficult to find such a tracker due to the fact that tracking an object without occlusion is still very challenging for both RGB and RGBD trackers. For the CAMShift tracker, calculating a robust back-projection is one of the most important issues for tracking a target. Indeterminate information always exists in the procedure of the CAMShift process. For instance, it exists when estimating the likelihood probability map, as well as localizing the target. In this work, we try to utilize the method of element-weighted neutrosophic correlation coefficient to handling decision-making problems when there is indeterminate information.

This work mainly exhibits four contributions. First, a method of element-weighted neutrosophic correlation coefficient is proposed. Secondly, three criteria are proposed for object seeds selection, and the corresponding membership functions, *T*, *I*, and *F*, are given. Thirdly, the proposed correlation

coefficient is applied for estimating a robust back-projection by fusing the information in both color and depth domains. Finally, for the scale adaption problem, two alternatives in the neutrosophic domain are proposed, and the corresponding correlation coefficient between the proposed alternative and the ideal one is employed for the identification of the scale.

The remainder of this paper is organized as follows: in Section 2, the element-weighted neutrosophic correlation coefficient is presented. In Section 3, main steps and basic flow of the improved CAMShift visual tracker is given first, and the details are illustrated in the following subsections. Experimental evaluations and discussions are presented in Section 4, and Section 5 is the conclusions.

2. Element-Weighted Neutrosophic Correlation Coefficient

2.1. Neutrosophic Correlation Coefficient

Let $A = \{A_1, A_2, \dots, A_m\}$ be a set of alternatives and $C = \{C_1, C_2, \dots, C_n\}$ be a set of criteria. We get

$$A_i = \left\{ \left\langle C_j, T_{C_j}(A_i), I_{C_j}(A_i), F_{C_j}(A_i) \right\rangle \mid C_j \in C \right\} \quad i = 1 \dots m, j = 1 \dots n, \quad (1)$$

where A_i is an alternative in A , $T_{C_j}(A_i)$ indicates the degree to which the alternative A_i satisfies the criterion C_j ; $I_{C_j}(A_i)$ denotes the indeterminacy degree to which the alternative A_i satisfies or does not satisfy the criterion C_j ; $F_{C_j}(A_i)$ denotes the degree to which the alternative A_i does not satisfy the criterion C_j , $T_{C_j}(A_i), I_{C_j}(A_i), F_{C_j}(A_i) \in [0, 1]$, $1 \leq T_{C_j}(A_i) + I_{C_j}(A_i) + F_{C_j}(A_i) \leq 3$, and A is a single-valued neutrosophic set.

The correlation coefficient under single-valued neutrosophic environment between two alternatives A_i and A_j is defined as [17]:

$$S_{C_k}(A_i, A_j) = \frac{T_{C_k}(A_i)T_{C_k}(A_j) + I_{C_k}(A_i)I_{C_k}(A_j) + F_{C_k}(A_i)F_{C_k}(A_j)}{\sqrt{T_{C_k}^2(A_i) + I_{C_k}^2(A_i) + F_{C_k}^2(A_i)}\sqrt{T_{C_k}^2(A_j) + I_{C_k}^2(A_j) + F_{C_k}^2(A_j)}}. \quad (2)$$

Considering the contribution of each criterion, the weighted correlation coefficient between A_i and A_j is defined by

$$W(A_i, A_j) = \sum_{k=1}^n w_k \frac{T_{C_k}(A_i)T_{C_k}(A_j) + I_{C_k}(A_i)I_{C_k}(A_j) + F_{C_k}(A_i)F_{C_k}(A_j)}{\sqrt{T_{C_k}^2(A_i) + I_{C_k}^2(A_i) + F_{C_k}^2(A_i)}\sqrt{T_{C_k}^2(A_j) + I_{C_k}^2(A_j) + F_{C_k}^2(A_j)}}, \quad (3)$$

where $w_k \in [0, 1]$ is the weight of each criterion C_k and $\sum_{j=1}^n w_j = 1$.

Assume A^* is the ideal alternative, $A^* = \left\{ \left\langle C_j, T_{C_j}(A^*), I_{C_j}(A^*), F_{C_j}(A^*) \right\rangle \mid C_j \in C \right\}$. Then the weighted correlation coefficient between any alternative A_i and the ideal alternative A^* can be calculated by

$$W(A_i, A^*) = \sum_{k=1}^n w_k \frac{T_{C_k}(A_i)T_{C_k}(A^*) + I_{C_k}(A_i)I_{C_k}(A^*) + F_{C_k}(A_i)F_{C_k}(A^*)}{\sqrt{T_{C_k}^2(A_i) + I_{C_k}^2(A_i) + F_{C_k}^2(A_i)}\sqrt{T_{C_k}^2(A^*) + I_{C_k}^2(A^*) + F_{C_k}^2(A^*)}}. \quad (4)$$

2.2. Element-Weighted Neutrosophic Correlation Coefficient

As seen in Equation (3), only the contribution of each criterion is considered. However, in some engineering applications, it is better to provide specific weights for the T , I , and F elements for an alternative A_i when considering the criteria C_j . The element-weighted correlation coefficient between A_i and A_j is defined by

$$S_e(A_i, A_j) = \frac{\alpha T_{C_k}(A_i)T_{C_k}(A_j) + \beta I_{C_k}(A_i)I_{C_k}(A_j) + \gamma F_{C_k}(A_i)F_{C_k}(A_j)}{\sqrt{\alpha T_{C_k}^2(A_i) + \beta I_{C_k}^2(A_i) + \gamma F_{C_k}^2(A_i)}\sqrt{\alpha T_{C_k}^2(A_j) + \beta I_{C_k}^2(A_j) + \gamma F_{C_k}^2(A_j)}}, \quad (5)$$

where $\alpha, \beta, \gamma \in [0,1], \alpha + \beta + \gamma = 1$. Then $S_e(A_i, A_j)$ is actually the cosine value between the vector $(\sqrt{\alpha}T_{C_k}(A_i), \sqrt{\beta}I_{C_k}(A_i), \sqrt{\gamma}F_{C_k}(A_i))$ and $(\sqrt{\alpha}T_{C_k}(A_j), \sqrt{\beta}I_{C_k}(A_j), \sqrt{\gamma}F_{C_k}(A_j))$. Hence, it is easy to find that S_e satisfies the following properties:

- (SP1) $0 \leq S_e(A_i, A_j) \leq 1$;
- (SP2) $S_e(A_i, A_j) = S_e(A_j, A_i)$;
- (SP3) $S_e(A_i, A_j) = 1$ if $A_i = A_j$;

Then the element-criterion-weighted correlation coefficient between A_i and the ideal alternative A^* is defined as

$$WS_e(A_i, A^*) = \sum_{k=1}^n w_k \frac{\alpha T_{C_k}(A_i)T_{C_k}(A^*) + \beta I_{C_k}(A_i)I_{C_k}(A^*) + \gamma F_{C_k}(A_i)F_{C_k}(A^*)}{\sqrt{\alpha T_{C_k}^2(A_i) + \beta I_{C_k}^2(A_i) + \gamma F_{C_k}^2(A_i)} \sqrt{\alpha T_{C_k}^2(A^*) + \beta I_{C_k}^2(A^*) + \gamma F_{C_k}^2(A^*)}} \tag{6}$$

3. Improved CAMShift Visual Tracker Based on the Neutrosophic Theory

The algorithmic details of the proposed tracker are presented in this section.

A rectangle bounding box [22,23] is always employed for the representation of the target location when dealing with the visual tracking problem. For a visual tracker, the critical task is to estimate the corresponding bounding box in each frame.

Algorithm 1 illustrates the basic flow of the proposed algorithm, as seen in Table 1. Details of each main step are given in the following subsections.

Table 1. Basic flow of the NeutRGBDs tracker.

Algorithm 1 NeutRGBDs
Initialization
Input: 1st video frame in the RGBD domain
(1) Select an object on the image plane
(2) Extract object seeds using both color and depth information
(3) Extract object region using object seeds and the information of the depth segmentation
(4) Calculate the corresponding color and depth histograms as target model
Tracking
Input: (t+1)-th video frame in the RGBD domain
(1) Calculate back-projections in both color (P_C) and depth domain (P_D)
(2) Calculate fused back-projection (P) using neutrosophic theory
(3) Calculate the bounding box of the target in the CAMShift framework
(4) Extract object region and update object model and seeds
(5) Modify the scale of the bounding box in neutrosophic domain
Output: Tracking location

3.1. Selecting Object Seeds

Assuming the tracker can calculate an adequate bounding box, and R is the extracted object region, then it is reasonable that we assume pixels located in R and the bounding box with smaller depth value are more likely to be a part of the target.

To select robust object seeds, the pixels located in R and the bounding box are first sorted by the descending order of the depth value. Then several pixels are selected as the candidate object seeds set S by sampling the sorted pixels with a fixed sampling space. Considering there may exist background regions in the corresponding area, only the top 50 percent pixels in the sorted pixel set are sampled.

For the proposition of depth value is a critical criterion for judging the object seed, $T_D, I_D,$ and F_D represent the probabilities when a proposition is true, indeterminate and false degrees, respectively. Then we can give the definitions:

$$T_D = \frac{r_{\max} - r_{si}}{r_{\max}} \tag{7}$$

$$I_D = \text{var}(D(\mathbf{ROI}_{si})) \tag{8}$$

$$F_D = 1 - T_D, \tag{9}$$

where r_{\max} is the lowest rank in \mathbf{S} , and r_{si} is the rank of the i -th candidate seed \mathbf{S}_i ; \mathbf{ROI}_{si} is a set of pixels that is located in the square area centered at \mathbf{S}_i , and the length of the sides of the square should be set to an odd value; $D(\mathbf{ROI}_{si})$ is the depth value set corresponding to \mathbf{ROI}_{si} , and $\text{var}(x)$ is the standard variance function.

As there may be other objects with a similar depth to the tracked target, the color similarity criterion is considered in this work to conquer such a problem. The corresponding three membership functions T_C, I_C and F_C are defined as follows:

$$T_C = \mathbf{P}_c(\mathbf{S}_i) \tag{10}$$

$$I_C = \text{var}(\mathbf{P}_c(\mathbf{ROI}_{si})) \tag{11}$$

$$F_C = 1 - T_C, \tag{12}$$

where \mathbf{P}_c is the back-projection calculated in color domain; $\mathbf{P}_c(\mathbf{S}_i)$ is the value of the back-projection located at \mathbf{S}_i . $\mathbf{P}_c(\mathbf{ROI}_{si})$ is the value set corresponding to \mathbf{ROI}_{si} in \mathbf{P}_c .

Besides the consideration of the depth and color feature separately, the fused color-depth information is also taken into consideration. The corresponding three membership functions T_{DC}, I_{DC} and F_{DC} for the fused color-depth criterion are defined as follows:

$$T_{DC} = \mathbf{P}(\mathbf{S}_i) \tag{13}$$

$$I_{DC} = \text{var}(\mathbf{P}(\mathbf{ROI}_{si})) \tag{14}$$

$$F_{DC} = 1 - T_{DC}, \tag{15}$$

where \mathbf{P} is the fused back-projection; $\mathbf{P}(\mathbf{S}_i)$ is the value of the back-projection located at \mathbf{S}_i ; $\mathbf{P}(\mathbf{ROI}_{si})$ is the value set corresponding to \mathbf{ROI}_{si} in \mathbf{P} .

By substituting the corresponding T, I , and F under the proposed three criteria into Equation (6), the probability of the reliability for the seed \mathbf{S}_i can be calculated by

$$\begin{aligned} pr_{si} &= w_D pr_{Dsi} + w_C pr_{Csi} + w_{DC} pr_{DCsi} \\ &= w_D \frac{\alpha T_D T_D(A^*) + \beta I_D I_D(A^*) + \gamma F_D F_D(A^*)}{\sqrt{\alpha T_D^2 + \beta I_D^2 + \gamma F_D^2} \sqrt{\alpha T_D^2(A^*) + \beta I_D^2(A^*) + \gamma F_D^2(A^*)}} \\ &\quad + w_C \frac{\alpha T_C T_C(A^*) + \beta I_C I_C(A^*) + \gamma F_C F_C(A^*)}{\sqrt{\alpha T_C^2 + \beta I_C^2 + \gamma F_C^2} \sqrt{\alpha T_C^2(A^*) + \beta I_C^2(A^*) + \gamma F_C^2(A^*)}}, \tag{16} \\ &\quad + w_{DC} \frac{\alpha T_{DC} T_{DC}(A^*) + \beta I_{DC} I_{DC}(A^*) + \gamma F_{DC} F_{DC}(A^*)}{\sqrt{\alpha T_{DC}^2 + \beta I_{DC}^2 + \gamma F_{DC}^2} \sqrt{\alpha T_{DC}^2(A^*) + \beta I_{DC}^2(A^*) + \gamma F_{DC}^2(A^*)}} \end{aligned}$$

where $w_D, w_C, w_{DC} \in [0,1]$ are the corresponding weights of criteria and $w_D + w_C + w_{DC} = 1$. Assuming the ideal alternative under all the three criteria is the same as $A^* = \langle 1, 0, 0 \rangle$, then Equation (16) can be simplified to

$$pr_{si} = \sqrt{\alpha} \left(\begin{aligned} &w_D \frac{T_D}{\sqrt{\alpha T_D^2 + \beta I_D^2 + \gamma F_D^2}} + w_C \frac{T_C}{\sqrt{\alpha T_C^2 + \beta I_C^2 + \gamma F_C^2}} \\ &+ w_{DC} \frac{T_{DC}}{\sqrt{\alpha T_{DC}^2 + \beta I_{DC}^2 + \gamma F_{DC}^2}} \end{aligned} \right). \tag{17}$$

After calculating all the probability of the reliability for the seed in \mathbf{S} , the first N seeds sorted by pr_{si} are selected as object seeds set \mathbf{OS} .

3.2. Extracting Object

The critical task for extracting object is to determine the extracted object region \mathbf{R} .

For each frame, the image is segmented into regions in the depth domain. In this work, the fast depth segmentation method [40] is employed for segmentation. Suppose there exist M regions in the t -th frame, \mathbf{C}_i represents the i -th region, and DC_i is the depth value at the location of the centroid of \mathbf{C}_i .

Suppose \mathbf{OS}_i is the i -th object seed in the previous frame and \mathbf{B} is the pixel set located in the area of the bounding box calculated by the tracker in the current frame. The extracted region \mathbf{R} is defined as:

$$\mathbf{R} = \{ \cup(\mathbf{C}_k \cap \mathbf{B}) | \min(|DC_k - DOS_i|) < T, i = 1, 2, \dots N \}, \tag{18}$$

where DOS_i is the depth value of \mathbf{OS}_i ; T is a threshold value; \cup and \cap corresponds to the set operation. All the regions \mathbf{C}_k belonging to \mathbf{R} constructs the candidate object region \mathbf{RC} .

3.3. Calculating the Fused Back-Projection

Back-projection is a probability distribution map with the same size as the input image. Each pixel value of the back-projection demonstrates the likelihood of the corresponding pixel belonging to the tracked object. Considering that using the discriminative feature is one of the most important factors for the visual tracker, both the color and depth information are employed for calculating the fused back-projection.

In [11], two criteria are proposed in both feature domain. For the near region similarity criterion, the corresponding neutrosophic elements are defined as follows:

$$T_{Cn} = \sum_{u=1}^{m^c} \sqrt{\hat{q}_u^c \hat{p}_u^c} \tag{19}$$

$$I_{Cn} = \sum_{u=1}^{m^c} \sqrt{\hat{q}_u^c \hat{p}_u^{c'}} \tag{20}$$

$$F_{Cn} = 1 - T_{Cn}, \tag{21}$$

where \hat{q}_u^c is the u -th bin of the feature vector for the target in the color domain, it is firstly calculated in the first frame by

$$\hat{q}_u^c = C \sum_{i=1}^n \delta[b^c(\mathbf{x}_i) - u], \tag{22}$$

where $\{\mathbf{x}_i\}_{i=1 \dots n}$ are the pixels located in the extracted object region \mathbf{R} ; b^c is the transformation function in color domain, and the function $b^c: R^2 \rightarrow \{1 \dots m^c\}$ associates to the pixel at location \mathbf{x}_i the index $b(\mathbf{x}_i)$ of its bin in the quantized feature space; δ is the Kronecker delta function; C is the normalization constant derived by imposing the condition $\sum_{u=1}^{m^c} \hat{q}_u^c = 1$; \hat{p}_u^c is the u -th bin of the feature vector corresponds to the extracted object region \mathbf{R} in the current frame, and $\hat{p}_u^{c'}$ corresponds to the feature vector in the annular region near the target, as defined in [11].

Similarly, for the near region similarity criterion, the corresponding neutrosophic elements in the depth domain can be computed by

$$T_{Dn} = \sum_{u=1}^{m^d} \sqrt{\hat{q}_u^d \hat{p}_u^d}, I_{Dn} = \sum_{u=1}^{m^d} \sqrt{\hat{q}_u^d \hat{p}_u^{d'}}, F_{Dn} = 1 - T_{Dn}, \tag{23}$$

where \hat{q}_u^d , \hat{p}_u^d and $\hat{p}_u^{d'}$ are the corresponding feature vector in the depth domain, similar to the calculation of \hat{q}_u^c , \hat{p}_u^c and $\hat{p}_u^{c'}$.

By applying the far region similarity criterion [11], the related functions are presented as follows:

$$\begin{aligned} T_{Cf} &= \sum_{u=1}^{m^c} \sqrt{\hat{q}_u^c \hat{p}_u^c}, \quad I_{Cf} = \sum_{u=1}^{m^c} \sqrt{\hat{q}_u^c \hat{p}_u^{c'}}, \quad F_{Cf} = 1 - T_{Cf} \\ T_{Df} &= \sum_{u=1}^{m^d} \sqrt{\hat{q}_u^d \hat{p}_u^d}, \quad I_{Df} = \sum_{u=1}^{m^d} \sqrt{\hat{q}_u^d \hat{p}_u^{d'}}, \quad F_{Df} = 1 - T_{Df} \end{aligned} \quad (24)$$

where T_{Cf} , I_{Cf} , and F_{Cf} correspond to the color domain; T_{Df} , I_{Df} and F_{Df} are the functions in the depth domain. As defined in [11], $\hat{p}_u^{c'}$ and $\hat{p}_u^{d'}$ are the feature vectors in the annular region far from the target in the color and depth domain, respectively.

As discussed in [11], the back-projection in the color domain is defined as

$$\mathbf{P}_C(\mathbf{x}) = \hat{q}_{b^c(\mathbf{x})}^c, \quad (25)$$

where b^c is the transformation function, as defined in Equation (22), and \mathbf{x} is the pixel coordinate on the image plane.

By using the assumption that the tracked target is with a relative low speed, the back-projection corresponding to the depth domain is defined as [11]:

$$\mathbf{P}_D(\mathbf{x}) = \frac{1}{2} \text{erfc}\left(4 \frac{d(\mathbf{x}, \mathbf{S})}{MAXD} - 2\right), \quad \mathbf{x} \in 2\mathbf{B}_{pre}, \quad (26)$$

where $d(\mathbf{x}, \mathbf{S})$ is the minimal depth distance between the pixel \mathbf{x} and the previous seed set \mathbf{S} ; $MAXD$ is the maximum depth displacement of the target between adjacent frames; $2\mathbf{B}_{pre}$ is the pixel set that is covered by a bounding box of twice the size of the previous object location, but with the same center.

By considering the robustness of \mathbf{P}_C and \mathbf{P}_D , the fused back-projection \mathbf{P} can be calculated as

$$\mathbf{P}(\mathbf{x}) = rc \times \mathbf{P}_C'(\mathbf{x}) + (1 - rc) \times \mathbf{P}_D'(\mathbf{x}), \quad (27)$$

where rc is calculated by

$$rc = \frac{ns_C}{ns_C + ns_D}, \quad (28)$$

where ns_C is the element-weighted correlation coefficient in the color domain. The ideal alternative under the near or far region similarity criterion is the same as $A^* = \langle 1, 0, 0 \rangle$. Then ns_C is defined as

$$ns_C = \sqrt{\alpha} \left(w_{Cn} \frac{T_{Cn}}{\sqrt{\alpha T_{Cn}^2 + \beta I_{Cn}^2 + \gamma F_{Cn}^2}} + w_{Cf} \frac{T_{Cf}}{\sqrt{\alpha T_{Cf}^2 + \beta I_{Cf}^2 + \gamma F_{Cf}^2}} \right), \quad (29)$$

where w_{Cn} , $w_{Cf} \in [0,1]$ are the corresponding weights of criteria and $w_{Cn} + w_{Cf} = 1$. Similarly, the element-weighted correlation coefficient in the depth domain is defined as

$$ns_D = \sqrt{\alpha} \left(w_{Dn} \frac{T_{Dn}}{\sqrt{\alpha T_{Dn}^2 + \beta I_{Dn}^2 + \gamma F_{Dn}^2}} + w_{Df} \frac{T_{Df}}{\sqrt{\alpha T_{Df}^2 + \beta I_{Df}^2 + \gamma F_{Df}^2}} \right), \quad (30)$$

where w_{Dn} , $w_{Df} \in [0,1]$ are the corresponding weights of criteria and $w_{Dn} + w_{Df} = 1$. For \mathbf{P}_C' and \mathbf{P}_D' defined in Equation (27), they are computed by

$$\mathbf{P}_C'(\mathbf{x}) = \begin{cases} \mathbf{P}_C(\mathbf{x}) & \text{if } \mathbf{P}_D(\mathbf{x}) > T_D \\ 0 & \text{else} \end{cases} \quad (31)$$

$$\mathbf{P}_{D'}(\mathbf{x}) = \begin{cases} \mathbf{P}_D(\mathbf{x}) & \text{if } \mathbf{P}_C(\mathbf{x}) > T_C \\ 0 & \text{else} \end{cases}, \quad (32)$$

where $T_D, T_C \in [0,1]$ are the threshold for filtering the noise in the color or depth back-projection. As seen in Equation (31), the information in \mathbf{P}_D is employed for enhancing the color back-projection \mathbf{P}_C , and vice versa for Equation (32).

3.4. Scale Adaption

In this work, the bounding box of the tracked object is firstly decided by the traditional CAMShift algorithm [26]. For each frame, the bounding box \mathbf{B}_{pre} in the previous frame is employed as the start location of the mean shift process. The current tracking location can be calculated by

$$x = \frac{M_{10}}{M_{00}}, \quad y = \frac{M_{01}}{M_{00}}, \quad (33)$$

where $M_{10} = \sum_{\mathbf{x} \in \mathbf{B}} x\mathbf{P}(\mathbf{x})$, $M_{01} = \sum_{\mathbf{x} \in \mathbf{B}} y\mathbf{P}(\mathbf{x})$, and $M_{00} = \sum_{\mathbf{x} \in \mathbf{B}} \mathbf{P}(\mathbf{x})$. Then, the size of the bounding box is $s = 2\sqrt{M_{00}/256}$. For convenience, this bounding box is called the initial bounding box bellow.

Getting the adequate scale of the bounding box is very important. If the scale cannot fit the object for a few frames, the tracker may finally drift from the target. Considering the size of the initial bounding box may be disturbed by the imprecision of the back-projection, a method in the neutrosophic domain is introduced into the scale identification process.

In this work, the color likelihood probability between the candidate object area and the object area is employed as the truth membership. The color likelihood probability between the candidate object area and the background area is employed as the indetermination membership. For the reducing scale alternative, the truth, indetermination, false membership functions are defined as

$$T_{rs} = \sum_{u=1}^{m^c} \sqrt{\hat{q}_u^c \hat{p}_{rsu}^c} \quad (34)$$

$$I_{rs} = \sum_{u=1}^{m^c} \sqrt{\hat{p}_{rsu}^c \hat{b}_{rsu}^c} \quad (35)$$

$$F_{rs} = 1 - T_{rs}, \quad (36)$$

where \hat{p}_{rsu}^c is the u -th bin of the feature vector in the initial bounding box with a reduced scale; \hat{b}_{rsu}^c corresponds to the feature vector in the annular region nearby the scale-reduced bounding box. It must be emphasized that all the pixels taken into consideration must located in the candidate object region **RC** in the current frame. By substituting Equations (34)–(36) into the Equation (6) with the assumption that the idea alternative $A^* = \langle 1, 0, 0 \rangle$, the probability for reducing scale is defined as

$$w_{rs} = \frac{\sqrt{\alpha} T_{rs}}{\sqrt{\alpha T_{rs}^2 + \beta I_{rs}^2 + \gamma F_{rs}^2}}. \quad (37)$$

Similarly, for the expanding scale alternative, the truth, indetermination, and false membership functions are defined as

$$T_{exs} = \sum_{u=1}^{m^c} \sqrt{\hat{q}_u^c \hat{p}_{exsu}^c} \quad (38)$$

$$I_{exs} = \sum_{u=1}^{m^c} \sqrt{\hat{p}_{exsu}^c \hat{b}_{exsu}^c} \quad (39)$$

$$F_{exs} = 1 - T_{exs}, \quad (40)$$

where \hat{p}_{exsu}^c and \hat{b}_{exsu}^c are with the similar meaning as \hat{p}_{rsu}^c and \hat{b}_{rsu}^c , but with an expanded scale for the initial bounding box. Then the probability for reducing scale is defined as

$$w_{exs} = \frac{\sqrt{\alpha}T_{exs}}{\sqrt{\alpha T_{exs}^2 + \beta I_{exs}^2 + \gamma F_{exs}^2}}. \quad (41)$$

Finally, the scale of the initial bounding box is updated by

$$\lambda_{new} = \begin{cases} 1 - \lambda_0 & \text{if } (w_{rs} > sw_{exs}) \\ 1 + \lambda_0 & \text{if } (w_{exs} > sw_{rs}) \\ 1 & \text{otherwise} \end{cases}, \quad (42)$$

where $\lambda_0 \in (0,1)$ is the step value for scale adaption, it should be set to a relatively small value, it is set to 0.04 in this work; $s > 1$ is a scaling factor, it is employed for avoiding the noise in the color domain, and it is set to 1.1 here.

After the scale identification process, a bounding box with the same center as the initial bounding box, but with the tuned width and height is employed as the location of the target in the current frame. The method proposed in [11] for updating the feature vector of the target is employed here.

4. Experiment Results and Analysis

Several challenging video sequences captured from a Kinect sensor are employed. Both color and depth information is provided in each frame. As mentioned at the beginning, several challenging factors like fast motion, blur, deformation, illumination variation, and camera jitter are considered, and those selected testing sequences are all without serious occlusion challenge.

For comparison, the NeutRGBD [11] algorithm is employed. Compared to the tracker proposed here (NeutRGBDs), the NeutRGBD is also a tracker using the CAMShfit framework, but employing different strategies for object region extraction, information fusion, and scale identification. In addition, we implemented a neutrosophic CAMShfit tracker based on the tangent correlation coefficient [4] in the SVNS domain, we call it NeutRGBDst. The only difference between NeutRGBDs and NeutRGBDst is the correlation model, which allows us to ensure that the selection of correlation model is the cause of the performance difference.

To gauge absolute performance, four other trackers, compressive tracking (CT) [33], LGT [34], IVT [35], and TLD [36], are selected. The tracking-by-detection scheme is employed by the three trackers except LGT. For the LGT tracker, features like color, shape, and apparent local motion are introduced into the local layer updating process, and the local patches are applied for representing the target's geometric deformation in the local layer.

4.1. Setting Parameters

For the proposed NeutRGBDs, the fixed sampling space for selecting candidate object seeds is set to 0.02. Due to the top 50% of pixels in the sorted pixel set being sampled, there will be 26 candidate seeds for each frame. After considering the neutrosophic criterion, $N = 6$ seeds are finally selected as object seeds. In order to emphasize the color information to some extent, w_D , w_C , and w_{DC} in Equation (17) are set to 0.3, 0.4, and 0.3, respectively. It is not suggested to set one or two of those three parameters to a very small value, because the information in the corresponding feature domain may be wrongly discarded. The threshold T , defined in Equation (18), which decides the accuracy of \mathbf{R} , is set to 50 mm. If T is set to a big enough value, the whole image region will be added into \mathbf{R} , and a too small value will lead to an incomplete object region. The accuracy of the region extracting result may influence the tracking result to some extent. According to the attribution of the testing sequences and the displacement of the target between adjacent frames, the parameter $MAXD$ in Equation (26) is set to 70 mm. Then w_{Cn} , w_{Cf} , w_{Dn} and w_{Df} in Equations (29) and (30) are all set to 0.5 equally. A relatively low value should be given to the parameters T_D and T_C defined in Equations (31) and (32); otherwise, most

useful information will be wrongly filtered out. In this work, both parameters are set to 0.1. For the element-weighted parameters α , β , and γ , they are set to 0.5, 0.25, and 0.25, respectively in Equations (17), (29), (30), (37) and (41) to enhance the truth element. Finally, all the values of these parameters are chosen by hand-tuning, and all of them are constant for all experiments.

4.2. Evaluation Criteria

Both the center position error and the success ratio are considered. The location error metric is employed for plotting the center position error curve. The center location Euclidean distance between the tracked target and the manually labeled ground truth is applied for calculating the center position error in each frame.

By setting an overlap score r , which is defined as the minimum overlap ratio, one can decide whether an output is correct or not. The success ratio R is calculated by the following formula:

$$R = \sum_{i=1}^N u_i / N, \quad u_i = \begin{cases} 1 & \text{if } s_i > r \\ 0 & \text{otherwise} \end{cases} \quad (43)$$

where N is the total number of frames, s_i is the overlap score, and r is the corresponding threshold. A robust tracker will earn a higher value for R . The overlap score can be calculated as

$$s_i = \frac{\text{area}(\text{ROI}_{T_i} \cap \text{ROI}_{G_i})}{\text{area}(\text{ROI}_{T_i} \cup \text{ROI}_{G_i})} \quad (44)$$

where ROI_{T_i} is the region covered by the target bounding box in the i -th frame, and ROI_{G_i} is the region covered by the corresponding ground truth bounding box.

4.3. Tracking Results

Several screen captures for the testing sequences are given in Figures 1–4. Success and center position error plots of each testing sequence are shown in Figures 5–8. A more detailed discussion is described in the following.

Wr_no1 sequence: This sequence highlights the challenges of rolling, blur, fast motion, and appearance change. As shown in Figure 1, all the trackers perform well until frame #24. However, the bounding box calculated by the NeutRGBD tracker has a relatively small scale. As seen in frames #50 and #56, only the NeutRGBDs tracker produces an adequate scale for the bounding box. As shown in frames #56, #75, and #150, a relatively small scale is estimated by the NeutRGBDst tracker. The CT, LGT, and IVT trackers failed to track the rabbit in frame #75 because of the wrongly estimated scale, and the TLD tracker failed due to the challenge of serious appearance change when the person tried to turn the rabbit back. As shown in Figure 5a, the NeutRGBDs tracker has a good success ratio when different overlap thresholds are selected. Due to the good performance when dealing with the scale adaption problem, the center location of the tracked object is closer to the ground truth in most frames, as seen in Figure 5b. In summary, during the whole tracking process, the NeutRGBDs tracker has the best performance.

Toy_no sequence: The challenges like blur, fast motion, and rotation are included in this sequence. As seen in frame #6, the CT, IVT, and TLD trackers have already failed due to the fast motion of the toy. For the LGT tracker, an improper scale is estimated on account of the update of the local patches cannot follow such a rapid change. For the NeutRGBD tracker, due to the fact that the toy covers a relatively wide range of depth information, and the object seeds only cover a small range, the extracted object region sometimes only covers parts of the toy. As seen in Figure 6, on account of this factor, the center location of the target produced by the NeutRGBD tracker is less stable than that of the NeutRGBDs tracker. Thanks to the seed selection scheme, the NeutRGBDst tracker performs well on this sequence. Unlike the NeutRGBDs tracker, when calculating a neutrosophic correlation coefficient, T , I , and F are treated as the same weight for NeutRGBDst. As seen in Figure 2, the scale produced by the NeutRGBDs tracker performs the best. As shown in Figure 6a, both the NeutRGBDs

and NeutRGBDst tracker perform well, and the NeutRGBDs tracker performs better when the overlap threshold is set to nearly 0.72.

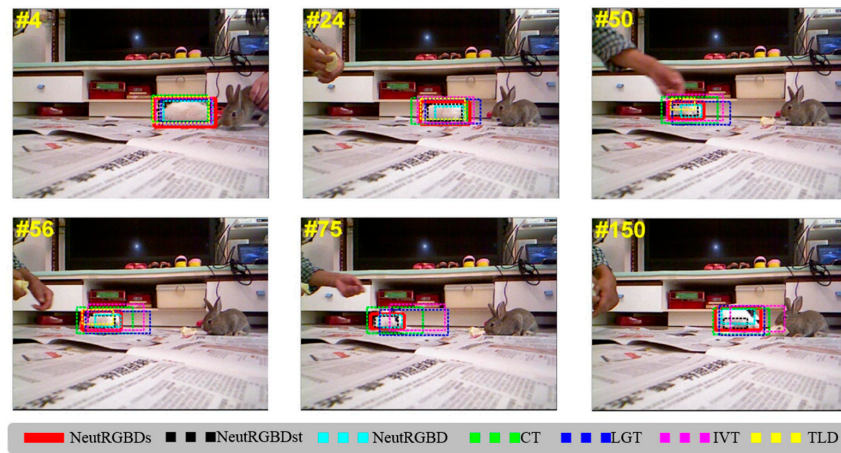


Figure 1. Performance on "wr_no1" sequence by seven trackers.

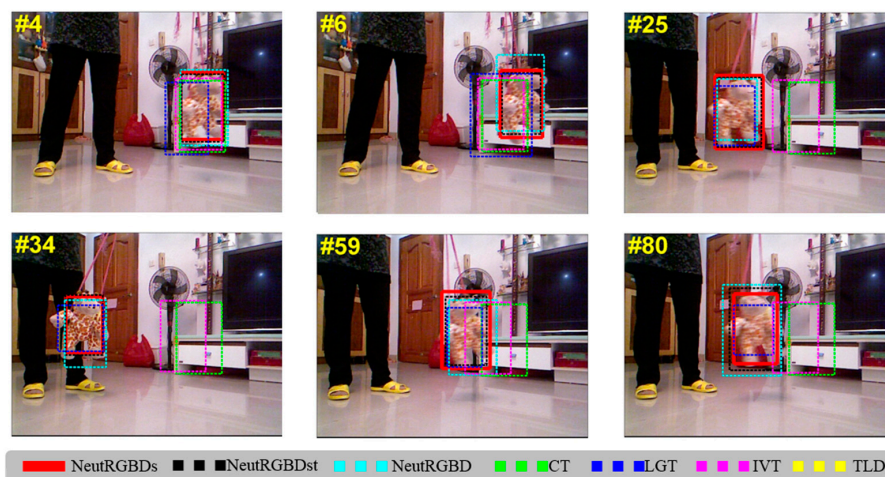


Figure 2. Performance on "toy_no" sequence by seven trackers.

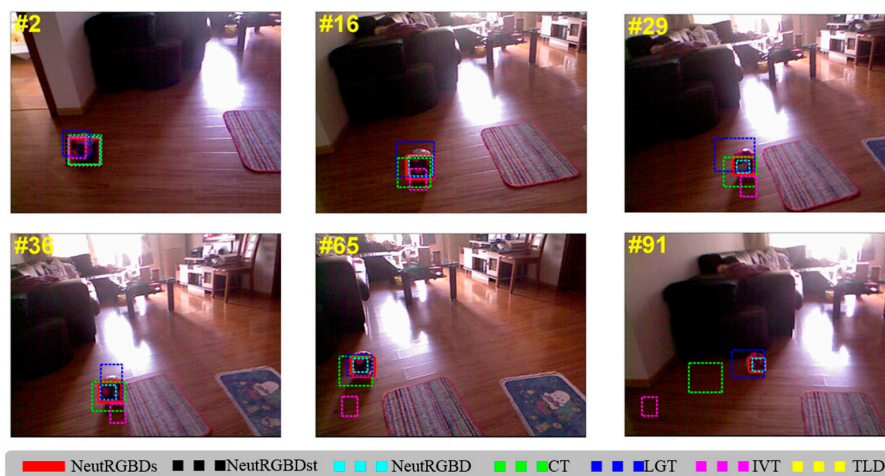


Figure 3. Performance on "zball_no1" sequence by seven trackers.

Zball_no1 sequence: Challenges of illumination variation, rapid motion and camera jitter existed in this sequence. Due to the challenge of appearance change and rapid motion, the TLD tracker fails soon, as seen in Figure 3. The IVT tracker also fails due to the similar surroundings, especially for the wood floor. Though a relatively large scale of the bounding box is estimated by the CT and LGT trackers, both of them can localize the ball properly before frame #65. With the challenges of similar surroundings, rapid motion, and camera jitter, the CT tracker has already failed before frame #91. As shown in Figure 7b, when judging the center position error evaluation criteria, owing to the information fusion in the neutrosophic domain, the NeutRGBDs, NeutRGBDst, and NeutRGBD tracker all perform well. Thanks to the scale adaption and seeds selection strategy, the NeutRGBDs tracker has a more appropriate scale than the other two, as seen in Figures 3 and 7a.

Hand_no_occ sequence: This sequence presents the challenges of illumination variation, deformation, out-of-plane rotation, and similar surroundings. As shown in frame #2 in Figure 4, a large background region is chosen as the object area at the phase of tracker initialization for the CT, LGT, IVT, and TLD tracker. All three trackers except LGT soon fail due to the weak initialization and the out-plane rotation challenge. The LGT tracker performs well throughout this sequence mainly due to the application of the scheme of apparent local motion. However, it is frequently disturbed by the similar surroundings, especially for regions with similar color and displacement. As seen in Figure 8b, a more accurate center has been produced by the NeutRGBDs tracker since frame #150. Although all three neutrosophic-based information fusion trackers perform well in this sequence, the NeutRGBDs tracker can produce a more accurate bounding box, as seen in Figures 4 and 8.



Figure 4. Performance on “hand_no_occ” sequence by seven trackers.

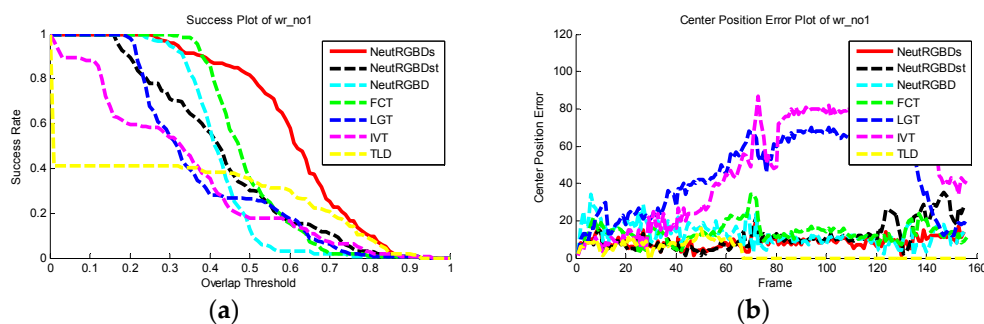


Figure 5. The quantitative plots for each tracker of the *wr_no1* sequence: (a) Success plots; (b) center position error plots.

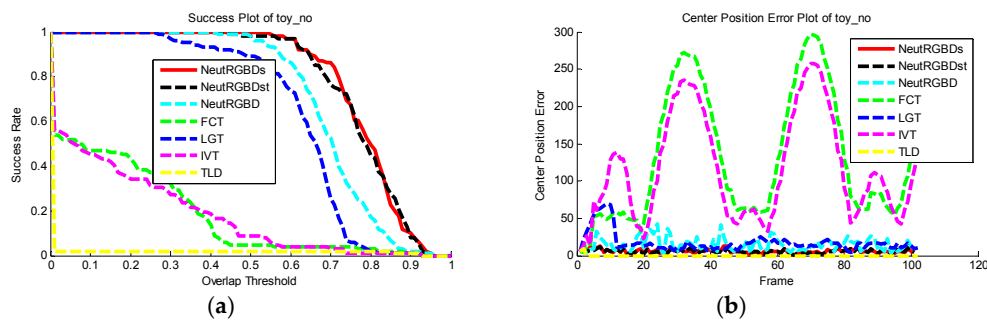


Figure 6. The quantitative plots for each tracker of the *toy_no* sequence: (a) Success plots; (b) center position error plots.

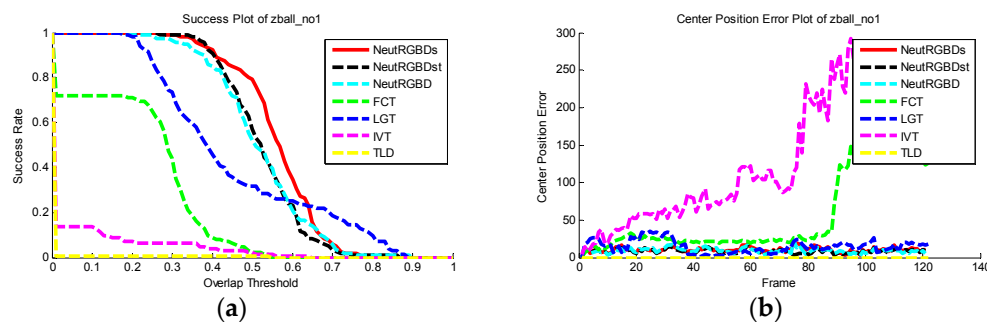


Figure 7. The quantitative plots for each tracker of the *zball_no1* sequence: (a) Success plots; (b) center position error plots.

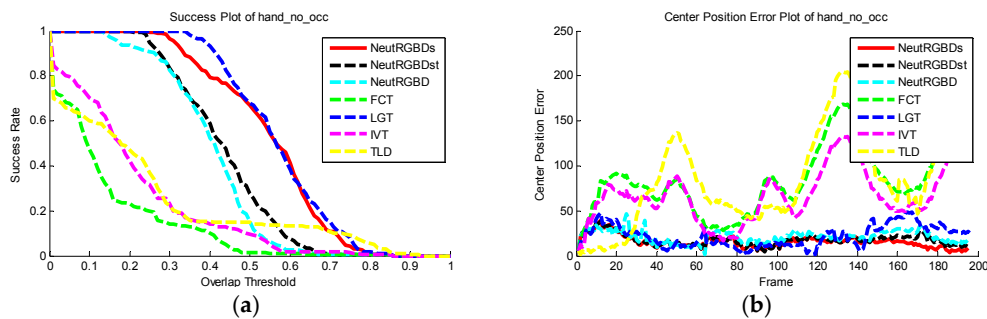


Figure 8. The quantitative plots for each tracker of the *hand_no_occ* sequence: (a) Success plots; (b) center position error plots.

4.4. Discussion

From the above illustrations of the tracking results, we see that a more accurate bounding box can be estimated when the element-weighted neutrosophic correlation coefficient is introduced into the CAMShift framework. Firstly, for the NeutRGBD tracker, only the depth information is considered for seed selection. A more robust seed selection scheme is applied by the NeutRGBDs tracker. Each seed is judged by using the T , I , and F elements in the neutrosophic domain, and the depth, color, and fused information are all taken into consideration. In addition, the truth element is emphasized for each criterion for the NeutRGBDs tracker. The object seeds play an essential role in the procedure of the NeutRGBD, NeutRGBDs, and NeutRGBDst tracker. The robust object seeds help the tracker earn a more accurate object region, as well as a more robust back-projection in the depth domain. Secondly, compared to the NeutRGBD tracker, the NeutRGBDs and NeutRGBDst trackers keep more useful information in both the color and depth domains in the final back-projection. Such a back-projection can provide a more discriminative feature when there are surroundings with similar color or depth to

the target. Finally, for the NeutRGBDs and NeutRGBDst trackers, a scale identification process is first proposed in the neutrosophic domain, and this method contributes a lot when the CAMShift scheme fails to estimate an adequate scale. The only difference between the NeutRGBDs and NeutRGBDst trackers is the calculation of the correlation coefficient. The tangent correlation coefficient employed by the NeutRGBDst tracker treats the T , I , and F elements equally. As can be seen from the above analysis, it is the main reason the NeutRGBDs tracker always produces a more robust target bounding box.

5. Conclusions

A method of element-weighted neutrosophic correlation coefficient is proposed, and it is successfully applied in improving the CAMShift tracker in RGBD video. The experimental results have revealed its robustness. For the selection of robust object seeds, three kinds of criteria are proposed, and each candidate seed is represented in the SVNS domain via three membership functions, T , I , and F . Then these seeds are employed for extracting the object region and calculating the depth back-projection. Furthermore, the proposed neutrosophic correlation coefficient is applied for fusing the likelihood probability in both the color and depth domains. Finally, in order to modify the scale of the bounding box, two alternatives in the neutrosophic domain are proposed, and the corresponding correlation coefficient between the proposed alternative and the ideal one is employed for the identification of the scale. As discussed in this work, challenges without serious occlusion are considered here. It will be our primary mission to try to tackle the occlusion problem through the RGBD information in the future.

Author Contributions: The algorithm was conceived and designed by K.H.; the experiments were performed and analyzed by all the authors; J.Y. and S.S. fully supervised the work and approved the paper for submission.

Funding: This work was supported by the National Natural Science Foundation of China under Grant Nos. 61603258, 61703280, 61601200, and the Plan Project of Science and Technology of Shaoxing City under Grant No. 2017B70056.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Smarandache, F. *Neutrosophy: Neutrosophic Probability, Set and Logic*; American Research Press: Rehoboth, Delaware, 1998; p. 105.
2. Wang, H.; Smarandache, F.; Zhang, Y.; Sunderraman, R. Single valued neutrosophic sets. *Multisp. Multistruct.* **2010**, *4*, 410–413.
3. El-Hefenawy, N.; Metwally, M.A.; Ahmed, Z.M.; El-Henawy, I.M. A review on the applications of neutrosophic sets. *J. Comput. Theor. Nanosci.* **2016**, *13*, 936–944. [[CrossRef](#)]
4. Ye, J.; Fu, J. Multi-period medical diagnosis method using a single valued neutrosophic similarity measure based on tangent function. *Comput. Methods Programs Biomed.* **2016**, *123*, 142–149. [[CrossRef](#)] [[PubMed](#)]
5. Guo, Y.; Şengür, A. A novel image segmentation algorithm based on neutrosophic similarity clustering. *Appl. Soft Comput. J.* **2014**, *25*, 391–398. [[CrossRef](#)]
6. Anter, A.M.; Hassanien, A.E.; ElSoud, M.A.A.; Tolba, M.F. Neutrosophic sets and fuzzy c-means clustering for improving ct liver image segmentation. *Adv. Intell. Syst. Comput.* **2014**, *303*, 193–203.
7. Karabatak, E.; Guo, Y.; Sengur, A. Modified neutrosophic approach to color image segmentation. *J. Electron. Imaging* **2013**, *22*, 4049–4068. [[CrossRef](#)]
8. Zhang, M.; Zhang, L.; Cheng, H.D. A neutrosophic approach to image segmentation based on watershed method. *Signal Process.* **2010**, *90*, 1510–1517. [[CrossRef](#)]
9. Guo, Y.; Şengür, A.; Ye, J. A novel image thresholding algorithm based on neutrosophic similarity score. *Meas. J. Int. Meas. Confed.* **2014**, *58*, 175–186. [[CrossRef](#)]
10. Guo, Y.; Sengur, A. A novel 3d skeleton algorithm based on neutrosophic cost function. *Appl. Soft Comput. J.* **2015**, *36*, 210–217. [[CrossRef](#)]

11. Hu, K.; Ye, J.; Fan, E.; Shen, S.; Huang, L.; Pi, J. A novel object tracking algorithm by fusing color and depth information based on single valued neutrosophic cross-entropy. *J. Intell. Fuzzy Syst.* **2017**, *32*, 1775–1786. [[CrossRef](#)]
12. Hu, K.; Fan, E.; Ye, J.; Fan, C.; Shen, S.; Gu, Y. Neutrosophic similarity score based weighted histogram for robust mean-shift tracking. *Information* **2017**, *8*, 122. [[CrossRef](#)]
13. Biswas, P.; Pramanik, S.; Giri, B.C. Topsis method for multi-attribute group decision-making under single-valued neutrosophic environment. *Neural Comput. Appl.* **2015**, *27*, 727–737. [[CrossRef](#)]
14. Kharal, A. A neutrosophic multi-criteria decision making method. *New Math. Nat. Comput.* **2014**, *10*, 143–162. [[CrossRef](#)]
15. Ye, J. Single valued neutrosophic cross-entropy for multicriteria decision making problems. *Appl. Math. Model.* **2014**, *38*, 1170–1175. [[CrossRef](#)]
16. Majumdar, P. Neutrosophic sets and its applications to decision making. In *Adaptation, Learning, and Optimization*; Springer: Cham, Switzerland, 2015; Volume 19, pp. 97–115.
17. Ye, J. Multicriteria decision-making method using the correlation coefficient under single-valued neutrosophic environment. *Int. J. Gen. Syst.* **2013**, *42*, 386–394. [[CrossRef](#)]
18. Baušys, R.; Juodagalvienė, B. Garage location selection for residential house by waspas-svns method. *J. Civ. Eng. Manag.* **2017**, *23*, 421–429. [[CrossRef](#)]
19. Zavadskas, E.K.; Bausys, R.; Juodagalviene, B.; Garnyte-Sapranaviciene, I. Model for residential house element and material selection by neutrosophic multimooora method. *Eng. Appl. Artif. Int.* **2017**, *64*, 315–324. [[CrossRef](#)]
20. Zavadskas, E.K.; Bausys, R.; Kaklauskas, A.; Ubarte, I.; Kuzminske, A.; Gudiene, N. Sustainable market valuation of buildings by the single-valued neutrosophic mamva method. *Appl. Soft Comput. J.* **2017**, *57*, 74–87. [[CrossRef](#)]
21. Guo, Y.; Sengur, A. Ncm: Neutrosophic c-means clustering algorithm. *Pattern Recognit.* **2015**, *48*, 2710–2724. [[CrossRef](#)]
22. Smeulders, A.W.M.; Chu, D.M.; Cucchiara, R.; Calderara, S.; Dehghan, A.; Shah, M. Visual tracking: An experimental survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 1442–1468. [[PubMed](#)]
23. Wu, Y.; Lim, J.; Yang, M.H. Online object tracking: A benchmark. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Portland, OR, USA, 23–28 June 2013; pp. 2411–2418.
24. Wang, X. *Tracking Interacting Objects in Image Sequences*; École Polytechnique Fédérale de Lausanne: Lausanne, Switzerland, 2015.
25. Comaniciu, D.; Ramesh, V.; Meer, P. Kernel-based object tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 564–577. [[CrossRef](#)]
26. Bradski, G.R. Real time face and object tracking as a component of a perceptual user interface. In Proceedings of the Fourth IEEE Workshop on Applications of Computer Vision, Princeton, NJ, USA, 19–21 October 1998; pp. 214–219.
27. Wong, S. *An Fpga Implementation of the Mean-Shift Algorithm for Object Tracking*; AUT University: Auckland, New Zealand, 2014.
28. Leichter, I. Mean shift trackers with cross-bin metrics. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *34*, 695–706. [[CrossRef](#)] [[PubMed](#)]
29. Zhu, C. Video Object Tracking Using Sift and Mean Shift. Master’s Thesis, Chalmers University of Technology, Gothenburg, Sweden, 2011.
30. Bousetouane, F.; Dib, L.; Snoussi, H. Improved mean shift integrating texture and color features for robust real time object tracking. *Vis. Comput.* **2013**, *29*, 155–170. [[CrossRef](#)]
31. Babenko, B.; Ming-Hsuan, Y.; Belongie, S. Visual tracking with online multiple instance learning. In Proceedings of the IEEE Conference on Computer Vision Pattern Recognition (CVPR), Miami, FL, USA, 20–25 June 2009; pp. 983–990.
32. Hu, K.; Zhang, X.; Gu, Y.; Wang, Y. Fusing target information from multiple views for robust visual tracking. *IET Comput. Vis.* **2014**, *8*, 86–97. [[CrossRef](#)]
33. Kaihua, Z.; Lei, Z.; Ming-Hsuan, Y. Fast compressive tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 2002–2015.
34. Cehovin, L.; Kristan, M.; Leonardis, A. Robust visual tracking using an adaptive coupled-layer visual model. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 941–953. [[CrossRef](#)] [[PubMed](#)]

35. Ross, D.; Lim, J.; Lin, R.-S.; Yang, M.-H. Incremental learning for robust visual tracking. *Int. J. Comput. Vis.* **2008**, *77*, 125–141. [[CrossRef](#)]
36. Kalal, Z.; Mikolajczyk, K.; Matas, J. Tracking-learning-detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 1409–1422. [[CrossRef](#)] [[PubMed](#)]
37. Almazan, E.J.; Jones, G.A. Tracking people across multiple non-overlapping RGB-D sensors. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Portland, OR, USA, 23–28 June 2013; pp. 831–837.
38. Han, J.; Pauwels, E.J.; De Zeeuw, P.M.; De With, P.H.N. Employing a RGB-D sensor for real-time tracking of humans across multiple re-entries in a smart environment. *IEEE Trans. Consum. Electr.* **2012**, *58*, 255–263.
39. Sridhar, S.; Oulasvirta, A.; Theobalt, C. Interactive markerless articulated hand motion tracking using rgb and depth data. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 2456–2463.
40. Camplani, M.; Hannuna, S.; Mirmehdi, M.; Damen, D.; Paiement, A.; Tao, L.; Burghardt, T. Real-time RGB-D tracking with depth scaling kernelised correlation filters and occlusion handling. In Proceedings of the British Machine Vision Conference (BMVC), Swansea, UK, 7–10 September 2015; pp. 145:1–145:11.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).