

Article

# Optical Flow Filtering-Based Micro-Expression Recognition Method

Junjie Wu <sup>1</sup>, Jianfeng Xu <sup>1,\*</sup>, Deyu Lin <sup>1</sup> and Min Tu <sup>2</sup>

<sup>1</sup> School of Software, Nanchang University, Nanchang 330047, China; 411727318007@email.ncu.edu.cn (J.W.); deylin@ncu.edu.cn (D.L.)

<sup>2</sup> Department of Safety Management, Jiangxi Police College, Nanchang 330004, China; tumin\_y@163.com

\* Correspondence: jianfeng\_x@ncu.edu.cn; Tel.: +86-13870861390

Received: 28 October 2020; Accepted: 1 December 2020; Published: 3 December 2020



**Abstract:** The recognition accuracy of micro-expressions in the field of facial expressions is still understudied, as current research methods mainly focus on feature extraction and classification. Based on optical flow and decision thinking theory, we propose a novel micro-expression recognition method, which can filter low-quality micro-expression video clips. Determined by preset thresholds, we develop two optical flow filtering mechanisms: one based on two-branch decisions (OFF2BD) and the other based on three-way decisions (OFF3WD). In OFF2BD, which use the classical binary logic to classify images, and divide the images into positive or negative domain for further filtering. Differ from the OFF2BD, OFF3WD added boundary domain to delay to judge the motion quality of the images. In this way, the video clips with low degree of morphological change can be eliminated, so as to directly improve the quality of micro-expression features and recognition rate. From the experimental results, we verify the recognition accuracy of 61.57%, and 65.41% for CASMEII, and SMIC datasets, respectively. Through the comparative analysis, it shows that the scheme can effectively improve the recognition performance.

**Keywords:** micro-expression recognition (me); two-branch decisions; three-way decisions; optical flow filtering

## 1. Introduction

Micro-expressions are always subtle and unconscious, which the duration of facial expression action is less than 0.2 s [1]. Micro-expressions can mine the true emotional state hidden in the human and demonstrates important application value in medical, polygraph, legal case and other fields. Therefore, micro-expressions recognition has gradually become a hot research issue. In early period, Ekman et al. [2] designed a Facial Action Coding System (FACS) based on the correlation between facial morphology changes to facilitate in-depth research in the field of it, and each of facial action coding is called an action unit (AU). After that, a lot of relative works have been researched on micro-expressions with the gradual maturity of artificial intelligence technology. However, the dataset of micro-expressions is difficult to collect because of the short duration and unconscious characteristics. Therefore, many micro-expression datasets have been constructed for the convenience of researchers for better in-depth research, such as SMIC [3], CASME II [4], SAMM [5], CAS(ME)2 [6], etc.

The main research idea of traditional micro-expression is to mine micro-expression information by extracting the features of the image combined with the classifier method. Pfister et al. [7] proposed an automatic micro-expression recognition system based on a temporal interpolation model and a spontaneous micro-expression corpus. This work is one of the earliest to realize automatic micro-expression recognition. Domestically, as representatives of micro-expression research and development, Fu et al. [8,9] made a lot of efforts to facilitate further research on micro-expression.

With the gradual maturity of deep learning technology, researchers at home and abroad began to use deep learning methods [10–12] for micro-expression image coding and micro-expression recognition. Those methods used CNN and extended methods to replace the traditional texture feature extraction method for feature extraction and recognition.

According to the research, it can be found that the current research on micro-expression recognition at home and abroad is mainly related to the following two links: The first is how to extract high-quality texture features, and the second is to choose which classifier has a good classification effect. However, there are several problems with the current research thinking. 1. Traditional research methods mainly focus on the internal connection between the feature value in the images. Actually, the micro-expression is essentially the combined behavior of the morphological semantics of the various organs of the face, so that the influence of the morphological semantic quality of micro-expression is ignored. 2. Because of the small amount in the micro-expression dataset, the potential of deep learning method in the field of micro expression can't be fully exploited, which also has low interpretability as a brute force method. In response to this problem, this paper introduces the decision-making thinking of rough set theory, which proposes an optical flow filtering method based on optical flow characteristics. This method eliminates video clips with low morphological change to improve the feature expression ability of the micro-expression to be recognized, thereby achieving the purpose of improving the performance of micro-expression recognition.

Therefore, the main contributions are as follows:

1. From the morphological change perspective, we propose a novel image weighting method based on optical flow to assign semantic quality to images between video clips;
2. Two optical flow filtering algorithms are proposed based on two-branch decisions and three-way decisions and a framework of micro-expression recognition, which establishes an interpretable and robust micro-expression recognition theory;
3. Compared with the state-of-the-art solution, the experiment proves the superiority of the effectiveness and efficiency of this method.

The rest of the paper is organized as follows: Section 2 conduct research work on related technologies. Section 3 describes the details of our method. Section 4 reports and discusses the experimental results. Section 5 gives conclusion and future work.

## 2. Related Works

Compared with macro-expression, the micro-expression has the characteristics of short duration and low intensity of facial movements, which leads to greater challenges in the facial features of micro-expressions. For the research on micro-expressions, the current research can be summarized as end-to-end research mode, which mainly focus on pre-processing, diversified feature extraction, and joint classifiers to micro-expression recognition. Nowadays, more and more new research methods are proposed with the maturity of science and technology.

I. Face detection and alignment. Face detection is essential for many facial applications (such as facial recognition and facial expression analysis). Accurate facial recognition through face detection methods can effectively reduce the impact of noise on micro-expressions in human face. The current face detection methods mainly include: 1. Use feature points to align faces and determine key points in the face area; 2. Use deep learning methods for face detection.

Cootes et al. [13] proposed the active shape model (ASM), which uses a 68-point subjective shape model to locate the key points of the human face to obtain the key points of the face area. This is an early widely used method for aligning face parts. Through the improvement of ASM, Cootes et al. [14] further proposed the active appearance model (AAM), which not only includes the establishment of shape models, but also the establishment of texture models and hybrid models. Subsequently, Cootes et al. [15] proposed a constrained local model (CLM), which abandoned the global texture

method of the AAM algorithm, and modeled the local area in the neighborhood of candidate matching feature points and Space position is constrained.

With the maturity of deep learning technology, deep learning is widely applied in different fields [16,17]. Yang et al. [18] proposed a cascaded structure to detect tiny faces in the complex environments. Zhang et al. [19] exploits the inherent correlation between the tasks of detection and alignment, and combined the CNN to proposed MTCNN. MTCNN utilizes a cascaded multi-task framework to boost up the performance of both face detection and alignment. In addition, fast R-CNN [20] or faster R-CNN [21] has also been applied to face detection, which has good improvement in detection speed.

II Feature extraction and derivation. Feature extraction has been a hot issue in the field of micro-expression research. The performance of micro-expression recognition is mainly reflected in the extraction of texture features according to current research. Ojala et al. [22] proposed a kind of binary vector encoding (LBP) of local pixels, and describes the texture information of the current image by generating a histogram. LBP generally used for feature extraction of static image. Expand based on LBP operator, Zhao et al. [23] proposed a dynamic spatiotemporal feature operator LBPTOP. Compared with LBP Operator, LBPTOP takes into account the influence of video spatiotemporal information, thus adding two new feature planes XT and YT of projection on T time axis. Based on the above two classic operators, a variety of feature extraction variant operators have been derived. Huang et al. [24] proposed a spatiotemporal Local Binary Pattern with Integral Projection (STLBP-IP), which uses a differential image-based integral projection method to obtain appearance and motion features in horizontal and vertical directions. Subsequently, Huang et al. [25] proposed a spatial and temporal complete local quantization module (STCLQP) for facial micro-expression analysis, which optimizes the efficiency of micro-expression recognition by extracting three levels of feature vectors of pixel coincidence, amplitude and direction. With the further research, the team [26] once again proposed the local binary mode based on the spatiotemporal radon transform. Guo et al. [27] proposed an extended local binary on the tri-orthogonal plane Value model (ELBPTOP). In order to improve the efficiency of dynamic feature extraction, the LBPSIP [28] (Local Binary Patterns with Six Intersection Points) operator is proposed through the optimization on the LBP-TOP operator. This method transforms the histogram features of the three planes of LBPTOP into a two-dimensional vector encoding between 6 points intersecting XY, XT and YT, and uses the generated histogram as the final dynamic texture feature. Guo et al. [29] proposed to replace the classic LBPTOP operator with CBPTOP operator for image feature extraction, and then use ELM for feature recognition of micro-expressions.

III Research on micro-expression framework. With the gradual maturity of feature operators, in order to break this solidified research model, researchers began to conduct in-depth research on the recognition framework of micro-expression. Jia et al. [30] used the LBP operator to extract the macro-information and the LBP-TOP operator to extract the micro-information to construct the conversion model of the micro expression macro and micro information. Yan et al. [31] restricted local model (CLM) algorithm combined with local binary mode (LBP) and optical flow (of) to identify and analyze the correlation between the frames of face region (ROIs). Wang et al. [32] combining the weights calculated by the cumulative optical flow with the spatiotemporal features extracted by LBPTOP obtained the weighted spatiotemporal features, and the weighted features are input into a support vector machine for micro-expression recognition. Wang et al. [33] convert a color video clip into a four-dimensional vector, where the first two dimensions are spatial information, the third dimension is temporal information, and the fourth dimension is color information. Liong et al. [34] extracted the features of the vertex frame by selecting the Apex frame and the offset frame image of each video and combining the bi-directional weighted optical flow (Bi-WOOF).

IV Application of decision thinking theory. The decision thinking in rough set theory is the core idea of two-branch decision and three-way decision. Two-branch decision emphasizes the right or wrong of things and is absolutely. However, the three-way decision is a decision-making model based on human cognition, which believes that people can immediately make judgments about things that

are fully certain to accept or reject in the actual decision-making process. Additionally, for things that cannot be made immediately, people tend to postpone the judgment of the event. Yang et al. [35] applied the three-way decision in medical diagnosis. Liang et al. [36] introduced the method of two-branch decision and three-way decision into the field of clustering, and defined a cluster calculation method based on clustering.

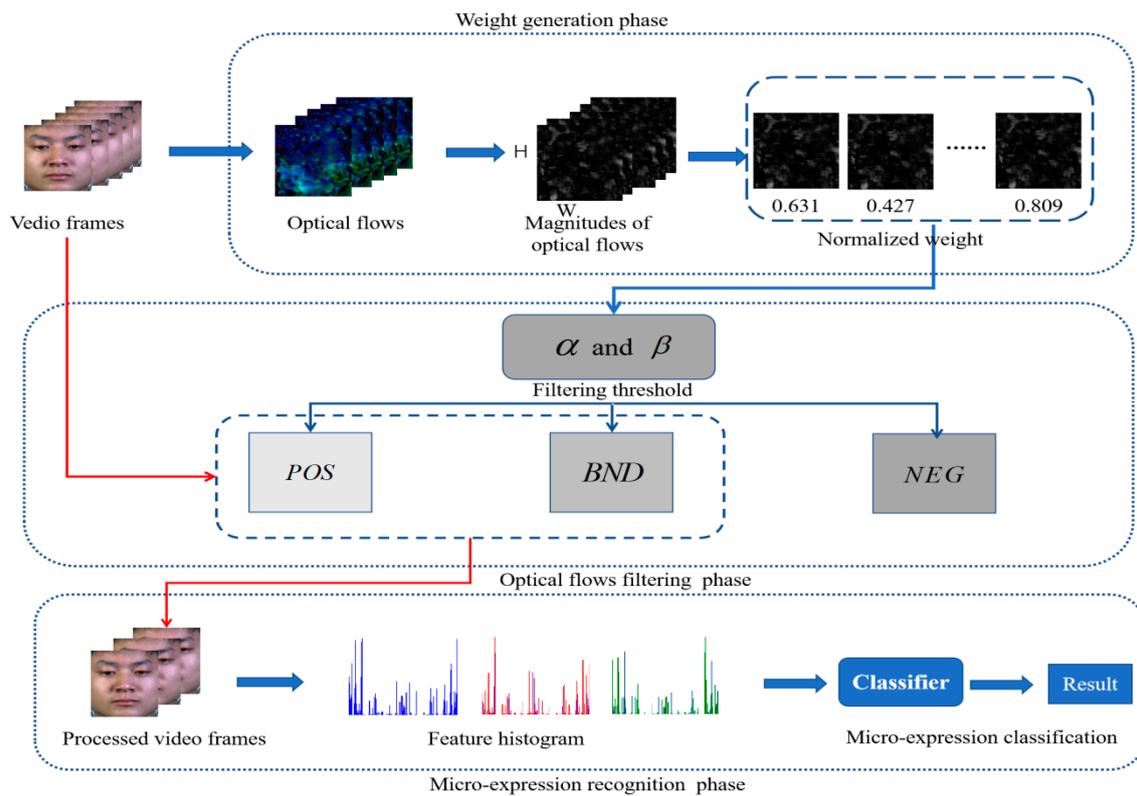
In addition, there are analysis some state-of-the-art-review in the fields of micro-expression, as shown in Table 1. According to the above research results, this paper combines multiple decision-making methods with classic feature extraction methods to construct a new micro-expression research framework.

**Table 1.** Analysis of the state-of-the-art-review for micro-expression recognition.

Methods	Description	Advantage
STCLQP [25]	Extracts the features of sign, magnitude and orientation components information and fused.	For facial micro-expression appearance and motion feature extraction has a better effect.
TAOF [32]	The motion intensity is calculated by accumulating the neighboring optical flow in the time interval, and the feature are further weighted locally.	Enhanced the discrimination of the weighted features for micro-expression recognition.
BI-WOOF [34]	Created a new feature extractor, which to encode essential expressiveness of the apex frame.	Capable of representing the apexframe in a discriminative manner which emphasizes facial motion information at both bin and block levels.
FHOFO [37]	Construct suitable angular histograms from optical flow vector orientations and using histogram fuzzification to encode the temporal pattern.	Insensitive to change of illumination conditions among video clips and the extracted features are robust to the variation of expression intensities
RIMDS [38]	Explored the shallower-architecture and lower-resolution input data, shrinking model and input complexities simultaneously for the composite-database.	Achieved the superior performance compared to the micro-expression recognition approaches.
FMBH [39]	Combing both the horizontal and the vertical components of the differential of optical flow as inspired from the motion boundary histograms (MBH).	The unexpected motions, caused by residual mis-registration that appearsbetween images cropped from different frames, can be removed.
FDM [40]	Characterize the movements of a micro-expression in different granularity.	Provided an effective and intuitive understanding of human facial expressions.
STRCN [41]	A novel deep recurrent convolutional networks based micro-expression recognition approach and capturing the spatiotemporal deformations of micro-expression sequence.	Conquered the “limited and imbalanced training samples” problem and considered the spatiotemporal deformations.

### 3. Proposed Methods

Our proposed method consists of three main phases: Weight generation phase, optical flows filtering phase and micro-expression recognition phase. The main framework of our method is shown in Figure 1.



**Figure 1.** Framework of our proposed micro-expression recognition based on OFF3WD.

### 3.1. Weight Generation Phase

The analysis of SMIC [3] and CASMEII [4] can be found that the video clips contains a large number of images with no expression and low morphological semantic change. The region of the expression is mainly focus on the beginning of Onset-Frame to the end of Offset-Frame, and the peak of semantics is reached in Apex-Frame. Therefore, the features extracted through low motion semantic images will result in a large amount of low-quality and imbalance data. In order to conquered the problem, this paper proposes a low motion semantic image filtering method based on optical flow and decision thinking theory, which is used for the experimental preprocessing of video clips.

A pre-processing step was needed to ensure that have common features because of the resolution and other external features of each dataset used in experiments are various. First, Multi-task Cascaded Convolutional Networks (MTCNN) [19] module was adopted to detect face areas from the input images. Second, it crops only facial area from the entire image to make them the same size and not affected by unnecessary parts such as hair or accessories. Additionally, the cropped face images are resized to a size of  $300 \times 250$ . In this manner, the variations of spatial appearance among different video clips are normalized. Subsequently, the dense optical flows [42] are extracted from the aligned video clips to represent motion information of each pixel in micro-expression video clips. The optical flows  $\xi_{o_i}(x, y)$  are computed from each pair of neighboring images. When input  $t$ -frames video  $V = \{v_1, v_2, \dots, v_t\}$  into the method,  $t - 1$  frames optical flows  $O = \{o_1, o_2, \dots, o_{t-1}\}$  will be obtained, Where  $o_i$  is the optical flow image between  $v_i$  and  $v_{i+1}$ . The optical flows  $\xi_{o_i}(x, y)$  consist of horizontal displacements and vertical displacements. Therefore, the optical flows  $\xi_{o_i}(x, y)$  can be expressed as:

$$\xi_{o_i}(x, y) = (h_{o_i}(x, y), v_{o_i}(x, y)) \tag{1}$$

where  $h_{o_i}(x, y)$  and  $v_{o_i}(x, y)$  indicate the horizontal and vertical displacements respectively.

According to the optical flow of obtaining  $t - 1$  frames, for each image of optical flow, the Euclidean distance of the horizontal and vertical displacement of each pixel is obtained as the motion intensity of

the current pixel. Then, the optical flow motion intensity  $W_{o_i}$  of the entire optical flow image of  $W \times H$  pixels can be calculated as:

$$W_{o_i} = \sum_{x=0}^W \sum_{y=0}^H \sqrt{h_{o_i}(x, y)^2 + v_{o_i}(x, y)^2} \quad (2)$$

where  $o_i$  are the  $i$ -frame optical flow and  $i = 1, 2, \dots, t-1$ ,  $W$  represents for optical flow  $o_i$  width,  $H$  represents for optical flow  $o_i$  height.

For the optical flow  $o_i$  in the set  $O$ , given a semantic quality metric  $\bar{W}_{o_i}$  called the optical flow average motion intensity to measure the quality of  $o_i$ .  $\bar{W}_{o_i}$  represents the average unit pixel semantic quality of the optical flow  $o_i$  w.r.t the  $v_i$  to  $v_{i+1}$ . The calculation formula as shown in Equation (3):

$$\bar{W}_{o_i} = \frac{W_{o_i}}{W \times H} \quad (3)$$

where  $W_{o_i}$  represents the motion intensity of  $i$ -frame optical flow,  $W$  represents for optical flow  $o_i$  width,  $H$  represents for optical flow  $o_i$  height.

Therefore, the motion intensity of optical flow can be measured by the average motion intensity value of optical flow in Equation (3). The average motion intensity reflects the strength of each group of neighboring video images. That is, the larger the average motion intensity of the optical flow, the larger the semantic quality between neighboring video frames. Namely, the smaller the average motion intensity of the optical flow, the smaller the semantic quality between neighboring video frames. Therefore, the average motion intensity of optical flow can be used to measure semantic quality information.

In order to better classify the semantic quality, the semantic quality of optical flow is evenly distributed in  $[0, 1]$  by defining a normalization weighting formula. That is to say, the larger the average motion intensity of the optical flow, the larger the normalized semantic weight value; on the contrary, the smaller the average motion intensity of the optical flow, the smaller the normalized semantic weight value. For each optical flow frame of the average motion intensity  $\bar{W}_{o_i}$ , the evaluation weighting of optical flow weighting filtering algorithm as shown in Equation (4).

$$\omega_{o_i} = -e^{-\frac{\bar{W}_{o_i}}{\theta}} + 1 \quad (4)$$

where  $\theta$  is a parameter to adjust the influence of the weighted over the index.

Therefore, all the video frames obtained the semantic quality weight information  $\omega_{o_i}$  ( $i = 1, 2, \dots, t-1$ ) between neighboring video frames, the weight information represent the strength of the face area change between the neighboring video frames.

### 3.2. Optical Flows Filtering Phase

#### 3.2.1. Optical Flows Filtering Algorithm Based on Two-Branch Decisions

The two-branch decisions have important application value in the choice of uncertainty about the data removal and retention. Therefore, determine the semantic quality of the video frames by defining the weight threshold  $\delta$  ( $\delta \in (0, 1)$ ), which facilitates the classification of high-quality or low-quality semantic images. In solving uncertainty problems, classic binary logic embodies its important role. In distinguishing the motion intensity, the weight threshold is used to judge the semantic quality of the video frames, so that the optical flow weight between the video frames is divided into an object set  $POS_{\delta}(O)$  or  $NEG_{\delta}(O)$ .

When  $\omega_{o_i} \geq \delta$ , the optical flow frame  $o_i$  is divided into the positive domain of the set  $\omega_{o_i}$  and this domain is denoted as  $POS_{\delta}(O)$ , recorded as  $o_i \in POS_{\delta}(O)$ ;

When  $\omega_{o_i} < \delta$ , the optical flow frame  $o_i$  is divided into the negative domain of the set  $\omega_{o_i}$ , and this domain is denoted as  $NEG_{\delta}(O)$ , recorded as  $o_i \in NEG_{\delta}(O)$ .

Through the above method, the optical flow frames with high semantic change can be distinguished from the optical flow frames with low semantic change, so that the high semantic quality optical flow frames in set  $POS_\delta(O)$ , and the low semantic quality optical flow frames in set  $NEG_\delta(O)$ .

With the average motion intensity weight  $\omega_{o_i}$  ( $i = 1, 2, \dots, t - 1$ ) and threshold  $\delta$  ( $\delta \in (0, 1)$ ), the motion intensity of the optical flow frame is divided into a set  $POS_\delta(O)$  and a set  $NEG_\delta(O)$ . The more essential meaning is to distinguish images with small morphological changes between neighboring video frames from pictures with large morphological changes. This method uses the image of the first frame as the semantic benchmark to mine the morphological semantic quality of time series. Corresponding to this, the video frames  $V$  can be divided as follows according to the optical flow sets  $POS_\delta(O)$  and  $NEG_\delta(O)$  as follows:

When  $o_i \in POS_\delta(O)$ , the video frame  $v_{i+1}$  is divided into the high semantic quality domain, recorded as  $v_{i+1} \in POS_\delta(V)$ ;

When  $o_i \in NEG_\delta(O)$ , the video frame  $v_{i+1}$  is divided into the low semantic quality domain, recorded as  $v_{i+1} \in NEG_\delta(V)$ ;

At this point, any video frame  $v_i \in V$  in the video frames set  $V$  given semantic quality w.r.t  $V$  and based on optical flow weighting of the image semantic quality filtering algorithm is completed, which lays a solid foundation for the establishment of the morphological semantic database in Section 4 below. The pipeline of OFF2BD algorithm is illustrated in Algorithm 1.

---

**Algorithm 1** Optical Flows Filtering Algorithm Based on Two-Branch Decisions (OFF2BD)

---

**Input:** video frames  $V = \{v_1, v_2, \dots, v_t\}$ , threshold  $\delta$

**Output:** High-quality semantic image set  $V$  after optical flow filtering

1. For  $v_i = 1, 2, \dots, |V| - 1$  /\*Get initial optical flow set  $O$ \*/;
  2. For  $v_j = 2, 3, \dots, |V|$ ;
  3. Generate optical flow  $o_i$  for  $v_i$  and  $v_j$ ;
  4. End for
  5. End for
  6. Obtain optical flow set  $O$ ;
  7. For  $o_i = 1, 2, \dots, |O|$  /\*Two-branch decisions iteration for optical flow in  $O$ \*/;
  8. Calculate the average optical flow motion intensity  $\bar{W}_{o_i}$  of optical flow  $o_i$  w.r.t. Formulae (2) and (3);
  9. Calculate the optical flow weight  $\omega_{o_i}$  based on the average optical flow motion intensity  $\bar{W}_{o_i}$  w.r.t. formula 4;
  10. If  $\omega_{o_i} \geq \delta$  then;
  11. put  $o_i$  into  $POS_\delta(O)$ ;
  12. Else;
  13. put  $o_i$  into  $NEG_\delta(O)$ ;
  14. End if;
  15. End for;
  16. Delete the optical flows set of  $NEG_\delta(O)$  and update the optical flow set  $O$  with  $POS_\delta(O)$ ;
  17. Update video set  $V$  according to the updated optical flow set  $O$ ;
  18. End for;
  19. Output high-quality semantic image set  $V$ .
- 

### 3.2.2. Optical Flows Filtering Algorithm Based on Three-Way Decisions

As the most important extension of the two-branch decisions, the three-way decisions [43] is an important granular calculation method derived from the rough set theory in machine learning, which is a new theory for solving uncertain problems proposed by the famous granular computing expert Yao [44] for the first time based on rough set theory and time research. Compared with the traditional two-branch decisions which requires the iterative incremental solution to solve the uncertainty of the problem, the three-way decision increases the delay behavior when the decision making is not accurate or cannot make accurate positive and negative decisions. The three-way decisions strategy not only reasonably explain the division behavior of the three decision domains (positive domain,

negative domain and boundary domain) in the rough set theory, but also reduce the time consumption (abandoning incremental decision-making strategy (i.e., adding new knowledge constantly to make sure the uncertainty part turn to certain domain) in two-branch decisions.)

In dealing with uncertain problems, the three-way decisions method developed by the two-decision method is more suitable for human inertial thinking in solving problems. This approach is more representative of the problem-handling theory of contradictory relation and contrary relation problems. Therefore, in order to improve the semantic relationship between video frames and improve the quality of feature extraction, the quality thresholds  $\alpha$  ( $\alpha \in (0, 1)$ ) and  $\beta$  ( $\beta \in (0, 1) \wedge \beta > \alpha$ ) of the video frames are defined for optical flow filtering. And then, the optical flows can be divided into an object set  $POS_{(\alpha,\beta)}(O)$  or  $BND_{(\alpha,\beta)}(O)$  or  $NEG_{(\alpha,\beta)}(O)$ .

When  $\omega_{o_i} \geq \beta$  the optical flow frame  $o_i$  is divided into the positive domain of the set  $\omega_{o_i}$  and this domain is denoted as  $POS_{(\alpha,\beta)}(O)$ , recorded as  $o_i \in POS_{(\alpha,\beta)}(O)$  ;

When  $\alpha \leq \omega_{o_i} < \beta$ , the optical flow frame  $o_i$  is divided into the positive domain of the set  $\omega_{o_i}$  and this domain is denoted as  $BND_{(\alpha,\beta)}(O)$ , recorded as  $o_i \in BND_{(\alpha,\beta)}(O)$  ;

When  $\omega_{o_i} < \alpha$ , the optical flow frame  $o_i$  is divided into the negative domain of the set  $\omega_{o_i}$ , and this domain is denoted as  $NEG_{(\alpha,\beta)}(O)$ , recorded as  $o_i \in NEG_{(\alpha,\beta)}(O)$ .

Three-way decisions methods for optical flow quality division are defined for  $(\alpha, \beta)$  threshold, wherein the  $(\alpha, \beta)$  are usually set to  $0 < \alpha < \beta < 1$ . Compared with the two-branch decisions filtering method, this method adds the uncertainty domain  $BND_{(\alpha,\beta)}(O)$  as a buffer area for uncertain semantic quality processing. Then, the  $BND_{(\alpha,\beta)}(O)$  domain and the  $POS_{(\alpha,\beta)}(O)$  domain are jointly used as the update domain, and iterative optical flow filtering analysis is performed until convergence is reached or a certain number of stops.

As the same time, the first frame of the image is used as a semantic reference combined with a convergent high-quality optical watershed to divide the collection of video frames. Corresponding to this, the video frames  $V$  can be divided as follows according to the optical flow set  $POS_{(\alpha,\beta)}(O)$ ,  $BND_{(\alpha,\beta)}(O)$  and  $NEG_{(\alpha,\beta)}(O)$  as follows:

When  $o_i \in POS_{(\alpha,\beta)}(O)$ , the video frame  $v_{i+1}$  is divided into the high semantic quality domain, recorded as  $v_{i+1} \in POS_{(\alpha,\beta)}(V)$  ;

When  $o_i \in BND_{(\alpha,\beta)}(O)$ , the video frame  $v_{i+1}$  is divided into the fuzzy semantic quality domain, recorded as  $v_{i+1} \in BND_{(\alpha,\beta)}(V)$  ;

When  $o_i \in NEG_{(\alpha,\beta)}(O)$ , the video frame  $v_{i+1}$  is divided into the low semantic quality domain, recorded as  $v_{i+1} \in NEG_{(\alpha,\beta)}(V)$  ;

When a specified number of iterative filtering or convergence is achieved, the  $POS_{(\alpha,\beta)}(V)$  and  $\in BND_{(\alpha,\beta)}(V)$  video frames set are jointly updated to the video frames set of the target domain  $V$ , and the video frame is reordered to obtain a high-quality semantic sequence video combination. The pipeline of OFF3WD algorithm is illustrated in Algorithm 2.

---

**Algorithm 2** Optical flows filtering algorithm based on three-way decision (OFF3WD)

---

**Input:** video frames  $V = \{v_1, v_2, \dots, v_t\}$ , threshold  $\alpha, \beta$ , iteration coefficient  $\eta$  ;

**Output:** High-quality semantic image set  $V$  after optical flow filtering.

1. For  $k = 1, 2, \dots, \eta$  /\*Filtering iteration for video frames in set  $V$  \*/;
  2. For  $v_i = 1, 2, \dots, |V| - 1$  /\*Get initial optical flow set  $O$  \*/;
  3. For  $v_j = 2, 3, \dots, |V|$  ;
  4. Generate optical flow  $o_i$  for  $v_i$  and  $v_j$ ;
  5. End for
  6. End for
  7. Obtain optical flow set  $O$  ;
  8. For  $o_i = 1, 2, \dots, |O|$  /\*Three-way decision iteration for optical flow in  $O$  \*/;
  9. Calculate the average optical flow motion intensity  $\overline{W}_{o_i}$  of optical flow  $o_i$  w.r.t. Formulae (2) and (3);
-

10. Calculate the optical flow weight  $\omega_{o_i}$  based on the average optical flow motion intensity  $\overline{W}_{o_i}$  w.r.t. Formula (4);
11. If  $\omega_{o_i} \geq \beta$  then;
12. put  $o_i$  into  $POS_{(\alpha,\beta)}(O)$  ;
13. If  $\alpha \leq \omega_{o_i} < \beta$  then;
14. put  $o_i$  into  $BND_{(\alpha,\beta)}(O)$ ;
15. Else;
16. put  $o_i$  into  $NEG_{(\alpha,\beta)}(O)$  ;
17. End if;
18. When  $|NEG_{(\alpha,\beta)}(O)| = 0$  then; /\*Reach the convergence condition\*/
19. Combine  $POS_{(\alpha,\beta)}(O)$  and  $BND_{(\alpha,\beta)}(O)$  to update  $O$  ;
20. Break;
21. Increase the  $\alpha$  threshold naturally and delete the optical flows set of  $NEG_{(\alpha,\beta)}(O)$  ;
22. Combine  $POS_{(\alpha,\beta)}(O)$  and  $BND_{(\alpha,\beta)}(O)$  to update  $O$  ;
23. Update video set  $V$  according to the updated optical flow set  $O$  ;
24. End for;
25. End for;
26. Obtain the video frame set  $V$  according to the completed iteration condition or the converged optical flow set  $O$  ;
27. Output high-quality semantic image set  $V$  .

### 3.3. Micro-Expression Recognition Phase

#### 3.3.1. Feature Extraction

In the field of micro-expression recognition, Local Binary Pattern (LBP) [22] is a common symbiotic pattern method that encodes local pixels in an image. The original LBP operator is defined as within the window of  $3 \times 3$ , taking the center pixel of the window as the threshold, and comparing the gray value of the adjacent 8 pixels with the center pixel in a clockwise direction, if the surrounding pixel value is greater than or equal to the center pixel value, the position of the pixel is marked as 1, otherwise it is 0. In this way, 8 pixels in the neighborhood of  $3 \times 3$  can be compared to produce 8-bit binary numbers (usually converted to decimal numbers, that is, LBP codes, a total of 256 kinds), that is, the LBP value of the pixel at the center of the window is obtained, and this value is used to Reflect the texture information of this area. Such a  $3 \times 3$  picture can be transformed into 256-dimensional texture features. The formula is shown in Equation (5).

$$H(x_c, y_c) = \sum_{p=0}^{p-1} 2^p S(i_p - i_c) \tag{5}$$

$$S(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{else} \end{cases} \tag{6}$$

where  $(x_c, y_c)$  represents the central pixel,  $i_c$  is the gray value,  $i_p$  is the gray value of the adjacent point, and  $S(x)$  is a function that conforms to Equation (6). According to the above formula method, the binary extraction process of the face picture is shown in Figure 2.

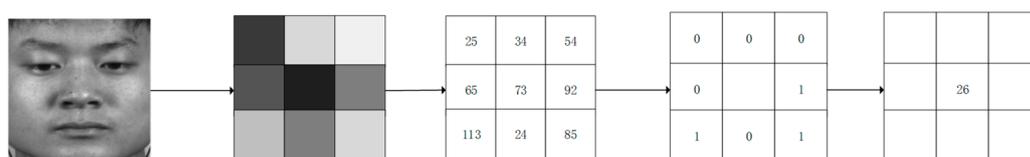


Figure 2. Example of encoding a LBP feature.

In order to solve the dynamic feature extraction, based on the LBP method proposed a dynamic feature extraction method LBPTOP (Local Binary Pattern from Three Orthogonal Planes). LBPTOP solves the problem that LBP can only extract two-dimensional features, and can be applied to feature extraction in three dimensions. LBPTOP extracts LBP features in three dimensions from the spatial plane XY and the spatial-temporal planes XT and YT respectively, and then combines them in series to form spatial-temporal features.

We use LBPTOP, LBPSIP, STCLQP, STLBPIP feature descriptors to encode the optical flow filtered video to express temporal and spatial information. These feature descriptors are all improved based on the LBPTOP method, and feature enhancement is performed on the three planes of the spatial plane XY and the spatial-temporal plane XT, YT. Therefore, the feature histograms of the based on the two-branch decisions spatial-temporal video  $V$  obtained after the optical flow filtering can be expressed as:

$$\begin{cases} V_{\delta} = POS_{\delta}(V) \\ H_{\delta} = [H_{V_{\delta}}^{XY}, H_{V_{\delta}}^{XT}, H_{V_{\delta}}^{YT}] \end{cases} \quad (7)$$

where  $V_{\delta}$  represents the set of high-quality semantic frame spatiotemporal sequences obtained after the convergence of the OFF2BD algorithm.  $H_{\delta}$  are the functions to get the feature histograms from the  $V_{\delta}$ .

To further improve the morphological semantics quality of the spatial-temporal video  $V$ , the feature histograms based on OFF3WD can be expressed as:

$$\begin{cases} V_{\alpha,\beta} = POS_{\alpha,\beta}(V) \cup BND_{\alpha,\beta}(V) \\ H_{\alpha,\beta} = [H_{V_{\alpha,\beta}}^{XY}, H_{V_{\alpha,\beta}}^{XT}, H_{V_{\alpha,\beta}}^{YT}] \end{cases} \quad (8)$$

where  $V_{\alpha,\beta}$  represents the set of high-quality semantic frame spatiotemporal sequences obtained after the convergence of the OFF3WD algorithm.  $H_{\alpha,\beta}$  are the functions to get the feature histograms from the  $V_{\alpha,\beta}$ .

### 3.3.2. Micro-Expression Recognition

For classic micro-expression feature recognition, Whitehill et al. [45] analyzed and compared it and found that SVM has relatively good recognition effect on micro-expression recognition compared with other traditional machine learning algorithms. FHOFO [37] uses linear SVM, 3point KNN and LDA to make correlation comparison among different micro-expression datasets. The experimental results show that compared with other classifiers, linear SVM has good generalization ability and stability. On the classification method, the leave-one-subject-out (LOSO) method was used to train and test the classifier. The LOSO-based validation is performed considering all images in the dataset. Given a dataset of  $n$  samples, the LOSO-based validation is performed with  $n$  iterations, such that in each iteration the classifier is trained with  $n-1$  samples and tested on the remaining one sample. Using this strategy to evaluate the classification effect can more intuitively and accurately reflect the effect of the current experimental method. Therefore, this classification test method is currently used in most machine vision fields.

## 4. Experiments and Discussion

In this paper, The SMIC and CASME II will be used for optical flow filtering related experiments. All experiments in this paper are running on a PC of Windows 10, intel i7-8750H, 16G memory, python 3.7 environments, MATLAB 2016a and the python IDE environment is pyCharm professional 2019.3.

### 4.1. Datasets

The experiment selects two most popular spontaneous datasets, including CASME II and SMIC are used to evaluate the performance of the proposed method. The datasets statistics are summarized in Table 2.

**Table 2.** A summary of the different features of the SMIC and CASME II datasets.

Feature	SMIC-HS	SMIC-VIS/NIR	CASME II
subjects	16	16	35
samples	164	71	247
Frames per sec.	100	25	200
Posed/spontaneous	Spontaneous	Spontaneous	Spontaneous
AU labels	no	no	yes
Emotion class	3(positive/negative/surprise)	3(positive/negative/surprise)	5(basic emotions)
Tagging	Emotion category	Emotion category	Emotion/FACS

SMIC [3]: SMIC was composed of 164 video clips of 16 subjects, which were divided into three different categories, such as Positive (51 samples), Negative (70 samples) and Surprise (43 samples). There are three versions of SMIC data: one is the high-speed camera (HS) of 100fps, the ordinary camera (VIS) version of 25 fps, and the near-infrared camera (NIR) version of 25 fps. HS cameras were used to record all data, while vis and NIR cameras were used to record data from the last eight subjects. Emotion classes are only based on participants' self-report.

CASME II [4]: CASME II contains 247 micro-expression video clips from 26 subjects. All samples were recorded by high-speed cameras at 200 fps. The sampling resolution is  $640 \times 480$  pixels. These samples were divided into five categories: Happiness (32 samples), Surprise (25 samples), Disgust (64 samples), Repression (27 samples) and Others (99 samples). Unlike SMIC, the AU label of CASME II follows the facial action coding system (FACS).

#### 4.2. Experimental Setting

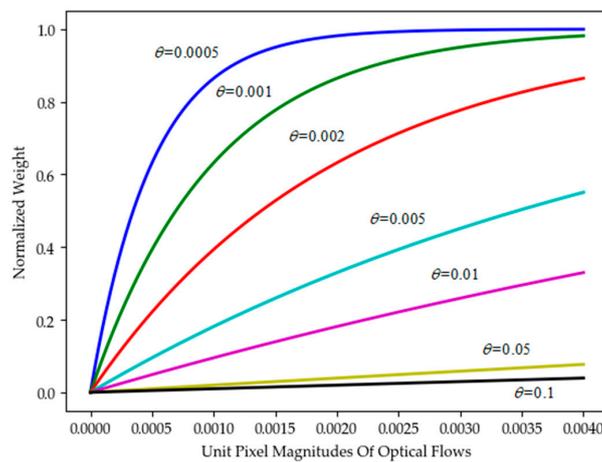
This section aims to verify the adaptability and performance of the method in this paper through relevant pairing and analysis experiments. MTCNN has good accuracy for face detection compared to traditional face alignment methods. Therefore, MTCNN is used for face positioning on the SMIC and CASME II data images, and the face region is cropped according to the final bounding box to obtain an average face area image size of  $300 \times 250$ .

Meanwhile, this article uses two optical flow filtering algorithms (OFF2BD, OFF3WD) to reduce the impact by low-quality semantic images. Through OFF2BD-LBPTOP, OFF3WD-LBPTOP, OFF2BD-LBPSIP, OFF3WD-LBPSIP, OFF2BD-STCLQP, OFF3WD-STCLQP, OFF2BD-STLBPIP, OFF3WD-STLBPIP methods and classic micro-expressions recognition methods LBP-TOP, LBP-SIP, STCLQP, STLBP-IP for performance comparison. In order to verify the effectiveness of the optical flow filtering algorithm proposed in this paper on micro-expression recognition, several groups of experiments are designed as follows:

- Sensitivity of the parameter  $\theta$  to normalization function  $\omega$ .
- Sensitivity of the threshold  $\delta$  to OFF2BD.
- Sensitivity of the threshold  $(\alpha, \beta)$  to OFF3WD.
- The comparison of performance scores between refined algorithms (OFF2BD-LBPTOP, OFF3WD-LBPTOP, OFF2BD-LBPSIP, OFF3WD-LBPSIP, OFF2BD-STCLQP, OFF3WD-STCLQP, OFF2BD-STLBPIP, OFF3WD-STLBPIP) and original algorithms (LBPTOP, LBPSIP, STCLQP, STLBP-IP).
- The confusion matrix of recognition accuracies for each micro-expression on CASMEII and SMIC dataset.

Choices of Parameters  $\theta$ : The size of the cropped face image is equalized to  $300 \times 250$  for SMIC and CASME II. A parameter  $\theta$  is adopted in the computation of normalized weighting to adjust the influence of the magnitudes of optical flows over the index. Among then, through the image calculation of CSAMEII and SMIC the unit pixel magnitudes of each optical flow belong to the interval of  $[3 \times 10^{-4}, 3.5 \times 10^{-3}]$ . Parameter  $\theta$  setting range is (0,1), and its relationship with the normalized weight can be represented by Figure 3. As shown in Figure 3, when setting  $\theta$  within a certain range

values, e.g., setting  $\theta < 0.005$ , the Normalized Weight steady increase as the unit pixel magnitudes of optical flow increases and the difference between high-quality semantic optical flow weights and low-quality semantic weights is evenly expanded. When setting  $\theta$  to larger values, the weight gap of high-quality semantic optical flow and that of low-quality semantic optical flow will be narrowed down, this will directly lead to the effect of the next optical flows weight filtering. Empirically, it is suggested that the parameter  $\theta$  be set in the interval of  $[5 \times 10^{-4}, 0.01]$ . Whereas  $\theta = 0.001$ , the weighted interval with a more uniform distribution will be obtained. Therefore, this paper chooses  $\theta = 0.001$  as the weight coefficient parameter of the experiment.



**Figure 3.** Correlation between normalized weight and unit pixel magnitudes of optical flows with different parameters  $\theta$ .

Performance Evaluation Index: Accuracy is a classic method to judge the performance of the model in machine learning. At present, most judgments on micro-expression recognition models are judged by recognition accuracy. The formula is as follows:

$$Acc = \frac{n^*}{n^* + m^*} \tag{9}$$

where  $n^*$  is the number of correct micro-expressions predicted by the model,  $m^*$  is the number of incorrect micro expressions predicted by the model,  $n^* + m^*$  is the total number of micro expressions in the micro-expression dataset.

F1 score (F1 Score) is an indicator used to measure the accuracy of the binary classification model in statistics, and it also takes into account the precision rate  $p$  and recall rate  $r$  of the classification model. Therefore, by calculating the recall rate  $r$  combined with the precision rate  $p$  to further calculate F1-score. In the prediction of the  $i$ -th micro-expression class, the number  $N_i^{TP}$  of true positive cases of the  $i$ -th micro-expression class that is predicted to be correct, the number  $N_i^{FP}$  of false positive cases that the  $i$ -th micro-expression class is predicted to be incorrect and the number  $N_i^{FN}$  of false negative cases that the non-  $i$ -th micro-expression class are predicted as  $i$ -th micro-expression class. The precision rate  $p_i$  and the recall rate  $r_i$  is calculated as follows:

$$p_i = \frac{N_i^{TP}}{N_i^{TP} + N_i^{FP}} \tag{10}$$

$$r_i = \frac{N_i^{TP}}{N_i^{TP} + N_i^{FN}} \tag{11}$$

where  $N_i^{TP} + N_i^{FP}$  is the total number of the real  $i$ -th micro-expression class,  $N_i^{TP} + N_i^{FN}$  is the number of  $i$ -th micro-expression class after prediction.

Calculation of micro-expression precision  $p$  and recall rate  $r$  of  $N$  ME classes, F1-score is defined as:

$$F = \frac{1}{N} \sum_{i=1}^N \frac{2p_i \times r_i}{p_i + r_i} \quad (12)$$

where  $p_i$  and  $r_i$  are the precision and recall of the  $i$ -th micro-expression class.

The values of the above Accuracy and F1-score performance evaluation indicators are in the range  $[0, 1]$ , and the larger the indicator value, the better the recognition performance of the framework model for micro-expressions.

#### 4.3. Experiments and Analysis

The LOSO method is used to the CASMEII and SMIC data are experimentally reproduced based on the LBP-TOP feature, and the selection of optimized relevant parameters value to further complete the parameter adjustment of the OFF2BD and OFF3BD algorithms. The video clips are divided to  $N \times N \times 2$  blocks. The recognition results of micro-expressions are shown in Table 3, the best accuracy and F1-score for the CASME II dataset is achieved at  $7 \times 7$  blocks, the radius for the spatial plane (XY plane) and the radius for the spatial-temporal planes (XT and YT planes) is achieved at  $(R_x, R_y, R_t) = (1, 1, 4)$ . Whereas the best accuracy and F1-score for the SMIC is achieved at  $5 \times 5$  blocks and  $(R_x, R_y, R_t) = (2, 2, 3)$ .

**Table 3.** The performance of different uniform LBP-TOP features on CASME II and SMIC.

$(R_x, R_y, R_z)$	CASME II			SMIC		
	$3 \times 3$	$5 \times 5$	$7 \times 7$	$3 \times 3$	$5 \times 5$	$7 \times 7$
	Acc (%) / F1-Score			Acc (%) / F1-Score		
(1,1,2)	43.43/0.34	46.88/0.38	45.82/0.38	49.42/0.48	50.48/0.50	45.58/0.45
(1,1,3)	45.27/0.36	47.04/0.39	46.98/0.39	48.62/0.48	49.28/0.49	47.85/0.48
(1,1,4)	46.91/0.39	47.65/0.41	48.24/0.42	47.57/0.47	46.04/0.45	43.25/0.42
(2,2,2)	45.14/0.36	45.32/0.37	46.24/0.39	43.61/0.43	47.32/0.46	49.54/0.51
(2,2,3)	45.43/0.33	44.92/0.32	47.15/0.40	47.76/0.49	51.83/0.51	44.38/0.47
(2,2,4)	42.15/0.32	46.31/0.38	44.31/0.37	45.17/0.43	49.55/0.47	43.47/0.45
(3,3,2)	43.47/0.35	43.28/0.34	45.69/0.37	46.04/0.47	44.71/0.42	47.85/0.49
(3,3,3)	46.38/0.37	46.93/0.39	43.74/0.35	46.41/0.48	42.52/0.41	46.12/0.48
(3,3,4)	47.22/0.41	45.82/0.35	45.49/0.36	45.79/0.44	43.45/0.45	44.91/0.44

##### 4.3.1. Sensitivity of the Threshold $\delta$ to OFF2BD

In this part, the dataset is divided into train and test sets multiple times with different specifications by CASME II and SMIC. Through  $\theta$  the comparison of the micro-expression recognition improvement rate obtained by the various threshold  $\delta$  of multiple models in different dataset, and the robustness and generalization of the model of the OFF2BD algorithm under different thresholds are verified by the curve comparison of the test set and train set maps. LBPTOP and OFF2BD-LBPTOP are used as the benchmark method to compare the experimental algorithm. The improvement rate of the improved algorithm and the baseline algorithm is used as the evaluation index. The two-dimensional curve graph of threshold and improvement rate is fitted by many experiments on the data under different thresholds.

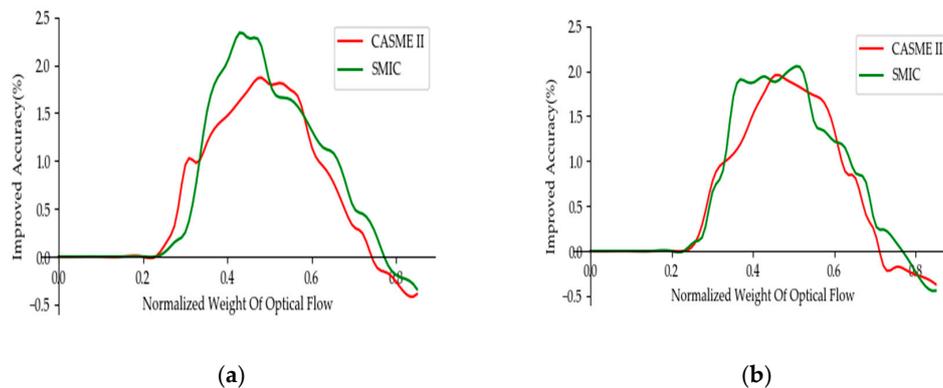
In this experiment, the CASME II and SMIC data are divided into multiple different dimensions of the dataset, and the effect of fitting the training data and test data of each dimension can be obtained, as shown in Figure 4. From the overall perspective, the analysis of the sets of data shows that the test data has a similar performance curve to the train data. In other words, the OFF2BD algorithm model has good generalization ability. From a partial analysis, when the threshold is at  $0.5 \pm 0.1$ , the optical flow processing of OFF2BD algorithm has a good performance improvement. Among them, SMIC peaked around 0.43, and CASME II peaked around 0.5. When the threshold is (0.2, 0.4),

the micro-expression recognition performance gradually increases with the increase of the optical flow threshold, and when the threshold is greater than 0.55, the micro-expression recognition performance gradually decreases. Meanwhile, when the threshold is too high, as a large number of video clips are eliminated, some high-quality clips will be classified into negative domain, which will lead to a negative improvement rate values compared with the baseline. In summary, the OFF2BD algorithm further proves that the two-branch decision method is suitable for solving the uncertainty of image semantic quality.

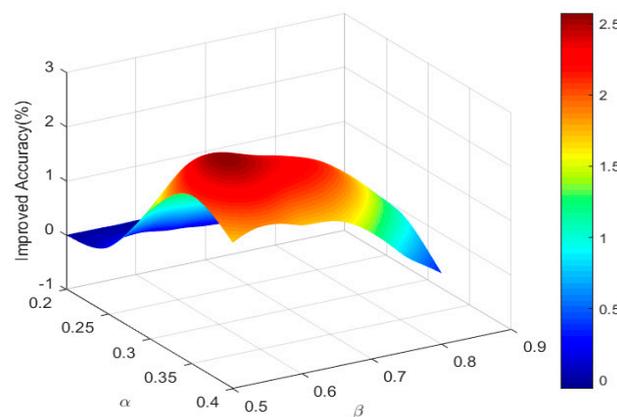
#### 4.3.2. Sensitivity of the Threshold ( $\alpha, \beta$ ) to OFF3WD

Through the analysis of the experimental results of OFF2BD, it is found that the threshold between (0.2, 0.4) has a significant improvement, and the performance gradually decreases after 0.55. The uncertainty domain BND is mainly used to solve the uncertainty of low-quality and high-quality data, and is used as a buffer domain of data to further distinguish the judgment after combining. Based on the above experience, this experiment chooses the threshold  $\alpha \in (0.2, 0.4)$  and threshold  $\beta \in (0.5, 0.8)$  to experiment on the SMIC. Three-dimensional fitting is through the results generated by the OFF3WD algorithm, and the effect diagram is shown in Figure 5.

The  $\alpha$  and  $\beta$  increase naturally according to the specified step length, and the range of the uncertainty domain is adjusted to select the optical flow intensity quality. As shown in Figure 5, the recognition rate of ME under the OFF3WD algorithm changes when the threshold range changes. Among them, it is a better effect improvement when the value of  $\alpha$  is around 0.35 and the value of  $\beta$  is around 0.6. Additionally, the current threshold based on the OFF3WD algorithm also has a better improvement rate on CASME II ME recognition compared to the benchmark method.



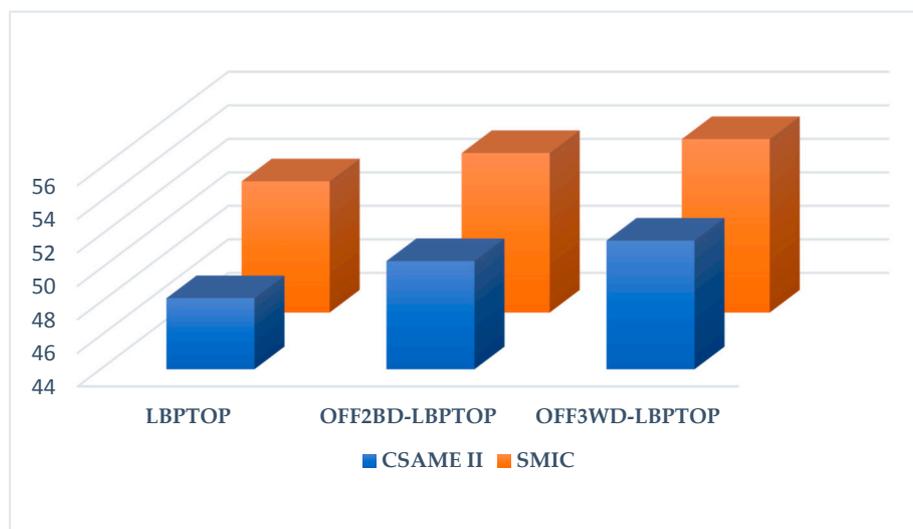
**Figure 4.** The sensitivity of OFF2BD with various thresholds  $\delta$ . (a) Represents the train-set; (b) represent the test-set.



**Figure 5.** The sensitivity of OFF3WD with various thresholds ( $\alpha, \beta$ ).

#### 4.3.3. Comparison with the State-of-the-Art Methods

LBPTOP was reproduced in this experiment and combined with the OFF2BD algorithm and the OFF3WD algorithm to obtain the ME recognition performance score. The experimental results are shown in Figure 6. The effects of OFF2BD-LBPTOP and OFF3WD-LBPTOP are significantly better than the traditional LBPTOP method on the test data. The accuracy has improved on the CASME II and SMIC about 2–3% by OFF2BD compared with the benchmark method. The OFF3WD algorithm has a more accurate judgment on the quality of optical flow semantics, which has increased by 0.5–2% compared with the OFF2BD algorithm. All in all, the experimental data preliminarily shows that the OFF2BD and OFF3WD algorithms can effectively mine the semantic information of micro-expressions and complete the image noise with low morphological semantic quality.



**Figure 6.** Recognition accuracies of different algorithms.

This section compares derivative algorithms (OFF2BD-LBPSIP, OFF3WD-LBPSIP, OFF2BD-STCLQP, OFF3WD-STCLQP, OFF2BD-STLBPIP, OFF3WD-STLBPIP) and original algorithms (LBP-SIP, STCLQP, STLBP-IP) to further verify the improvement of the two filtering algorithms on the ME recognition. Two evaluation criteria (F1-score and Accuracy) were used to evaluate the performance scores. The experimental results are shown in Table 4. The proposed method achieves comparable recognition performances with the state-of-the-art methods on SMIC and CASME II. Compared with LBPSIP, the accuracy of OFF2BD and OFF3WD On CASME II is increased from 49.73% to 52.77% and 53.45%, which is 3–4% higher, and the F1 score is improved from 0.47 to 0.51. The accuracy of SMIC was improved by 2.94% and 4.15%, and the F1 score was improved from 0.46 to 0.5. Compared with STCLQP and STCLBPIP, the accuracy of the two algorithms on CASME II is improved by 2.48%, 2.89% and 2.06%, 1.02% respectively, and 0.82%, 1.39% and 1.79%, 3.90% on SMIC.

Among the feature methods of the current experiment, the two algorithms in this paper have achieved the greatest recognition accuracy and F1 score on STCLQP features. Especially, after three-way decision optical filtering, the recognition accuracy of STCLQP reached 65.41% on SMIC and 61.28% on CASME II, which have higher accuracies and F1-score than the state-of-the-art methods such as BI-WOOF and FHOFO. Simultaneously, the micro-expression feature obtained by the algorithm also has a good improvement effect on the F1-score.

**Table 4.** Comparison with different ME recognition methods.

Method	CASME II		SMIC	
	Acc (%)	F1-Score	Acc (%)	F1-Score
LBPTOP (reproduced)	48.24	0.42	51.83	0.51
LBPSIP [28]	49.73	0.47	47.42	0.46
STCLQP [25]	58.39	0.58	64.02	0.63
STCLBP-IP [24]	59.51	-	57.93	-
BI-WOOF [34]	58.46	0.60	61.29	0.61
FHOFO [37]	60.52	0.61	58.61	0.58
OFF2BD-LBPTOP	50.46	0.44	53.52	0.52
OFF2BD-LBPSIP	52.77	0.51	50.36	0.50
OFF2BD-STCLQP	60.87	0.62	64.84	0.63
OFF2BD-STCLBP-IP	61.57	-	59.18	-
OFF3WD-LBPTOP	51.68	0.46	54.35	0.54
OFF3WD-LBPSIP	53.45	0.51	51.57	0.52
OFF3WD-STCLQP	61.28	0.61	65.41	0.65
OFF3WD-STCLBP-IP	60.53	-	61.83	-

In order to further evaluate the performance of the method in this paper to identify different micro-expressions, we perform optical flow filtering algorithm application and performance testing on the STCLQP features, and use the LOSO cross-validation method to generate a confusion matrix with the best recognition accuracy. Figures 7 and 8 show the confusion matrix of OFF2BD-STCLQP and OFF3WD-STCLQP on SMIC and CASMEII respectively.

On the task of micro-expression detection, STCLQP can easily detect most micro-expression videos from non-micro-expression videos, and the OFF2BD and OFF3WD algorithms select high-quality semantic images, which further deepens the feature expression ability of STCLQP. For the micro-expression recognition on the SMIC database, OFF2BD has a corresponding improvement in the recognition accuracy of “positive” and “negative” classes compared with STCLQP, while the recognition of “surprise” shows a stable trend. OFF3WD further enhances the expression of morphological and semantic features of the “negative” class and make the recognition accuracy of the “negative” category reach 0.7. On the CASME II database, the micro-expression categories have become more complex and diverse, and the CASME II database contains five categories. From the confusion matrix of the CASMEII data set, it is found that the recognition and expression ability of the two algorithms proposed in this paper on STCLQP has generally increased. It shows the best improvement in identifying “happy” and “disgust”, and there are also minor improvements in “depressed”, “surprise” and “other”. Among the five classes, “depression” and “disgust” are the most difficult to identify micro-expression, while the classes of “surprise”, “happiness” and “other” are easy to identify. The algorithm improves the morphological semantic information of different classes, enhances the expression ability of different classes of features, and reduces the probability that the samples of “happiness”, “surprise”, “disgust” and “depression” are wrongly classified as “other” micro expression classes.



**Figure 7.** The confusion matrix of OFF2BD on SMIC and CASME II.

	Negative	Positive	Surprise		Happiness	Disgust	Surprise	Repression	Others
Negative	0.70	0.08	0.22	Happiness	0.64	0.03	0.03	0.04	0.26
Positive	0.25	0.63	0.12	Disgust	0.02	0.48	0.00	0.02	0.48
Surprise	0.22	0.14	0.64	Surprise	0.08	0.00	0.76	0.02	0.14
				Repression	0.19	0.14	0.00	0.43	0.24
				Others	0.04	0.19	0.00	0.00	0.76

Figure 8. The confusion matrix of OFF3WD on SMIC and CASME II.

## 5. Conclusions

We propose an optical flow filtering micro-expression recognition algorithm (OFF2BD and OFF3WD) based on two-branch decision and three-way decision. The algorithm integrates the three-way decision method of rough set theory and adopts the three-way division idea of uncertain affairs, and improves the judgment of POS, BND and NEG domain conditions. The algorithm divides the morphological semantic quality of temporal images by redefining the weight value of optical flow to obtain a more robust and high-quality video images collection. We have seen that these two algorithms have improved their effects on the classic feature extraction methods. For STCLQP features, the optimized recognition rate and F1 score of the test are achieved on the SMIC data set. Optical flow filtering strengthens the feature intensity information on the time axis and eliminates images with low morphological semantic quality.

For future work, we pay more attention to the changes brought about by the facial action areas, and combine the idea of the OFF3WD algorithm to conduct semantic mining on the XY axis and the time axis T to reduce the impact of noise in different dimensions. In addition, not only limited to the method test of different features, we are designing the connection between different classifiers with our algorithm.

**Author Contributions:** Conceptualization, J.W. and J.X.; methodology, J.W., D.L., J.X. and M.T.; software, J.W. and D.L.; validation, J.W., D.L. and J.X.; formal analysis, D.L. and J.X.; investigation, J.W. and D.L.; resources, J.W. and J.X.; data curation, J.W. and J.X.; writing—original draft preparation, J.W.; writing—review and editing, D.L., J.X. and M.T.; visualization, J.W.; supervision, D.L. and J.X.; project administration, D.L. and J.X.; funding acquisition, J.X. and M.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by National Natural Science Foundation of China, grant number 61763031, Jiangxi Provincial Natural Science Foundation, grant number 20202BAB202018, Key Project of Jiangxi Provincial Department of Education Science and Technology Project “Public Security Big Data Analysis for Early Warning and Pre-control” under Grant GJJ191026.

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

## References

- Ekman, P.; Friesen, W.V. Nonverbal leakage and clues to deception. *Psychiatry Interpers. Biol. Process.* **1969**, *32*, 88–106. [[CrossRef](#)] [[PubMed](#)]
- Ekman, R. *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*; Oxford University Press: New York, NY, USA, 1997.
- Li, X.; Pfister, T.; Huang, X.; Zhao, G.; Pietikäinen, M. A spontaneous micro-expression database: Inducement, collection and baseline. In Proceedings of the 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Shanghai, China, 22–26 April 2013; pp. 1–6.
- Yan, W.J.; Li, X.; Wang, S.J.; Zhao, G.; Liu, Y.J.; Chen, Y.H.; Fu, X. CASME II: An improved spontaneous micro-expression database and the baseline evaluation. *PLoS ONE* **2014**, *9*, e86041. [[CrossRef](#)] [[PubMed](#)]
- Davison, A.K.; Lansley, C.; Costen, N.; Tan, K.; Yap, M.H. SAMM: A spontaneous micro-facial movement dataset. *IEEE Trans. Affect. Comput.* **2016**, *9*, 116–129. [[CrossRef](#)]

6. Qu, F.; Wang, S.J.; Yan, W.J.; Li, H.; Wu, S.; Fu, X. CAS(ME)<sup>2</sup>: A Database for Spontaneous Macro-Expression and Micro-Expression Spotting and Recognition. *IEEE Trans. Affect. Comput.* **2017**, *9*, 424–436. [[CrossRef](#)]
7. Pfister, T.; Li, X.B.; Zhao, G.Y.; Pietikinen, M. Recognising spontaneous facial micro-expressions. In Proceedings of the 2011 IEEE International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 1449–1456.
8. Wang, S.J.; Yan, W.J.; Zhao, G.; Fu, X.; Zhou, C.G. Micro-Expression Recognition Using Robust Principal Component Analysis and Local Spatiotemporal Directional Features. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2014; pp. 325–338.
9. Wu, Q.; Shen, X.; Fu, X. The machine knows what you are hiding: An automatic micro-expression recognition system. In *International Conference on Affective Computing and Intelligent Interaction*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 152–162.
10. Khor, H.Q.; See, J.; Phan, R.C.W.; Lin, W. Enriched long-term recurrent convolutional network for facial micro-expression recognition. In Proceedings of the 2018 13th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2018), Xi'an, China, 15–19 May 2018; pp. 667–674.
11. Wang, S.J.; Li, B.J.; Liu, Y.J.; Yan, W.J.; Ou, X.; Huang, X.; Xu, F.; Fu, X. Micro-expression recognition with small sample size by transferring long-term convolutional neural network. *Neurocomputing* **2018**, *312*, 251–262. [[CrossRef](#)]
12. Zhi, R.; Xu, H.; Wan, M.; Li, T. Combining 3D Convolutional Neural Networks with Transfer Learning by Supervised Pre-Training for Facial Micro-Expression Recognition. *IEICE Trans. Inf. Syst.* **2019**, *102*, 1054–1064. [[CrossRef](#)]
13. Hill, A.; Cootes, T.F.; Taylor, C.J. Active shape models and the shape approximation problem. *Image Vis. Comput.* **1996**, *14*, 601–607. [[CrossRef](#)]
14. Edwards, G.J.; Cootes, T.F.; Taylor, C.J. Advances in active appearance models. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–25 September 1999; pp. 137–142.
15. Cristinacce, D.; Cootes, T. Automatic feature localisation with constrained local models. *Pattern Recognit.* **2008**, *41*, 3054–3067. [[CrossRef](#)]
16. Yang, Z.; Leng, L.; Kim, B.G. StoolNet for Color Classification of Stool Medical Images. *Electronics* **2019**, *8*, 1464. [[CrossRef](#)]
17. Leng, L.; Yang, Z.; Kim, C.; Zhang, Y. A Light-Weight Practical Framework for Feces Detection and Trait Recognition. *Sensors* **2020**, *20*, 2644. [[CrossRef](#)]
18. Yang, Z.; Li, J.; Min, W.; Wang, Q. Real-Time Pre-Identification and Cascaded Detection for Tiny Faces. *Appl. Sci.* **2019**, *9*, 4344. [[CrossRef](#)]
19. Zhang, K.; Zhang, Z.; Li, Z.; Qiao, Y. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Process. Lett.* **2016**, *23*, 1499–1503. [[CrossRef](#)]
20. Sun, X.; Wu, P.; Hoi, S.C.H. Face detection using deep learning: An improved faster RCNN approach. *Neurocomputing* **2018**, *299*, 42–50. [[CrossRef](#)]
21. Jiang, H.; Learned-Miller, E. Face Detection with the Faster R-CNN. In Proceedings of the 2017 12th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2017), Washington, DC, USA, 30 May–3 June 2017; pp. 650–657.
22. Ojala, T.; Pietikäinen, M.; Maenpää, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 971–987. [[CrossRef](#)]
23. Zhao, G.; Pietikäinen, M. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 915–928. [[CrossRef](#)] [[PubMed](#)]
24. Huang, X.; Wang, S.J.; Zhao, G.; Pietikäinen, M. Facial micro-expression recognition using spatiotemporal local binary pattern with integral projection. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Santiago, Chile, 11–18 December 2015; pp. 1–9.
25. Huang, X.; Zhao, G.; Hong, X.; Zheng, W.; Pietikäinen, M. Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns. *Neurocomputing* **2016**, *175*, 564–578. [[CrossRef](#)]
26. Huang, X.; Zhao, G. Spontaneous facial micro-expression analysis using spatiotemporal local radon-based binary pattern. In Proceedings of the 2017 International Conference on the Frontiers and Advances in Data Science (FADS), Xi'an, China, 23–25 October 2017; pp. 159–164.
27. Guo, C.; Liang, J.; Zhan, G.; Liu, Z.; Pietikäinen, M.; Liu, L. Extended Local Binary Patterns for Efficient and Robust Spontaneous Facial Micro-Expression Recognition. *IEEE Access* **2019**, *7*, 174517–174530. [[CrossRef](#)]

28. Wang, Y.; See, J.; Phan, W.; Oh, Y.H. LBP with Six Intersection Points: Reducing Redundant Information in LBP-TOP for Micro-expression Recognition. In *Asian Conference on Computer Vision*; Springer: Cham, Switzerland, 2014; pp. 525–537.
29. Guo, Y.; Xue, C.; Wang, Y.; Yu, M. Micro-expression recognition based on CBP-TOP feature with ELM. *Optik* **2015**, *126*, 4446–4451. [[CrossRef](#)]
30. Jia, X.; Ben, X.; Yuan, H.; Kpalma, K.; Meng, W. Macro-to-micro transformation model for micro-expression recognition. *J. Comput. Sci.* **2018**, *25*, 289–297. [[CrossRef](#)]
31. Yan, W.J.; Chen, Y.H. Measuring dynamic micro-expressions via feature extraction methods. *J. Comput. Sci.* **2018**, *25*, 318–326. [[CrossRef](#)]
32. Wang, L.; Xiao, H.; Luo, S.; Zhang, J.; Liu, X. A weighted feature extraction method based on temporal accumulation of optical flow for micro-expression recognition. *Signal Process. Image Commun.* **2019**, *78*, 246–253. [[CrossRef](#)]
33. Wang, S.J.; Yan, W.J.; Li, X.; Zhao, G.; Zhou, C.G.; Fu, X.; Yang, M.; Tao, J. Micro-expression recognition using color spaces. *IEEE Trans. Image Process.* **2015**, *24*, 6034–6047. [[CrossRef](#)] [[PubMed](#)]
34. Liong, S.T.; See, J.; Wong, K.S.; Phan, C.W. Less is more: Micro-expression recognition from video using apex frame. *Signal Process. Image Commun.* **2018**, *62*, 82–92. [[CrossRef](#)]
35. Li, Z.; Xie, N.; Huang, D.; Zhang, G. A three-way decision method in a hybrid decision information system and its application in medical diagnosis. *Artif. Intell. Rev.* **2020**, *25*, 1–30. [[CrossRef](#)]
36. Liang, W.; Zhang, Y.; Xu, J.; Lin, D. Optimization of Basic Clustering for Ensemble Clustering: An Information-theoretic Perspective. *IEEE Access* **2019**, *7*, 179048–179062. [[CrossRef](#)]
37. Happy, S.L.; Routray, A. Fuzzy Histogram of Optical Flow Orientations for Micro-Expression Recognition. *IEEE Trans. Affect. Comput.* **2019**, *10*, 394–406. [[CrossRef](#)]
38. Xia, Z.; Peng, W.; Khor, H.Q.; Feng, X.; Zhao, G. Revealing the Invisible with Model and Data Shrinking for Composite-Database Micro-Expression Recognition. *IEEE Trans. Image Process.* **2020**, *29*, 8590–8605. [[CrossRef](#)] [[PubMed](#)]
39. Lu, H.; Kpalma, K.; Ronsin, J. Motion descriptors for micro-expression recognition. *Signal Process. Image Commun.* **2018**, *67*, 108–117. [[CrossRef](#)]
40. Xu, F.; Zhang, J.; Wang, J.Z. Micro-expression identification and categorization using a facial dynamics map. *IEEE Trans. Affect. Comput.* **2017**, *8*, 254–267. [[CrossRef](#)]
41. Xia, Z.; Hong, X.; Gao, X.; Feng, X.; Zhao, G. Spatiotemporal Recurrent Convolutional Networks for Recognizing Spontaneous Micro-Expressions. *IEEE Trans. Multimed.* **2020**, *22*, 626–640. [[CrossRef](#)]
42. Black, M.J.; Anandan, P. The Robust Estimation of Multiple Motions: Parametric and Piecewise-Smooth Flow Fields. *Comput. Vis. Image Underst.* **1996**, *63*, 75–104. [[CrossRef](#)]
43. Xu, J.; Miao, D.; Zhang, Y.; Zhang, Z. A three-way decisions model with probabilistic rough sets for stream computing. *Int. J. Approx. Reason.* **2017**, *88*, 1–22. [[CrossRef](#)]
44. Yao, Y.Y. An outline of a theory of three-way decisions. In *Rough Sets and Current Trends in Computing*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 1–17.
45. Whitehill, J.; Littlewort, G.; Fasel, I.; Bartlett, M.; Movellan, J. Toward practical smile detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 2106–2111. [[CrossRef](#)] [[PubMed](#)]

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 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/>).