Visual identification of oscillatory two-phase flow with complex flow patterns

Yuqi Huang, Dominique Li, Haoyi Niu, Donatello Conte

Measurement 186 (2021) 110148

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Visual identification of oscillatory two-phase flow with complex flow patterns

Yuqi Huang [a], Dominique H. Li [b,*], Haoyi Niu [a], Donatello Conte [b]

[a] College of Energy Engineering, Zhejiang University, Hangzhou 310027, China
[b] LIFAT Laboratory, University of Tours, 37000 Tours, France

ARTICLE INFO

Keywords:
Two-phase flow
Pattern identification
Optical flow
Offset vector and sequence
Reynolds number classification
Computational fluid dynamics

ABSTRACT

We present an approach based on computer vision and machine learning methods to identify two-phase flow with complex flow patterns in oscillatory conditions. A visualization experiment bench was designed, constructed, and used to simulate the actual reciprocating motion of the cooling gallery inside the piston of low-speed diesel engines. The results of our proposed approach show that the feature vectors extracted from the optical flow images provides a valuable reference for the velocity vectors in two-phase flow. We show that it is possible to identify oscillatory two-phase flow videos with respect to Reynolds numbers from 10568 to 31704 using a Bayesian Network classifier, with the best accuracy of 94%. The approach purposed in this paper can not only be used to present the validating sources for numerical simulation results, but also be widely applied in the visualization of multiphase flow, which is a key area to be developed on the basic research of heat transfer systems.

1. Introduction

Identification and analysis of complex two-phase flow patterns are essential to many engineering domains including chemicals, energy, and automobiles that concern gas–liquid or liquid–solid flows, of which testing, analyzing, and identifying are key factors to understand the working mechanism of components and to optimize the structure and process [1,2]. Visualized experiments have been much introduced to analyze two-phase flows, however, because the two-phase flow interface is complex and the flow patterns are turbulent, the limited clarity caused by experimental conditions degrades the experimental results. For instance, the pictures obtained in visual experiments often contain a large amount of discrete information, which makes it difficult to carry out simple description and identification of two-phase flows [3]. Hence, in recent years, the application of machine learning methods has received a lot of attention in the experimental research of two-phase flow in different fields [4,5].

In automobile industries, as one of the most important in-cylinder components of internal combustion engines, the piston operates under the most severe conditions with the highest thermal load, which requires proper thermal controlling methods to maintain its lifetime and prevent fatigue damages. With respect to the study reported by [6], there are three mainstream methods to cool pistons where the *cooling gallery* with internal oscillating flow has been recognized as the most efficient method for the heat dissipation. Technically, the heat transfer coefficient inside the cooling gallery is the key factor to analyze the piston thermal state so the accuracy of its assignment is therefore directly related to the overall evaluation of the engines' thermal balance. The *two-phase flow* heat exchange process in the cooling gallery definitively affects the piston thermal state, however, due to the special working environment of pistons, including high temperature, high strength, and reciprocating motion, most of the early experimental studies could only measure the temperature of the piston as a calibration basis, but could not directly observe the movement of the two-phase flow in the piston. Hence, many studies have been conducted to investigate the oscillating flows via numerical simulation and experiments, as well as the studies presented in [7,8].

The oscillatory two-phase flow patterns and heat transfer enhancements are closely related to various factors, including but not limited to the size, the shape, the heating state of the container, the surface tension of the medium, and the shear stress between phase interfaces. Therefore, it is difficult to construct a common mathematical model to describe the oscillatory two-phase flow that takes all of the above variables into account. Actually, Computational Fluid Dynamics (CFD) approach, which can study fluid flow and heat transfer of the gallery, has been extensive employed in providing effective assistance for the evaluation of pistons' thermal state and structural design, such as the

* Corresponding author.
E-mail addresses: huangyuqi@zju.edu.cn (Y. Huang), dominique.li@univ-tours.fr (D.H. Li), niuhaoyi@zju.edu.cn (H. Niu), donatello.conte@univ-tours.fr (D. Conte).

https://doi.org/10.1016/j.measurement.2021.110148
Received 17 June 2021; Received in revised form 21 August 2021; Accepted 6 September 2021
Available online 20 September 2021

research word presented by [9,10]. Due to the complicated features of the structure of piston, our previous study presented by [11] shows that most of the existing methods can only be observed from one direction, that is, along the line of sight, the depth of field is 10–20 cm; moreover, the liquid also fluctuates during the movement, which makes it difficult to achieve a high definition. Thus, it is necessary to make the visualization equipment reach a large viewing range, high resolution, and high frame rate at the same time. However, the performance of existing equipment basically cannot meet all these requirements.

Most of the existing approaches are designed to study the two-phase flow with sensors, however, when the fluids' container is moving constantly in a very large area at high speed, the installation of contact sensors is no longer secure, and the quality of flow pattern images would be affected by the oscillation behavior of container, which cause great difficulties for the analysis of internal flow patterns. Although the high-speed Particle Image Velocimetry (PIV) technology can be used to obtain the visualization of the internal field flow in some experiments, for instance the work reported by [12] and [13], PIV and other accurate flow field measurement techniques are also difficult to apply to the experiment presented in this paper due to the experimental design and model clarity limitations. As well, under the reciprocating motion, a crowd of droplets accumulates in the oil cavity, which further affects the visual observation effect. Hence, the flow field therefore has more random characteristics and poor repeatability that make it difficult to compare accurately with the simulation results, and it is therefore necessary to develop new video data analysis methods to conduct an in-depth analysis of the results of visualization experiments, seeking for the relationship between motor speed and internal two-phase flow status, and to provide more stable and reliable verification data for analysis with the CFD results, effectively by machine learning.

In this paper, we present a video-based machine learning application to identify and classify complex oscillatory two-phase flow patterns in the context of the cooling gallery of internal combustion engines. In our study, gas–liquid flow is investigated, of which the liquid is a kind of diesel engine lubricating oil exactly same as the one used in actual engine and the gas is the air from environment, flowing inside the gallery accompany with the oil spray. The reported experiment used the structure, size and stroke of the real piston so that the *Reynolds numbers*[1] concerned in this paper are in the same range as real working conditions. Because the experimental conditions of our study are close to the real working conditions, the clarity of visualization results is actually affected. Therefore, the machine learning methods are introduced to analyze and classify the captured videos, seeking for the relationship between Reynolds numbers and internal two-phase flow status, and to provide more stable and reliable verification data for analysis with the CFD results.

On the other hand, it is precisely because the flow in the cavity is invisible, the local transient temperature is unable to be measured in a real engine, verifying the CFD results becomes a critical issue with extensive applications of numerical simulation method in analyzing the oscillatory two-phase flow in the cooling gallery. The video taken by flow visualization approach can offer the image to observe the flow patterns inside the piston gallery. Therefore, it is interesting to develop the experiments at first, then to simulate the flow in experimental conditions. By comparing simulation and experimental results, the CFD approach can be verified or adjusted accordingly. Only the verified CFD approach can be further applied in analyzing real engine. Based on image processing and machine learning techniques that makes the visualization results of oscillatory two-phase flow with complex flow patterns convinced, the presented method can be summarized as the following 3 steps.

(1) Extract two-phase flow patterns from traced optical flow points of video segments.
(2) Represent two-phase flow patterns as a set of sequences of discretized offsets on the 2D X/Y axis.
(3) Characterize two-phase flow patterns by feature vectors consisting of n-grams generated from offset sequences.

A video classification process on Reynolds numbers is mentioned in order to evaluate the effectiveness of generated feature vectors that identify two-phase flow patterns, of which the mean accuracy is up to 94%. The proposed method has also been successfully applied to verify CFD simulation results on the oscillating cooling gallery. Our experimental evaluation shows that the obtained results could be used not only to depict the motion trends of flow pattern images but also to deeply analyze and excavate the visualization results of oscillatory two-phase flow with complex flow patterns, so that can be widely applied for visualization research on multiphase flow systems.

The reminder of this paper is organized as follows. Section 2 introduces the industrial background of our research. In Section 3, we detail our protocol for experiments on identifying two-phase flow patterns. We present our method in Section 4, which consists of the entire processing chain that generates two-phase flow pattern descriptors. Section 5 reports the video classification assessed evaluation and the CFD verification of our method. Finally, we conclude in Section 6.

2. Background

Current trend in internal combustion engine industry is to increase the power density of engines, and as one of the most important components, the piston suffers a higher thermal load than others. There exist different ways to cope with the increasing thermal load, in which the most effective one is to improve the piston cooling ability where spray cooled, oil jet, and cooling gallery are the most used types. Since the piston head zone suffers the most thermal loads and is the place having the highest temperature, the cooling gallery is indeed the most effective cooling device. Because the oscillated gas–oil flow status in pistons makes a great influence in the cooling efficiency, it is significant to investigate the principles of two-phase flow in such devices.

The oscillatory two-phase flow patterns and heat transfer enhancements are closely related to various factors, including but not limited to the size and the shape of the cavity, the heating state of the container, the surface tension of the medium, and the shear stress between phase interfaces. Many research work has been dedicated to use sensors to study and classify the gas–liquid flow, and most of such work focuses on using the parameters of two-phase flow as features. For instance, [14] used the gamma-ray absorption technique to acquire the feature of the two-phase flow; [15] applied a conductance ring coupled cone meter to estimate individual flow rate of oil–water two-phase flow; in most recent studies, a special implementation method of phase-isolation implemented by [16] and a method based on multi-scale marginal spectrum entropy proposed by [17] were mentioned to rectify two-phase flow patterns. All those studies depend on the sensitivity and accuracy of sensors so the achievement is limited.

The study presented by [18] concerns the two-phase flow generated by mixing compressed air and water, and identifies the flow patterns by using the multi-scale marginal spectrum entropy method to analyze the pressure-difference signal. The results indicate that the signal recognition method is feasible for the identification of gas–liquid two-phase flow patterns, and the pressure-difference signal reflects the characteristics of gas–liquid flow status, therefore signal recognition method is useful and suited. However, in our study, the gas–liquid flow is generated by the oscillation of cavity, thus, not only the fluid, but also the entire cavity is continuously vibrating under particular frequencies. The changing of pressure signal might be dominated by the oscillatory, it is difficult to reflect the changing of flow pattern. The main advantages of signal recognition are fast calculating speed and comparatively

[1] We use the Reynolds number (**Re**) to generalize and to measure the two-phase flow patterns within a cooling gallery.

high accuracy. To verify the results of signal recognition, the images or some other information is required. The signal recognition method is more suitable for solving mature and classic problems, such as the gas–liquid flow in channels. On the contrary, the image recognition method can deal with most visible conditions, but it requires very large amount of image samples, and spends more time and storage spaces.

The CFD approach is widely applied to two-phase flow analysis, it has been used conveniently to achieve the temperature distribution presented by [6], the oil filled ratio presented by [19], and the heat transfer coefficient presented by [20] in a cooling gallery of the piston at different crank angles and speeds. It also helps to anticipate the effects of the geometry and oil supply strategy, such as the work presented by [21], to offer a critical reference for further optimization of engine cooling gallery for vehicle application.

Some visualized experiments adopting high-speed camera have been developed to capture the flow patterns at different oil fill ratio and different engine speed to directly analyze the heat transfer effects, for instance the work presented by [22] and by [23]. [11] presented visualized study of two-phase flow patterns within an open cooling gallery by using a high-speed camera at various crank angles and obtained the correlation formula considering the influence of motor speed, the cavity shape and the material properties to the heat transfer in the cavity. The period-doubling phenomenon was observed in this study when the speed was around 400–500 rpm, which demonstrated the transition from laminar flow to turbulent flow. Meanwhile, the flow state directly affected the heat transfer capability. It indicates the classification of two-phase flow status would be significantly valuable for estimating the intensity of heat dissipation. However, at a high rotation speed, the oscillatory two-phase flow in such a gallery is always in a strong turbulent state that prevents accurate analysis of flow patterns. Compare to previous studies, the cavity structure in our presented test is more complex, the size is bigger, and the piston stroke is longer. Under the action of reciprocating motion the oscillatory two-phase flow in such a gallery (Fig. 2) is always in a strong turbulent state with different motor speeds, a crowd of droplets accumulates in the oil cavity, which further affects the visual observation effect. The flow field therefore has more random characteristics and poor repeatability that make it difficult to compare accurately with the simulation results. It is therefore necessary to develop new video data analysis methods to conduct an in-depth analysis of the results of visualization experiments, seeking for the relationship between motor speed and internal two-phase flow status, and to provide more stable and reliable verification data for analysis with the CFD results, effectively by machine learning.

3. Protocol design

In this section, we detail the design of our experimental protocols, including the experimental apparatus, the calculation of Reynolds number, and the learning and evaluation process to identify two-phase flow patterns.

3.1. Experimental apparatus

We simulate the movement of a piston within a simplified and transparent model of cooling gallery under motor-driven forced high-speed oscillations, so that the two-phase flow patterns can be visualized by a high-speed camera. In our experiments, the design of the apparatus is shown in Fig. 1, which includes two parts: (a) the experiment console that could set different motor speeds; and (b) the visualization facility, in which we use a high-speed camera to get the videos of two-phase flow patterns in the piston. The cavity made by 3D printing technology contains a 1:1 model of the cooling gallery in the diesel engine piston studied in our research, where the cooling medium is continuously sprayed. Note that the transparency of material used by 3D printing is not as good as Plexiglass. The high-speed camera employed in this measurement is Photron Fastcam Mini AX100, which can provide a

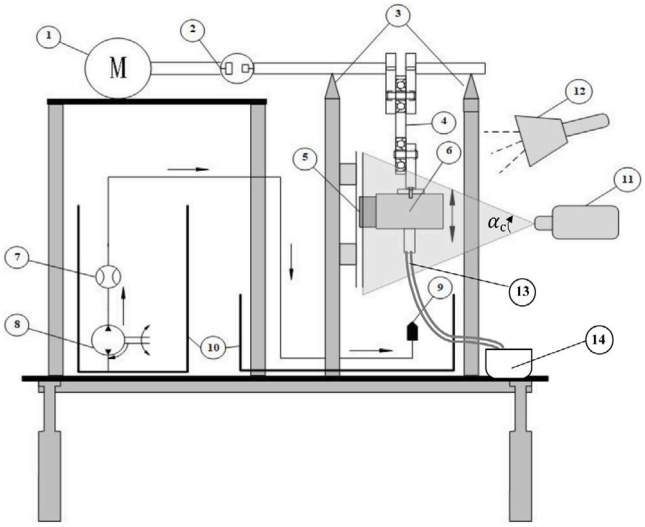

Fig. 1. Design of experimental apparatus: (1) motor, (2) rigid coupling, (3) fixture, (4) crank rod, (5) guide rail and sliding block, (6) cavity, (7) flowmeter, (8) oil pump, (9) nozzle, (10) oil tank, (11) high-speed camera, (12) laser source, (13) collecting pipe, and (14) collecting can.

resolution of 1024 × 1024 pixels at a maximum speed of 4000 frames per second (4000 fps). The playback speed of recorded videos is 30 fps (about 133:1) so we are able to trace the flow patterns.

With such apparatus, we can capture two-phase flow patterns by the high-speed camera and analyze recorded video clips to characterize them. Fig. 2 shows a schematic of the crank rod mechanism. The cavity is driven by the motor to make up and down reciprocating oscillations due to the connecting with the crank and connecting rod. In Fig. 2(a); α and r represent the crankshaft angle and radius respectively. The designed piston stroke (the distance between the TDC and BDC positions of the cavity) is 210 mm. With different motor speed, the oscillation speed of the cavity will change so the liquid oscillates back and forth with the cavity, which will lead to different results of the flow state, and will produce different oscillatory two-phase flow patterns. The oscillatory two-phase flow differs from the tube flow, and the installation of contact sensors is difficult because they are strongly influenced by the oscillating behavior. Besides, because of the oscillation speed is too fast, and it is necessary to clearly capture the flow patterns when the oscillation occurs, the high-speed camera and a laser source were set in front of the device, of which the height is set to the middle of the cavity's oscillatory movements, and the distance between the cavity and the camera is 900 mm. The maximum positive and negative capture angle (α_c in 1) can be therefore calculated by $2 \cdot \tan^{-1}(105/900)$, that is, ±6.65°. The influence imposed to the light and the shadow at the phase interface by the change of capture angle is limit.

The piston studied in this paper is employed in a large marine diesel engine. Unlike gasoline engine, the motor speed of large diesel engine is comparatively low: the working speed is 150 rpm in low-speed engine, about 800 rpm in middle-speed, and maximum 1800 rpm in high-speed engine. But even under the low motor speed, the two-phase flow in gallery is turbulence because of the long moving distance (210 mm). As shown in Table 1, when the motor speed is 400 rpm, the Reynolds number reaches to 21137. The experimental results under such turbulent mode can help to validate the models selected in CFD simulation, such as turbulence model, two-phase flow model, etc. In order to capture the both status in TDC and BDC, the range in photography need to cover the whole stroke, which limits the shooting speed (4000 fps). In further faster motions, the clarity and focus will be affected directly. Besides, the piston module in this test is big (Fig. 2)

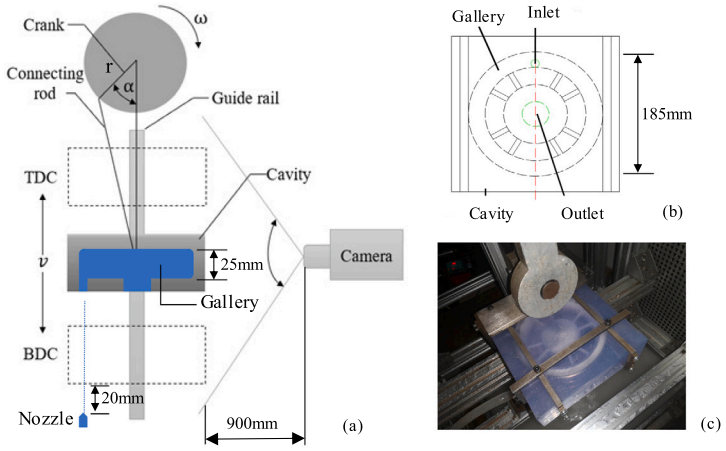

Fig. 2. (a) The movement of the cavity. (b) The design of the cavity. (c) A photo of the standby cavity.

Table 1
Velocity curves of pistons with respect to motor speed.

$\omega(rpm)$	$\overline{v}(m \cdot s^{-1})$	R_e
200	2.31	10568
250	2.89	13221
300	3.46	15830
350	4.04	18484
400	4.62	21137
450	5.20	23791
500	5.77	26398
550	6.35	29052
600	6.93	31704

and heavy (about 1.8 kg), faster speed brings serious noise and the risk of instability. Hence, in our experiments, the oscillation motion starts to be observed in the cavity while the motor speed reaches 200 rpm, so we select the minimum speed as 200 rpm; due to the security concerns, we limit the maximum speed as 600 rpm. Indeed, this setting fits our studies on marine low-speed diesel engines. In order to obtain enough sample data and verify the effectiveness of our proposed flow pattern characterization method, the experimental motor speed gradient is set at 50 rpm. Hence, totally 9 video clips of 30 fps are recorded in our experiments, corresponding to the motor speeds at 200 rpm, 250 rpm, 300 rpm, till to 600 rpm, where each clip lasts about 10 min.

3.2. Calculation of Reynolds numbers

The Reynolds number is a dimensionless constant that can clearly distinguish the patterns and boundary of oscillatory two-phase flow. Based on the studies by [24], the corresponding expression is

$$R_e = \frac{\overline{v} D_e \rho}{\mu}, \tag{1}$$

where R_e is a dimensionless Reynolds number of the cavity under forced oscillation, $\overline{v}$ is the mean oscillating speed of cavity with a diameter D_e, and ρ and μ represent the density and viscosity of the glycerin into the cavity, respectively. The instantaneous oscillation speed v of cavity is defined as

$$v = r\omega(\sin\alpha + \frac{\lambda}{2}\sin 2\alpha), \tag{2}$$

where ω is the motor speed, λ is the connecting rod ratio, and α and r denote the angle and radius of the crank.

In our experiments, we have $\rho = 1260$ kg/m^3 and $\mu = 0.0168$ kg/m s were finally chosen based on the properties of glycerin. Furthermore, the velocity curves of the piston and Reynolds number were calculated as shown in Fig. 3 and Table 1 according to Eq. (1) and Eq. (2).

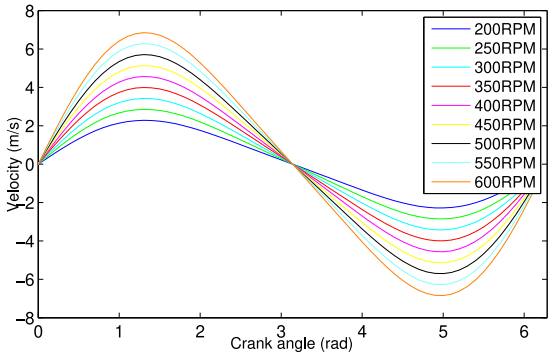

Fig. 3. Velocity curves of pistons with respect to motor speed. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 4 shows examples of nine different oscillatory two-phase flow patterns with respect to different Reynolds number during the experiment analyzed in this paper from the videos at various Reynolds number when the cavity reached the Top Dead Center (TDC). The quality of the figures was severely affected because the semitransparent material and complicated structure weakened the light transmission; the reciprocating motion of the objects requires a large filming range, which decreases the resolution of the photograph; the splashed droplet and bubbles in the cavity further influence the observation. Obviously, the characteristics of a flow field at various Reynolds number is difficult to identify and classify directly because the images shown in Fig. 4 are similar and there are no significant visible differences in the two-phase flow patterns contained in them. However, it is a manual observation result in the static context (single images) instead of the dynamic procedure of the two-phase flow. That is why we use video data captured by high-speed camera to study two-phase flow patterns. Our study and results described in next sections clearly shows that in the dynamic context (videos), the machine learning methods can identify complex two-phase flow patterns, and the mean classification accuracy obtained in our experiments show that our presented method allows to find the differences between different flow patterns, which are represented by different Reynolds numbers.

3.3. Learning and evaluation process

In this paper, we consider that two-phase flow pattern descriptors can be able to characterized by the features used by video classification approaches since such features allow to classify video data

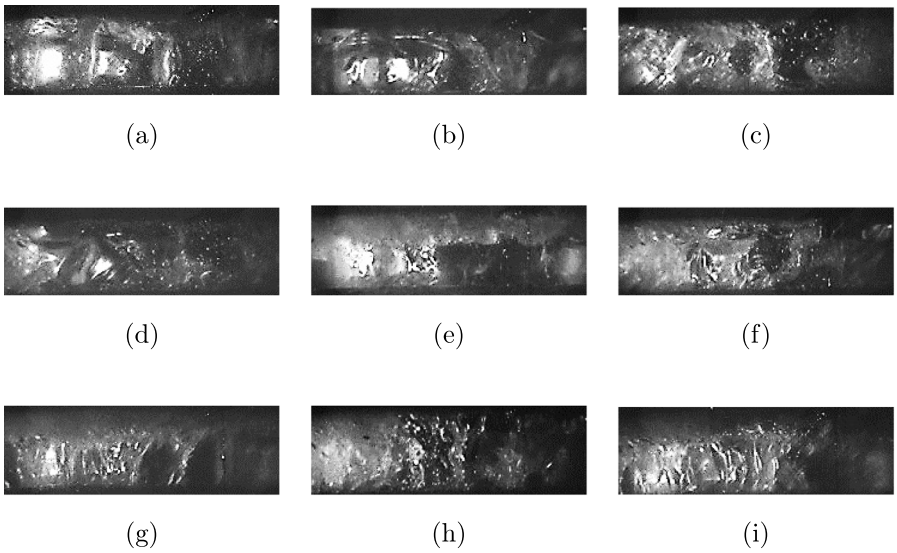

Fig. 4. Analyzed flow patterns under the Reynolds number: (a) 10568, (b) 13221, (c) 15830, (d) 18484, (e) 21137, (f) 23791, (g) 26398, (h) 29052, and (i) 31704. Each image represents the 185 mm × 25 mm cooling gallery inside the cavity.

and their performance can be effectively evaluated by the accuracy of classification results. Note that although deep neural networks, such as, typically, convolutional neural networks (CNN) can accurately accomplish video classification tasks, the iterations inside the multi-layer network make it impossible to exploit learned features in other applications, such as CFD in the context of our research.

The data processing chain of our method is shown in Fig. 5. First, at the step (1), a captured video clip is converted to a sequence of frames; then, at the following step (2), according to the method developed by [25], a set of moving points are extracted by optical flow detection that represent the movements of phase interfaces; at the step (3), we represent the traced moving points by consecutive offsets on the 2D X/Y axis, hence, such offset sequences depict two-phase flow patterns; finally, we extract classification assessed features as two-phase flow pattern descriptors at the step (4). Obviously, an offset sequence is a time series-like sequence of bidimensional numeric values, of which, however, the form is very different than that of time series, as shown in Fig. 6, where 4 randomly selected sample offset sequences are plotted with respect to Reynolds numbers 10568, 21137, and 31704. Therefore, our problem is different than the time series classification problem although there exist many efficient methods with respect to a recent review of [26]. In fact, the two-phase flow pattern characterization problem studied in this paper is rather close to the feature selection problem in general sequence classification surveyed by [27], where the numeric value discretization-based sequential patterns are used as features, such as in the work of [28] and of [29]. However, the data model of traditional sequential patterns does not fit our requirements because of the lack of temporal continuance while defining patterns: undetermined gaps between the elements in a sequence are allowed, which make the description of two-phase flow imprecise. Hence, we finally consider frequent discretized n-grams as two-phase flow pattern descriptors.

The quality of extracted two-phase flow pattern descriptors is measured by N-fold cross validation of video classification as shown in Fig. 7. Traditionally, each video clip is split to N independent segments without overlap, of which $N-1$ segments are used as training set and 1 segment is used as testing set for each fold. Flow pattern descriptors for training and testing the classifier are discretized only based on training data (that will be detailed in the next section) at the step (1), then at step (2) we construct per-segment feature vectors so that any classification methods can be used to evaluate flow pattern descriptors as at steps (3) and (4).

4. Identification of two-phase flow patterns

In this section, we present our approach to two-phase flow pattern identification, which includes optical flow point extraction, time increment interval estimation, segmented motion detection, and the offset vector sequence representation. We finally precise the classification-assessed method used to validate our two-phase flow pattern identification approach.

4.1. Optical flow detection

To detect two-phase flow patterns, we use *corner points* contained in consecutive sampled frames of each video clip and calculate the optical flow for each point by using the method of [25], which is a technique widely used in computer vision. The feature points used in our approach are therefore corner points detected and tracked in frames that are the best for estimating the motions in the video.

The corner points are tracking in order to calculate *optical flow* for these points. Optical flow is the pattern of apparent motion of image objects between two consecutive frames caused by the movement of object or camera, which intuitively reflects the two-phase flow pattern in our study. An optical flow consists of X/Y offset vectors showing the movement of points from the first sampled frame to the second (intuitively reflecting the two-phase flow patterns) to which the Lucas–Kanade method [30,31] is the most widely used one. The Lucas–Kanade method can provide an estimate of the movement of interesting features in successive images of a scene, with some implicit assumptions: (1) two frames are separated by a small time increment Δt (see Section 3.2 for the estimation), in such a way that objects have not displaced significantly; (2) textured objects exhibiting shades of gray which change smoothly. Assume that the increase of brightness per pixel at position (x, y) is respectively $I_x(x, y)$ and $I_y(x, y)$ in X/Y directions, the total increase of brightness after a movement by u and v pixels respectively in the X/Y directions can be calculated by

$$I_x(x, y) \cdot u + I_y(x, y) \cdot v$$

that matches the local difference in intensity which we call $I_t(x, y)$, so that

$$I_x(x, y) \cdot u + I_y(x, y) \cdot v = -I_t(x, y).$$

In global, the goal is to find all those pixels in each frame that represent a corner, that is, a corner lies on a neighborhood in which

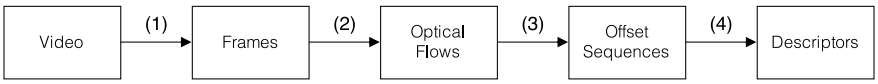

Fig. 5. Extraction of flow pattern descriptors.

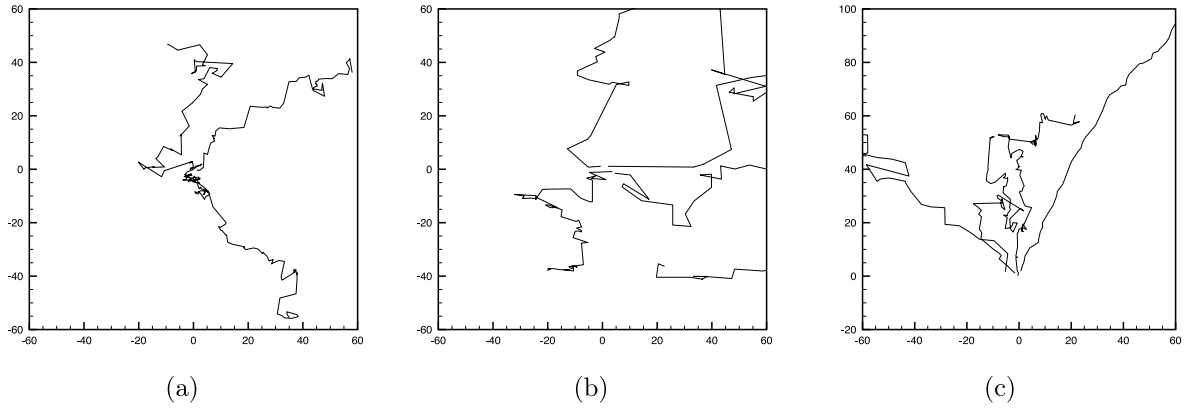

Fig. 6. Sample optical flow points at motor speed (a) $R_e = 10568$, (b) $R_e = 21137$, and (c) $R_e = 31704$. The units of axes are pixel-distance in captured videos.

Fig. 7. Evaluation of flow pattern descriptors.

edges in several directions appear in the image. Notice that the corners detected at the this step are suppressed if they are not a local maximum neighborhood in terms of intensity value, and those with the minimal eigenvalue λ_m less than a threshold are also rejected. Given a corner pixel $p = (x, y)$, its minimal eigenvalue λ_m is computed on the covariance matrix of derivatives centered on p, that is,

$$\lambda_m = \begin{bmatrix} \sum_{S(p)} (dF/dx)^2 & \sum_{S(p)} dF/dx \cdot dF/dy \\ \sum_{S(p)} dF/dx \cdot dF/dy & \sum_{S(p)} (dF/dy)^2 \end{bmatrix},$$

where $S(p)$ is the neighborhood of the pixel p and function dF/dx is the derivative of the frame F at the point with respect to x and to y. Let λ_{max} the maximum value of λ_m over the whole frame F, it is suggested to retain 10% or 5% of the image pixels that have a λ_m value larger than a percentage of λ_{max} [31]. The remaining corners are sorted by the quality measure in the descending order, and finally, the algorithm removes each corner for which there is a stronger corner at a distance less than a threshold. In our approach, the percentage is set to 10%.

By default, Lucas–Kanade method takes a 3×3 patch around an identified point (x, y), so all the 9 points shall have the same motion. Therefore, we can find (f_x, f_y, f_t) for these 9 points (the derivatives along vertical, horizontal axis and along the intensity dimension) so the problem becomes solving 9 equations with two unknown variables which is over-determined, to which a solution is obtained with least square fit method, that is,

$$\begin{bmatrix} u \\ w \end{bmatrix} = \begin{bmatrix} \sum_i {f_{x_i}}^2 & \sum_i f_{x_i} f_{y_i} \\ \sum_i f_{x_i} f_{y_i} & \sum_i {f_{y_i}}^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum_i f_{x_i} f_{t_i} \\ -\sum_i f_{y_i} f_{t_i} \end{bmatrix},$$

where (x_i, y_i) are points belonging to the neighborhood of (x, y). The solution u and w of the above equation is therefore the optical flow vector of the point (x, y), we also call it a *offset vector*.

For each traced optical flow point, the time-ordered list of all offset vectors is an *offset sequence*, so we can represent a video as a set of offset sequences.

4.2. Time increment interval estimation

In this section, we discuss how to estimate a reasonable time increment interval, denoted by N, by establishing the relationship between the physical motion duration, denoted by T, as well as the total number of frames, denoted by F.

In our visualized experiments, the stroke of the guide rail is 210 mm and the shutter speed of our high-speed camera is optimized to 4000 fps that the fluid inside the transparent cavity can be clearly photographed as it moves. With respect to the total number of frames F and the physical motion duration T of the video, the interval of each frame is clearly $\Delta f = T/F$. The velocity of the cooling piston in our experimental video respects multiple sine curves shown in Fig. 3 with respect to our previous work reported by [11], hence, as long as the values of 0 and of π are known, we can estimate the value of N by $N = n/i = (F/T)(1/i)$, where $n = 1/\Delta f$ and i is constant to the extraction of frames at equal intervals that can be constrained as follows: based on our previous analysis, $N \cdot \Delta f$ shall satisfy $\Delta f \le N \cdot \Delta f \le T/2$ so that $2/T \le i \le F/T$.

According to our experiment setup, the ratio of frame number and motion duration is about 30 (frames per second), so the value of frame interval is taken as $N = (F/T)(1/i) = 30(1/i)$, where $i = 2$ ($i = 6$ for *segmented motion*), is finally selected with considering the mean length of extracted offset sequences, and therefore, the Δt mentioned in Lucas–Kanade is finally reduced to be every 15 frames (every 5 frames for segmented motion).

4.3. Segmented motion detection

During our visualized experiments, we noticed the existence of incomplete flow patterns caused by a plastic pipe (the component 13 shown in Fig. 1) that occasionally adheres to the cavity at the bottom position and at the upward stage, shown as Fig. 8, where the red and blue boxes represent the tube and the cavity respectively. The plastic pipe shown in the video is scalable, one side is sticked at the outlet of piston model, and another side is fixed on a collection can to collect the oil through the cavity. Indeed, our experiments confirms the influence of such block of visibility in flow pattern classification, hence, it is important to reduce such influences. In our approach, we segment video clips into four physical motions including *top*, *downward*, *bottom*, and *upward* with respect to the position of the cavity, and each video

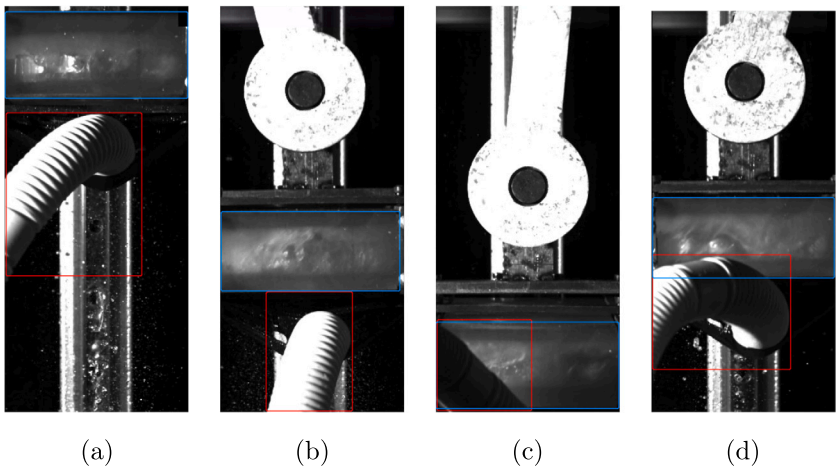

Fig. 8. Motion segments: (a) top, (b) downward, (c) bottom, and (d) upward.

clips contains several cycles of motions, in order to further prove the usefulness and the efficiency of our proposed two-phase flow pattern descriptors: the classification results obtained from different motion segments must be different (which is confirmed by our experimental results shown in Section 5).

Let λ_1^t and λ_2^t be the time points of the physical top position of the cavity between two consecutive cycles, the motion duration per cycle $\sigma = \lambda_2^t - \lambda_1^t$ shall be a constant since the motor speed is fixed. The total number k of cycles contained in a video clip is therefore $k = F/(30\sigma)$. Let the physical top and bottom positions of the cavity between two contiguous cycles be denoted by λ_m^t, λ_{m+1}^t, λ_m^b, and λ_{m+1}^b, where $m \in \mathbb{Z}$ and $1 \leq m < k$, and in considering that the cavity stays at the top and bottom for a short time δ, the physical motion of a video can be therefore divided to 4 segments:

$$T_{top} = [\lambda_m^t - \delta, \lambda_m^t + \delta],$$
$$T_{downward} = [\lambda_m^t + \delta, \lambda_m^b - \delta],$$
$$T_{bottom} = [\lambda_m^b - \delta, \lambda_m^b + \delta],$$
$$T_{upward} = [\lambda_m^b + \delta, \lambda_{m+1}^b - \delta].$$

For instance, before estimating time increment, the video clip of 200 rpm contains 16939 frames, let $\delta = 1s$, we get respectively 1901, 5391, 1291, and 7819 frames for the above 4 segments without counting incomplete cycles. Our experiments show that the *top* and *downward* segments can individually characterize the flow patterns in our cooling gallery model and outperform full length video clips.

4.4. Offset vector representation

Assume that a video clip P contains n traced points of which the movement reflects the gas–liquid border, each point p_i $(1 \leq i \leq n)$ characterizes P, denoted as $p_i \vDash P$. Then, the movement of a point p is identified as an ordered list of offsets (x, y) of contiguous values, that is, an *offset sequence*. An offset sequence depicts a two-phase flow pattern. Since a video clip can contain multiple two-phase flow patterns, in this context, a video clip can be considered as a set of offset sequences. Therefore, to characterize the two-phase flow patterns appearing in a video clip is indeed to feature a set of offset sequences.

The contiguous values contained in offset sequences prevents the collection of common values, which is a necessity of featuring offset sequences. A general way to discretize numerical vector attributes (i.e. the offset vectors) is to use a clustering method to generate symbolic attributes with respect to predefined constraints (for instance, as k-means, the number of attributes) so that a limited number of feature attributes can cover all contiguous values.

We shown in Fig. 9 a comparison between different clustering methods with respect to the clusters generated from all offset vectors gathered from a training video at the motor speed 200 rpm, of which the Reynolds number $R_e = 10568$. Fig. 9(a) shows the spatial distribution of all offset vectors; Fig. 9(b), (c), and (d) respectively show the clusters generated by the k-means method with $k = 100$, by the agglomerative method that stops at 100 clusters, and by the DBSCAN method developed by [32] with $\epsilon = 0.157$ and $MIN = 2$. In fact, the tuning of parameters does not make the application of DBSCAN in our problem. We have tested different parameters of DBSCAN, no meaningful clusters can be generated. The reason is clear: the difference between the central density and the peripheral density is too large. We have also applied DBSCAN to central regions, however, the distribution of offset vectors' values is almost uniformed in this region, as shown in Fig. 9(a). We note that k-means and agglomerative methods discretize offset vectors in a representative way however the resulting clusters obtained by the DBSCAN method are not applicable to our approach.

There is no definitive difference between the results obtained by k-means and Agglomerative methods, hence, we have finally adopted k-means to convert offset vectors to k discretized offset attributes because of its simplicity and wide usage. Although the selection of the k value is essential to k-means, we show (in Section 5) that each k value in a particular range performs as well as each others, and a reasonable range of k values is enough to ensure the representability of feature attributes. By such a k-means based conversion makes an offset sequence discretized to symbolic sequence, where common features can be extracted.

As described in the precedent section, the contiguous movements of two-phase flows are much interesting than a frequent set or a frequent subsequence of all discretized offset vectors, which are very probably discontinued movements. We therefore propose a very simple n-gram based encoding of *feature vectors* that ensures the effectiveness of our method by keeping the efficiency (the counting of n-gram can be done in linear time by n position pointers): given a set S of discretized offset sequences that represents a set P of two-phase flow patterns (i.e. a video), each feature is the frequency of a n-gram g present in a sequence $s \in S$, where g is a n-gram that exists in any (at least one) sequence in the set S.

4.5. Reynolds number classification in video data

According to the above statement, we are two parameters to setup in order to find the best feature vectors to two-phase flow patterns: the k value for k-means and the n value for n-gram. The above task can be resolved by video classification processes.

In our approach, each video clip corresponds to a Reynolds number, that is, a particular class of two-phase flow patterns. Since a set of two-phase flow patterns $P \to C$ can be represented as a set S_k^n of discretized

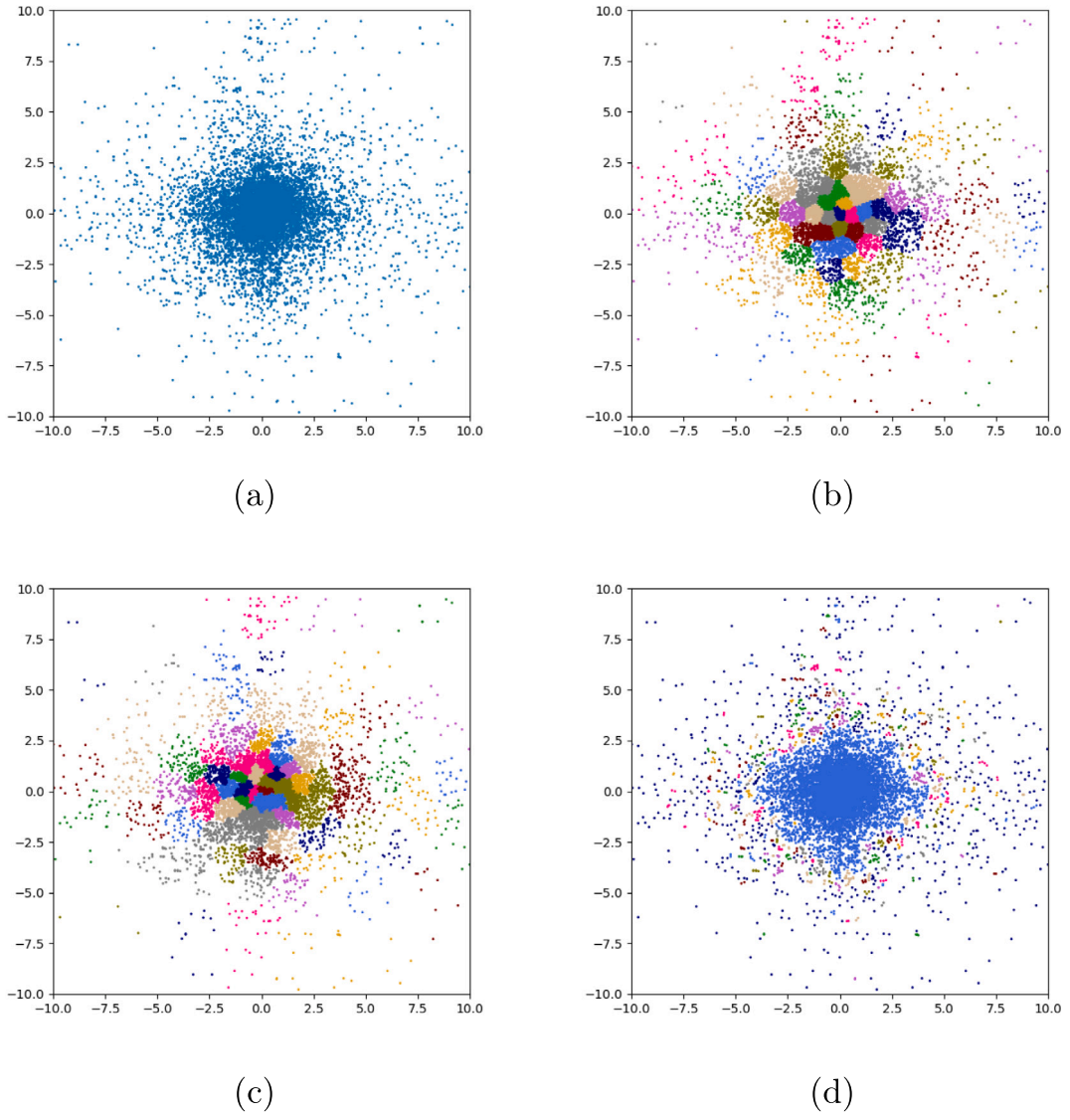

Fig. 9. (a) All offset vectors. (b) The clusters generated by the k-means method ($k = 100$). (c) The clusters generated by the Agglomerative method ($c = 100$). (d) Clusters generated by the DBSCAN method ($\epsilon = 0.157$ and $MIN = 2$). The units of axes are pixel-distance in captured videos. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

offset sequences (where k and n are the parameters of k-means and of n-gram), denoted as $S_k^n \vDash P \rightarrow C$, and since $P \rightarrow C$ is a subjective mapping, we have $S_k^n \vDash C$, that is, the ensemble of all moving points identifies the class C of two-phase flow patterns P. Let $\mathcal{P}$ be a set of two-phase flow patterns represented as video clips and $\mathcal{C}$ be a set of classes, each video clip $P_i \in \mathcal{P}$ is assigned to a class $C_i \in \mathcal{C}$, our goal is to learn a factor $f(k,n) : S \rightarrow C$ where $S_k^n \vDash P$, where $S_k^n \vDash P$. The problem mentioned in this paper can be therefore described as: learn a best factor $f(k,n) : \mathcal{S}_k^n \rightarrow \mathcal{C}$, where $\mathcal{S}_k^n = \{S_k^n \vDash P \mid P \in \mathcal{P}\}$.

We propose a simple strategy to classify offset sequence set represented video data. Let $\mathcal{C}$ be the set of all classes, $\mathcal{S}_k^n$ be a sequence set with respect to the two parameters k and n, h be a classifier such that for each sequence $s \in S_k^n$, $h(s) = C$ where $C \in \mathcal{C}$ is a class. Given a classifier h, we define the *sequence-level score* of a class C with respect to a sequence set S_k^n, denoted by $\sigma_h(C, S_k^n)$, as the number of all sequences $s \in S_k^n$ classified to C, that is,

$$\sigma_h(C, S_k^n) = \frac{|\{s \in S_k^n \mid h(s) = C\}|}{|S_k^n|} \times 100.$$

A class C is called the *dominant class* of a sequence set S_k^n with respect to a classifier h if for all other classes $C' \in \mathcal{C}$ we have $\sigma_h(C, S_k^n) > \sigma_h(C', S_k^n)$; otherwise we say that there is no dominant class so the sequence set S_k^n cannot be classified by the classifier h. Within the typical context of classification, a score is a sequence-level accuracy of a class, which can be calculated from the confusion matrix, and the computational task is to find the best pairwise (k,n) values that maximizes the number of correctly classified videos.

The principal idea of the classification is to validate the features that we extracted from video data that characterize two-phase flow patterns. Although there exist many classifiers, we consider that probabilistic classifiers, such as the Bayesian Network (BN) classifier, shall be suitable to classify offset sequence represented two-phase flow patterns since there are spatial–temporal correlations raised by the characteristics of two-phase flow between a large number of selected features within the offset sequences.

Briefly, the Bayesian Network is a probabilistic graphical model that encodes the probabilistic relationships among the variables of interest. It is a natural way to express causal information and to discover hidden patterns from data. The Bayesian Network avoids using joint probability to reason directly, but uses independent relationship between variables to decompose joint distribution. Bayesian Network

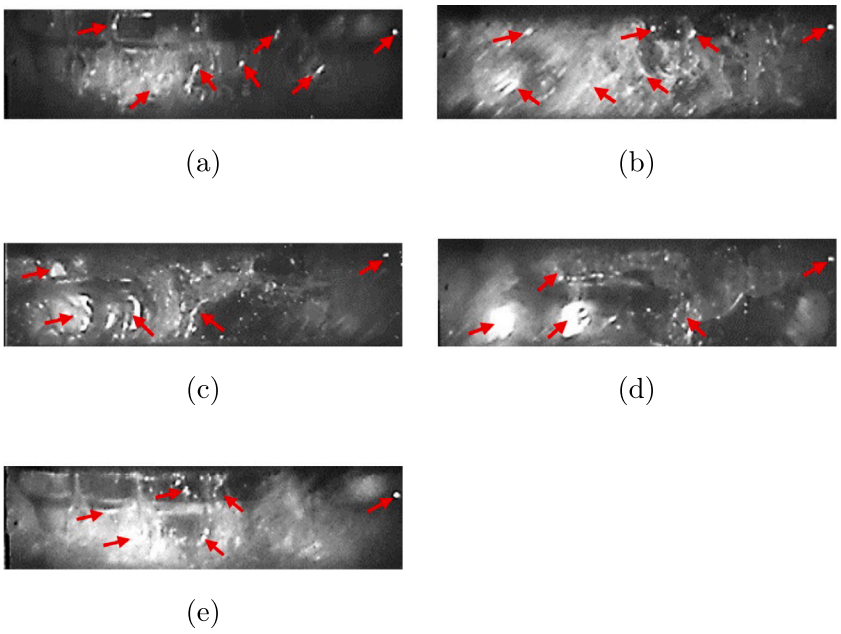

Fig. 10. Oscillatory two-phase flow patterns when $R_g = 15830$ with respect to the angle of crank at (a) 0°, (b) 90°, (c) 180°, (d) 270°, and (e) 360°. Each image represents the 185 mm × 25 mm cooling gallery inside the cavity. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

can analyze the dependence between features qualitatively and quantitatively, and then establish the network structure for probability reasoning. A Bayesian Network consists of two parts: network structure and parameter structure. The network structure is a directed acyclic graph, in which the nodes represent random variables, and the edges between nodes represent the dependence between variables; the parameter structure refers to the local probability dependence, and each node has a probability distribution. In brief, a Bayesian Network is learned to determine a network model that could best represent the dependent relationships of the variables in the data. In our approach, each variable treated by the Bayesian Network is a *n*-gram, so the dependency among *n*-grams can be handled.

On the other hand, the large number of features compose a high-dimensional search space, in which the performance of other classifiers such as the Support Vector Machine (SVM) classifier or the Decision Trees (DT) classifier might be very limited. Our experimental results have confirmed this selection of classifier. Although recent work [33] shows that deep learning on video data can also classify two-phase flow patterns, however the iterations inside the multilayer network make it impossible to exploit learned features in other applications, such as CFD in the context of our research.

5. Experimental evaluation

In this section, we report our experimental results on the visualization experiments, on the Reynolds number classification, and on the comparison with CFD results.

5.1. Visualization experiments

To record videos, we first turn on the motor, then start the shooting at the speed of 4000 fps. In detail, we observe the flowing condition, when the oil flow through gallery is continuous, we start recording the video. Each recording lasts for about 4.5 s, which includes about 45 periods at 600 rpm and about 15 periods at 200 rpm condition. Each playback video lasts for about 10 min at the speed of 30 fps. The oscillatory two-phase flow patterns captured by the high-speed camera under a complete cycle present complex and changeable turbulent flow states.

Fig. 10 shows that a large number of droplets and bubbles interfere with observation, as shown by the red arrow. Indeed, the entire appearance is vague and difficult to identify, which was also noted in some previous studies, presented by [23] and by [22], that abandoned the circular structure and adopted a straight structure cavity for visualization research to obtain clear images. However, as the Reynolds number increases, the image quality decreases significantly. One inevitable reason is that the cavity in the piston is circular, which would affect light refraction; in addition, the cavity's maximum relative velocity reaches 4.37 m/s when the calculated Reynolds number is about 20,000, and the splashed droplets and transient bubbles further damage the picture quality. Therefore, it is not easy to identify flow patterns directly via such pictures, and, nevertheless, the motion of the gas–liquid interface can be distinguished by observation at various angles. Considering that the optical flow algorithm has good flow feature extraction effect in the field of flow visualization, such as the work presented by [34] and by [35], it can be considered that the algorithm can be applied to the oscillatory two-phase flow pattern of continuous frames to generate offset vector.

5.2. Reynolds number classification

5.2.1. Overview

To make this paper less complex, we only report the results of BN and SVM. All classification tasks have been done by the `BayesNet` (the BN classifier) and `SMO` (the SVM classifier with Sequential Minimal Optimization, using polynomial kernel) classifiers under Weka 3.9.3 with default parameters. The results of the DT classifier is very similar to the results of the SVM classifier. The results shown in our experimental evaluation are average values obtained by 5-fold of cross validation where the discretization of offset sequences is strictly independent between training sets and test sets.

Each video clip is first convert to a sequence of frames so that the optical flow detection method can be applied to extract offset sequences, and a Round Robin mechanism is used to distribute all offset sequences extracted from all video clips to 5 independent datasets without overlaps. Then, for each fold, we apply the *k*-means method only to each training set to generate symbolic offset attributes and then all offset vectors in each pair of training and test sets are converted to the nearest symbolic offsets, so all test sets are absolutely excluded

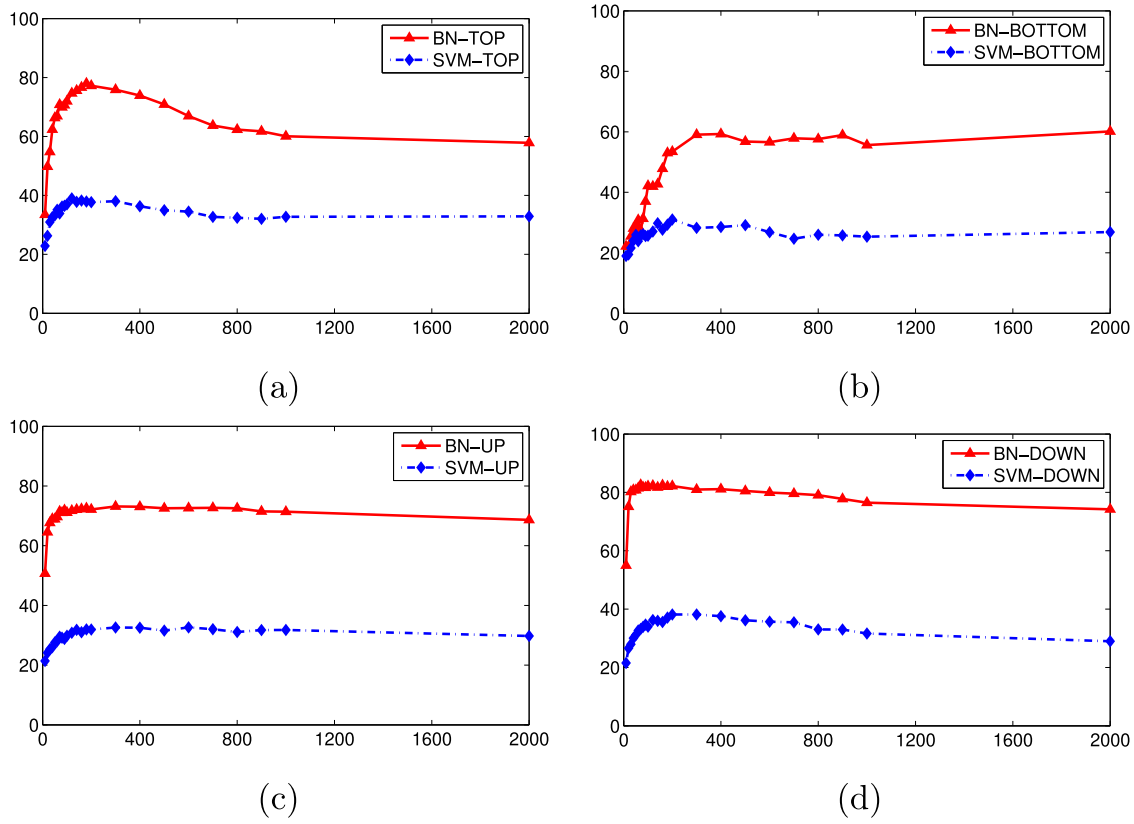

Fig. 11. Reynolds number classification mean sequence-level score (Y axes) and k values (number of features, X axes). (a) Top segments. (b) Bottom segments. (c) Upward segments. (d) Downward segments.

Table 2
Statistical information of full cycle segments (f), top segments (t), downward segments (d), bottom segments (b), and upward segments (u) with respect to the Reynolds numbers. The mean values of Test sets and Train sets are respectively listed.

R_e	$\|S\|_f$	$\|\|S\|\|_f$	$\|s\|_f$	$\|S\|_t$	$\|\|S\|\|_t$	$\|s\|_t$	$\|S\|_d$	$\|\|S\|\|_d$	$\|s\|_d$	$\|S\|_b$	$\|\|S\|\|_b$	$\|s\|_b$	$\|S\|_u$	$\|\|S\|\|_u$	$\|s\|_u$
10568	1028	64154	62	382	22920	60	1158	69480	60	78	4692	60	896	48436	54
13221	355	20864	58	133	6558	49	492	25950	52	148	3724	25	589	30476	51
15830	556	40084	72	311	18191	58	560	30800	55	182	11843	65	630	41774	66
18484	494	32437	65	198	12276	62	744	46128	62	187	11619	62	654	40548	62
21137	469	35685	76	134	7877	58	340	16866	49	87	6333	72	612	39154	63
23791	285	19907	69	127	8340	65	366	21152	57	78	4172	53	328	21561	65
26398	426	36783	86	264	25872	98	537	52626	98	233	22373	96	506	49278	97
29052	386	31563	81	177	10975	62	303	14055	46	107	8377	78	398	33693	84
31704	377	31337	83	178	11228	63	376	22694	60	128	9365	73	424	35612	83
TRAIN	2857	221933	77	1232	84014	68	3051	196336	64	796	54961	69	3498	255081	72
TEST	720	55954	77	312	21347	68	769	49526	64	206	14312	69	878	64154	73

from the training process. Table 2 lists basic statistical information about all sequence sets, where we denote $|S|$ the number of sequences, $\|S\|$ the number of distinct discretized offsets, $|s|$ the mean number of discretized offsets per sequence. We use $\{f, t, d, b, u\}$ to identify different segmented motions, f represents the full-cycle segments; t, d, b, and u represent respectively top, downward, bottom, and upward segments.

5.2.2. Parameter estimation for k-means

A fixed k value is necessary to evaluate the performance of our approach to two-phase flow identification by classification method, since this k value must be applied to both training data and test data. Fig. 11 shows the performances of the BN classifier with unigram represented segmented motions to determine the best k values of the k-means method, where the results of the SVM classifier is shown as a reference. Each unigram is the centroid of one of the k clusters of offset vectors extracted from the offset sequences that represent concerned two-phase flow patterns. All these data show a comparison between mean sequence-level score of Reynolds number classification with different k values, where $k = 80$ and $k = 160$ are two interesting ones. In the curves, we see that there are much more dots in the range $0 < k < 200$ than the range $k > 200$, indeed the dichotomy is used to avoid test every possible k value. The curves also show that the *range* of k values does make the impact, but the sequence-level scores do not depend on individual k values. Besides, the results from the bottom (b) and upward segments (c) confirm the influence of the pipe, and so that the results from full cycle segments are affected, shown in Fig. 12.

Table 3 shows the best k values (number of features) for video-level Reynolds number (from 10568 to 31704) classification accuracy with respect to each segment. In top and downward segments, all the 9 Reynolds numbers are correctly identified and the best score is 100% however the worst one is 15.38% (upward segment of $Re = 23791$). Although the plastic pipe does not appear in all clips of bottom and upward segments, the exceptionally high number of features corroborated the instability of flow identification caused by this pipe.

Table 3
Best k values (**Features**) of video-level classification accuracy of Reynolds numbers from **10568** to **31704** with unigram model and the BN classifier. The values in bold are the best ones in the column.

Segment	Features	Accuracy	10568	13221	15830	18484	21137	23791	26398	29052	31704	Mean
Full	400	8/9	97.70	94.37	51.72	73.56	37.93	**63.16**	64.71	**62.34**	21.33	63.07
Top	180	**9/9**	**100.00**	80.77	92.31	**100.00**	61.54	44.00	97.44	51.43	40.00	77.97
Downward	70	**9/9**	96.91	70.10	**98.97**	**100.00**	73.53	36.99	**100.00**	48.33	**90.67**	**82.54**
Bottom	2000	7/9	66.67	**100.00**	50.00	**100.00**	**94.12**	20.00	73.08	23.81	32.00	60.12
Upward	400	8/9	73.15	91.67	71.30	**100.00**	55.56	15.38	97.03	51.90	64.29	73.03

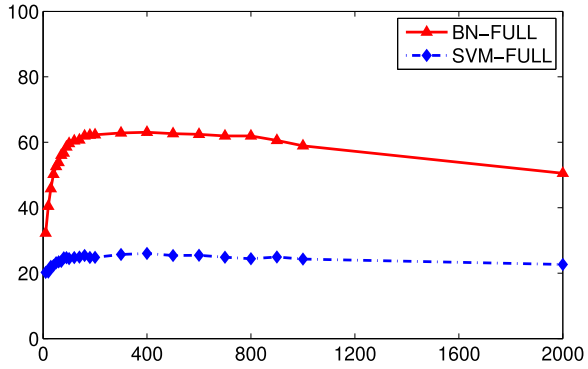

Fig. 12. Reynolds number classification mean sequence-level score (Y axis) and k values (number of features, X axis) in full-cycle segments.

5.2.3. Bigram and frequent multi-gram based evaluation

Based on the above results, we tested bigram and frequent multi-gram based models by two fixed k values, 80 and 160, where $k = 80$ is the best trade-off value on balancing features/accuracy in all segments and $k = 160$ is a reference value. Not surprising, the results obtained by $k = 160$ are much worse than the results obtained by $k = 80$, so it makes no sense to further detail the case while $k = 160$. Nevertheless, we report the results of the SVM classifier in order to enrich the performance of probabilistic models (i.e, Bayesian Network in our research) in two-phase flow identification. The comparative results are shown in Tables 4 and 5.

The number of n-grams generated from N variables is N^n. Let $k = 80$, then 80^1 unigrams, $80^2 = 6400$ bigrams, $80^3 = 512000$ trigrams, etc. can be generated. It is clear that the number of multi-grams cannot be handled while $n \geq 3$, so we apply a *frequency* filter f to reduce the number of features, that is, only the multi-grams whose frequency in the data is superior or equal f, called *f-frequent multi-grams*, are kept as features. The f-frequent n-gram based models are generated by frequency $f \in \{2, 3, 4, 5\}$ and $n \in \{2, 3, 4, 5, 6\}$ where n stands for n-gram. In Tables 4 and 5, f and n are presented as explosion values f/n, for instance, $12259^{2/6}$ denotes that 12259 features obtained by $f = 2$ and $n = 6$. We did not apply the frequency filter to bigram, the number of features based on bigrams are varied because not all bigrams appear in the data. The comparative results show that bi-gram and frequent multi-gram based features improves the final results. In terms of video-level, the mean accuracy of Reynolds numbers classification of all models and segments reaches 89.9% and 94.4% if we exclude the evidently disturbed bottom segments; on sequence-level score of Reynolds numbers classification, our best mean accuracy of all models reaches 81.17% in downward segments (there are common patterns in different Reynolds numbers).

5.3. Comparison with CFD results

The comparison results of the oscillatory two-phase flow patterns acquired by our method and the velocity vectors simulated by CFD are shown in this section.

In this study, the numerical simulations were conducted using the commercial software ANSYS FLUENT 19.0 combined with user defined functions (UDFs) as the main solver. The CFD simulation method used in this study draws on the results of the method used by [23]. The initial boundary conditions set in this paper are described as the following: the injection hole is the flow inlet boundary condition; the flow rate is 0.025 kg/s; the bottom is the pressure outlet and the surface pressure is 0 Pa. Besides, the turbulence model used SST k-ω, and the two-phase flow model used VOF. Mixed grids were generated in the model, the total number of the grid elements was 1.86 million after the grid independence analysis. The transient calculation was adopted and the time step was set as the same as the moving time of 0.5° crankshaft angle.

Figs. 13, 14, and 15 show representative comparison results between the optical flow and CFD at different Reynolds numbers when the cavity at the top dead center (TDC) or the bottom dead center (BDC) positions. Although some differences can be seen between the gas–liquid interface of the images and the volume fraction interface of the CFD results, the vectors obtained by combining images with the optical flow method and compared with velocity vectors by CFD can further support the reliability of the simulation results. The results show that the vortices in the flow patterns (a, red box areas) can be displayed accurately according to the optical flow image (c, yellow box areas) processed by the optical flow algorithm. At the same time, the results of CFD simulation are compared with respect to the flow rules calculated in the corresponding regions (b, black box areas; d, orange box areas), which indicate that the results have a good fitting effect in reflecting the trend in motion of oscillatory two-phase flow patterns.

On the other hand, as we see in these figures, (a) and (b) show the photo and computational volume fraction. In many gas–liquid two-phase flow cases, the gas–liquid interface is very clear in the photo, therefore the simulation can be directly verified by comparing the photo and volume fraction. But in present study, the flow is strong chaotic in a moving part, with the interference of droplets and bubbles, the interface is unclear and difficult to be recognized. The images taken by camera show limited similarity with the computational volume fraction, as exhibited in (a) and (b). Then, we explore the method by comparing the optical flow vectors and simulated velocity vectors of liquid phase, which show better consistence. This phenomenon indicates that the optical flow can partially reflect the dominant moving trends of liquid, but the whole vector diagram is affected by a few factors such as the angle of light source, the reflection of bubble, etc.

6. Conclusion

Visualized experiments have been used for many years as an effective approach to explore two-phase flow that can produce immediate pictures to validate the results of numerical simulation. Normally, the interface between the two phases could be clearly observed in most flow conditions. However, in oscillatory two-phase flow, the cavity moved at high speed with splashed and bubbled internal flow, so the flow patterns cannot be directly identified and classified. This paper discusses the application of computer vision and machine learning methods in a visualized study of oscillatory two-phase flow with complex flow patterns. The results of the oscillatory two-phase flow patterns with nine Reynolds number were considered. We conclude that:

Table 4
Comparison of video-level Reynolds number (**10568** to **31704**) classification accuracy and sequence-level score w.r.t. unigram/bigram/multi-gram (UG/BG/MG) models and BN classifier, where $k = 80$. The values in bold are the best ones in the column.

Segment	Features	Accuracy	10568	13221	15830	18484	21137	23791	26398	29052	31704	Mean
UG_f	80	8/9	85.06	88.73	42.53	58.62	39.08	50.88	52.94	45.45	14.67	56.72
BG_f	5130	8/9	95.40	97.18	50.57	73.56	34.48	50.88	62.35	64.94	22.67	61.37
MG_f	$433^{5/3}$	8/9	95.40	94.37	52.87	78.16	34.48	63.16	62.35	61.04	21.33	62.87
UG_t	80	**9/9**	92.31	80.77	84.62	92.31	61.54	36.00	94.87	60.00	40.00	69.91
BG_t	4097	7/9	97.44	73.08	84.62	89.74	30.77	48.00	89.74	54.29	37.14	69.94
MG_t	$632^{3/2}$	**9/9**	97.44	80.77	89.74	92.31	53.85	48.00	100.00	57.14	62.86	77.53
UG_d	80	**9/9**	96.91	65.98	97.94	100.00	63.24	35.62	100.00	53.33	92.00	81.75
BG_d	4133	**9/9**	97.94	69.07	93.81	98.97	72.06	30.14	94.85	56.67	81.33	79.82
MG_d	$827^{3/2}$	**9/9**	100.00	72.16	93.81	100.00	67.65	32.88	96.91	56.67	88.00	81.94
UG_b	80	4/9	40.00	57.69	3.85	19.23	5.88	0.00	38.46	42.86	28.00	31.29
BG_b	3980	6/9	80.00	73.08	50.00	88.46	23.53	6.67	76.92	71.43	16.00	53.44
MG_b	$891^{3/3}$	8/9	66.67	69.23	61.54	84.62	29.41	53.33	88.46	61.90	28.00	60.40
UG_u	80	8/9	73.15	94.44	67.59	98.15	53.70	20.00	96.04	46.84	66.67	71.05
BG_u	4956	**9/9**	72.22	95.37	68.52	99.07	54.63	29.23	94.06	54.43	66.67	73.90
MG_u	$1764^{3/3}$	**9/9**	75.00	92.59	67.59	100.00	55.56	30.77	96.04	50.63	69.05	74.06

Table 5
Comparison of video-level Reynolds number (**10568** to **31704**) classification accuracy and sequence-level score w.r.t. unigram/bigram/multi-gram (UG/BG/MG) models and SVM classifier, where $k = 80$. The values in bold are the best ones in the column.

Segment	Features	Accuracy	10568	13221	15830	18484	21137	23791	26398	29052	31704	Mean
UG_f	80	2/9	59.77	28.17	14.94	14.94	20.69	8.77	38.82	14.29	16.00	24.69
BG_f	5130	5/9	41.38	32.39	18.39	14.94	21.84	15.79	22.35	20.78	9.33	22.84
MG_f	$433^{5/3}$	**8/9**	25.53	31.03	39.44	17.24	24.14	22.99	22.81	32.94	28.57	22.67
UG_t	80	5/9	61.54	15.38	46.15	41.03	15.38	0.00	56.41	40.00	20.00	36.21
BG_t	4097	4/9	58.97	30.77	46.15	10.26	15.38	24.00	64.10	20.00	17.14	35.56
MG_t	$4871^{2/4}$	6/9	40.36	53.85	38.46	53.85	38.46	19.23	24.00	69.23	37.14	20.00
UG_d	80	6/9	40.21	63.92	43.30	27.84	26.47	0.00	50.52	5.00	6.67	33.90
BG_d	4133	**8/9**	36.08	59.79	44.33	48.45	42.65	21.92	59.79	20.00	24.00	39.19
MG_d	$12259^{2/6}$	7/9	49.48	60.82	50.52	53.61	39.71	16.44	71.13	11.67	22.67	45.99
UG_b	80	3/9	6.67	50.00	46.15	15.38	5.88	26.67	23.08	23.81	20.00	26.16
BG_b	3980	5/9	13.33	46.15	65.38	23.08	11.76	6.67	38.46	33.33	44.00	30.47
MG_b	$1416^{3/6}$	7/9	37.32	40.00	26.92	73.08	38.46	29.41	20.00	50.00	33.33	32.00
UG_u	80	3/9	26.85	62.04	12.04	21.30	14.81	9.23	46.53	18.99	35.71	29.18
BG_u	4956	**8/9**	28.70	62.04	18.52	32.41	24.07	24.62	43.56	32.91	33.33	32.09
MG_u	$9459^{2/6}$	7/9	37.96	50.93	36.11	32.41	29.63	21.54	57.43	31.65	40.48	37.94

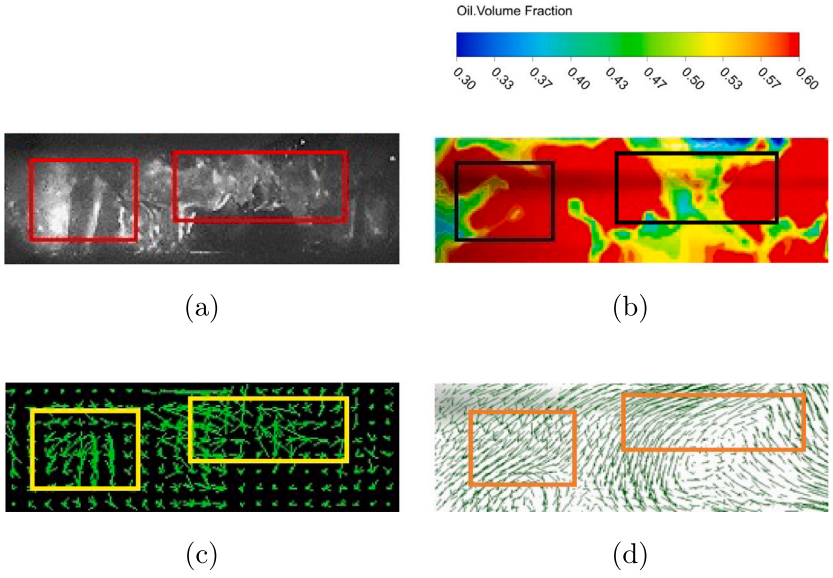

Fig. 13. Comparison of contrast area at TDC position for optical flow images (c, yellow box areas) with volume fraction (b, black box areas) and liquid-phase velocity vectors (d, orange box areas) by CFD results at $R_e = 10568$ (a, red box areas).

(1) Computer vision and machine learning methods were innovatively applied to oscillating two-phase flow pattern identification in providing more valuable analysis for visualization experiment

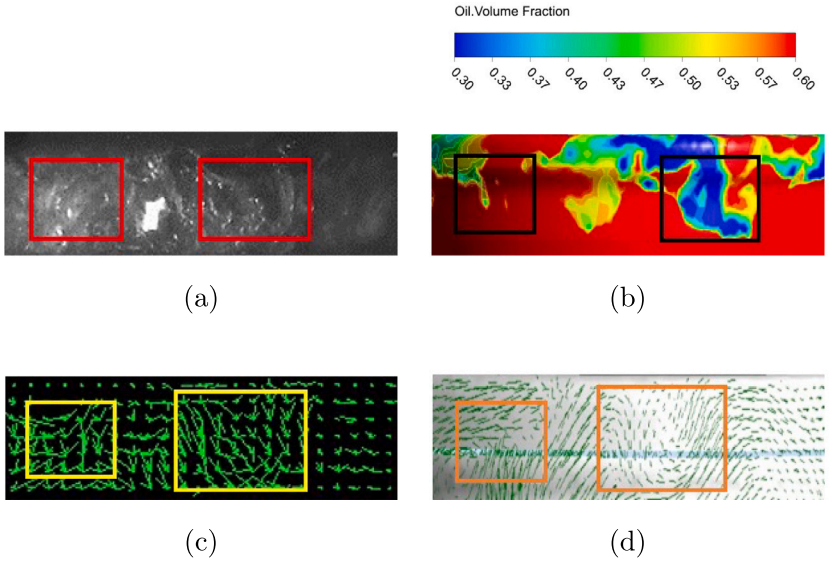

Fig. 14. Comparison of contrast area at BDC position for optical flow images (c, yellow box areas) with volume fraction (b, black box areas) and liquid-phase velocity vectors (d, orange box areas) by CFD results at $R_e = 21137$ (a, red box areas).

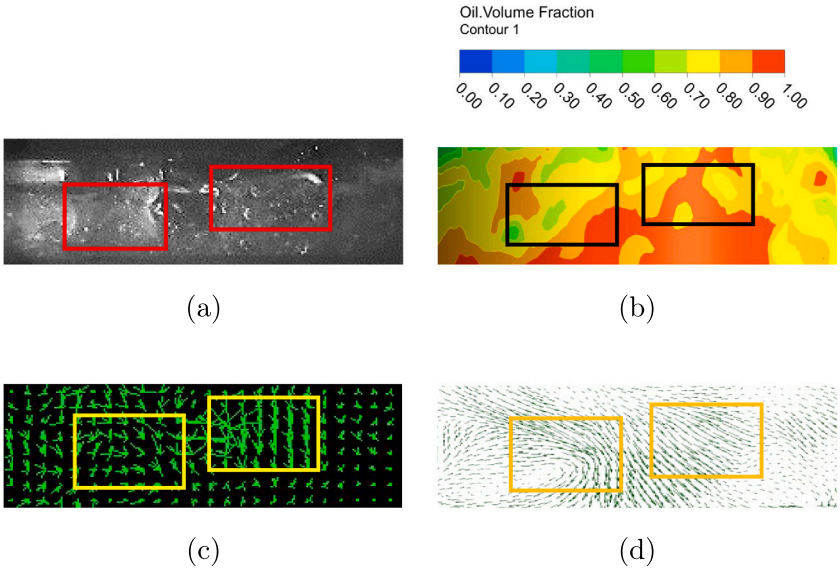

Fig. 15. Comparison of contrast area at BDC position for optical flow images (c, yellow box areas) with volume fraction (b, black box areas) and liquid-phase velocity vectors (d, orange box areas) by CFD results at $R_e = 31704$ (a, red box areas).

results. Our classification assessed experimental results showed the performance of the presented work.

(2) This presented work can help to obtain a clearer motion trend of oscillatory two-phase flow and provide verification for CFD simulation. The vectors obtained by combining images with the optical flow method could be taken as a reference to compare with the velocity vectors from the numerical simulation.

(3) It was found that it is possible to classify, with the mean accuracy of 94%, oscillatory two-phase flow patterns with Reynolds number from 10568 to 31704 using Bayesian network, which confirms the usefulness of a combination of visualization experimental results with machine learning methods.

The method provides a new strategy for non-contact two-phase flow field detection and can deeply analyze and excavate the visualization results of the oscillatory two-phase flow with complex flow patterns with high reliability and application value. The introduced research method can be widely applied in the visualization of multiphase flow which is a key area to be developed on the basic research of heat transfer systems. In our future work, we are interested in understanding and predicting complex two-phase flow patterns by probabilistic models.

CRediT authorship contribution statement

Yuqi Huang: Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Resources, Validation, Writing - original draft, Writing - review & editing. **Dominique H. Li:** Conceptualization, Formal analysis, Methodology, Validation, Writing - original draft, Writing - review & editing. **Haoyi Niu:** Data curation, Formal analysis, Writing - original draft. **Donatello Conte:** Formal analysis, Writing - original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

Dr. Y. Huang would like to thank for the financial support by National Natural Science Foundation of China (52076193).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.measurement.2021.110148.

References

[1] M. Fukuta, S. Morishita, K. Nishihata, M. Motozawa, N. Makimoto, Quality measurement of refrigerant two-phase flow in refrigeration cycles, Flow Meas. Instrum. 77 (2021) 101880.
[2] N. Ghendour, M. Meribout, A. Azzi, Review of measurement techniques for void fraction of two-phase flow through annulus, Measurement 165 (2020) 108196.
[3] L. Wang, S. Li, Y. Yuan, Z. Sun, T. Zhou, Measurement of flow rate in solid-liquid two-phase flow in pipes at low volume concentration with venturimeter, Measurement 138 (2019) 409–415.
[4] E. Khamehchi, A. Bemani, Prediction of pressure in different two-phase flow conditions: machine learning applications, Measurement 173 (2021) 108665.
[5] F. Liang, Y. Hang, H. Yu, J. Gao, Identification of gas-liquid two-phase flow patterns in a horizontal pipe based on ultrasonic echoes and RBF neural network, Flow Meas. Instrum. (2021) 101960.
[6] H. Kajiwara, Y. Fujioka, H. Negishi, Prediction of Temperatures on Pistons with Cooling Gallery in Diesel Engines Using CFD Tool, Tech. Rep. SAE Technical Paper, 2003.
[7] S. Guanghui, J. Dounan, K. Fukuda, G. Yujun, Theoretical and experimental study on density wave oscillation of two-phase natural circulation of low equilibrium quality, Nucl. Eng. Des. 215 (3) (2002) 187–198.
[8] G. Su, D. Jia, K. Fukuda, Y. GUO, Theoretical study on density wave oscillation of two-phase natural circulation under low quality conditions, J. Nucl. Sci. Technol. 38 (8) (2001) 607–613.
[9] X. Deng, J. Lei, J. Wen, Z. Wen, L. Shen, Numerical investigation on the oscillating flow and uneven heat transfer processes of the cooling oil inside a piston gallery, Appl. Therm. Eng. 126 (2017) 139–150.
[10] X. Deng, J. Lei, J. Wen, Z. Wen, L. Shen, Multi-objective optimization of cooling galleries inside pistons of a diesel engine, Appl. Therm. Eng. 132 (2018) 441–449.
[11] X. Yu, D. Yi, Y. Huang, Y. Lu, A.P. Roskilly, Experimental investigation of two-phase flow and heat transfer performance in a cooling gallery under forced oscillation, Int. J. Heat Mass Transfer 132 (2019) 1306–1318.
[12] P. Stansfield, G. Wigley, T. Justham, J. Catto, G. Pitcher, Piv analysis of in-cylinder flow structures over a range of realistic engine speeds, Exp. Fluids 43 (1) (2007) 135–146.
[13] S. Müller, B. Böhm, M. Gleiß ner, R. Grzeszik, S. Arndt, A. Dreizler, Flow field measurements in an optically accessible, direct-injection spray-guided internal combustion engine using high-speed PIV, Exp. Fluids 48 (2) (2010) 281–290.
[14] R. Hanusa, M. Zychb, M. Kusya, M. Jaszczurc, L. Petrykad, Identification of liquid-gas flow regime in a pipeline using gamma-ray absorption technique and computational intelligence methods, Flow Meas. Instrum. 60 (2018) 17–23.
[15] C. Tan, W. Dai, H. Yeung, F. Dong, A Kalman estimation based oil-water two-phase flow measurement with CRCC, Int. J. Multiph. Flow. 72 (2015) 306–317.
[16] Y. Yang, D. Wang, P. Niu, M. Liu, C. Zhang, Measurement of vertical gas-liquid two-phase flow by electromagnetic flowmeter and image processing based on the phase-isolation, Exp. Therm Fluid Sci. 101 (2019) 87–100.
[17] H. Yin, Y. Zhou, J. Zhao, Y. Du, Q. An, Y. Wang, L. Ma, Flow-pattern recognition and dynamic characteristic analysis based on multi-scale marginal spectrum entropy, Appl. Therm. Eng. 146 (2019) 30–38.
[18] H. Yin, Y. Zhou, J. Zhao, Y. Du, Q. An, Y. Wang, L. Ma, Flow-pattern recognition and dynamic characteristic analysis based on multi-scale marginal spectrum entropy, Appl. Therm. Eng. 146 (2019) 30–38.
[19] J. Pan, R. Nigro, E. Matsuo, 3-D Modeling of Heat Transfer in Diesel Engine Piston Cooling Galleries, Tech. Rep, SAE Technical Paper, 2005.
[20] Y. Yi, M. Reddy, M. Jarrett, P. Shyu, C. Kinsey, T. Alcenius, K. Inal, CFD Modeling of the Multiphase Flow and Heat Transfer for Piston Gallery Cooling System, Tech. Rep, SAE Technical Paper, 2007.
[21] W. Sander, B. Weigand, Shaker-based heat and mass transfer in liquid metal cooled engine valves, Int. J. Heat Mass Transfer 52 (11–12) (2009) 2552–2564.
[22] P. Wang, J. Lv, M. Bai, G. Li, K. Zeng, The reciprocating motion characteristics of nanofluid inside the piston cooling gallery, Powder Technol. 274 (2015) 402–417.
[23] J. Lv, P. Wang, M. Bai, G. Li, K. Zeng, Experimental visualization of gas–liquid two-phase flow during reciprocating motion, Appl. Therm. Eng. 79 (2015) 63–73.
[24] J.E. Bush, A.L. London, Design data for" cocktail shaker" cooled pistons and valves, SAE Trans. (1966) 446–459.
[25] J. Shi, C. Tomasi, Good features to track, in: Conference on Computer Vision and Pattern Recognition, CVPR 1994, 1994, pp. 593–600.
[26] A. Bagnall, J. Lines, A. Bostrom, J. Large, E. Keogh, The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances, Data Min. Knowl. Discov. 31 (3) (2017) 606–660.
[27] Z. Xing, J. Pei, E. Keogh, A brief survey on sequence classification, ACM Sigkdd Explor. Newslett. 12 (1) (2010) 40–48.
[28] E. Egho, D. Gay, M. Boullé, N. Voisine, F. Clérot, A user parameter-free approach for mining robust sequential classification rules, Knowl. Inf. Syst. 52 (1) (2017) 53–81.
[29] L. Diop, C.T. Diop, A. Giacometti, D. Li, A. Soulet, Sequential pattern sampling with norm-based utility, Knowl. Inf. Syst. 62 (5) (2020) 2029–2065.
[30] B.D. Lucas, T. Kanade, An iterative image registration technique with an application to stereo vision, in: Proceedings of the 7th International Joint Conference on Artificial Intelligence, IJCAI 1981, 1981, pp. 674–679.
[31] J.-Y. Bouguet, Pyramidal implementation of the affine Lucas Kanade feature tracker description of the algorithm, Intel Corp. 5 (1–10) (2001) 4.
[32] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, et al., A density-based algorithm for discovering clusters in large spatial databases with noise, in: KDD, 1996, pp. 226–231.
[33] L.P. Nguyen, J. Mille, D. Li, D. Conte, N. Ragot, Trajectory extraction and deep features for classification of liquid-gas flow under the context of forced oscillation, in: 15th International Conference on Computer Vision Theory and Applications, SCITEPRESS-Science and Technology Publications, 2020, pp. 17–26.
[34] S. Cai, S. Zhou, C. Xu, Q. Gao, Dense motion estimation of particle images via a convolutional neural network, Exp. Fluids 60 (4) (2019) 1–16.
[35] T. Liu, Openopticalflow: an open source program for extraction of velocity fields from flow visualization images, J. Open Res. Softw. 5 (1) (2017).