HEFAT2011
8th International Conference on Heat Transfer, Fluid Mechanics and Thermodynamics
11 July – 13 July 2011
Pointe Aux Piments, Mauritius

VIDEO BASED CLASSIFICATION OF TWO PHASE FLOWS USING LINEAR DISCRIMINANT ANALYSIS AND EXPECTATION MAXIMIZATION CLUSTERING

Bernd Ameel*, Kathleen De Kerpel, Hugo Canière, Henk Huisseune, Christophe T'Joen and Michel De Paepe
*Author for correspondence
Department of Flow, Heat and Combustion,
Ghent University
Ghent, 9000,
Belgium
Email: Bernd.Ameel@UGent.be

ABSTRACT

Two-phase adiabatic flow of refrigerant R134a was studied and an automated flow regime detection algorithm was developed. For a saturation temperature of 15°C video data was captured for mass fluxes from 100 kg/m²s to 400 kg/m²s in smooth horizontal tubes. The vapour quality varied between 0 and 1, the I.D. was 8 mm. Slug flow, intermittent flow and annular flow were discerned. Several parameters are computed for each video, characterizing the video with both temporal and spatial information. Linear discriminant analysis is used to reduce the dimension of the data. The classification is done by unsupervised clustering with the expectation maximization algorithm. All of the slug and annular flows are correctly recognized, intermittent flows are identified with 95% accuracy.

INTRODUCTION

A lot of research has already been done on automated flow regime detection. Several different principles can be used, but three steps can always be discerned.

Firstly, a signal is obtained from the flow. Tan et al. [1] used the difference in electrical resistance between the two phases of the refrigerant in a technique called electrical resistance tomography. The difference in electrical capacitance was used by Canière et al. [2]. The two-phase flow also produces pressure fluctuations on the tube wall. Liebenberg et al. [3] studied this fluctuation of the static pressure, Sun et al. [4] used the fluctuation of the differential pressure over a Venturi tube. The difference in refractive indices between vapour and liquid phase also allows for optical techniques to be used. By placing a light source and a detector on opposite sides of the tube, Keska et al. [5] studied the intensity of transmitted light. Van Rooyen et al. [6] recorded the flow with a high speed camera and computed the average pixel intensity for each frame. All spatial information contained in the video data is lost. All these techniques clearly produce a signal which varies in time. Several other researchers have also used video data as a basis for the flow regime classification. Both Zhou et al. [7] and Jassim et al. [8] use each frame of a video stream as a different data point. Each frame of a video is interpreted as an independent signal, which contains no time-frequency information.

Secondly, several characteristic parameters are derived from the measured signal. For time-signals several statistical parameters are used, like mean, standard deviation, probability density function, probability distribution function and so on. Time-frequency parameters have been useful in characterizing the signal. Ding et al. [9] compared the Fourier transform, wavelet transform and Hilbert-Huang transform.

Finally, an algorithm classifies the different signals according to flow regime, based on the derived parameters. There is a whole field of research dedicated to this problem. Many different algorithms exist and a lot of them have already been applied to the classification of flow regimes. Mahvash and Ross [10] used a continuous hidden markov model, Zhou et al. [7] classified their data by means of support vector machines. Both of these techniques need a dataset to train the algorithms where the researcher has to provide a dataset with the "correct" classification and hence these methods are called supervised. Rosa et al. [11] studied the use of neural networks, both supervised and unsupervised versions. Unsupervised algorithms have no need for a dataset used for training. Clustering algorithms group the data points together based on a criterion for distance between the points. This was used by Canière et al. [2].

It is apparent that all techniques except for the video frame-based ones use a signal that fluctuates in time, containing time-frequency information, whereas the video frame-based techniques use a single instant in time, containing no time-frequency information. Hence, a logical extension is treating every video stream as whole as a data point and assigning a flow regime to a video stream instead of to a single frame. This allows for using both time-frequency information that has proven to be useful for classification in the other techniques as well as spatial information from the video images. This should allow for a more accurate identification of the flow regime.

NOMENCLATURE

x	[-]	Data vector containing the parameters
X'_1	[m]	First LDA coordinate
X'_2	[m]	Second LDA coordinate
$\mathbf{x}_i$	[m]	Component i of vector **x**

EXPERIMENTAL SETUP

The experimental setup consists of a refrigerant loop, a hot water loop and a cold water loop. In Figure 1 only the refrigerant loop is shown. The preheater is a tube-in-tube heat exchanger. Its length can be varied between 1 and 15 m by adjusting shut-off valves. Hot water flows through the annulus of the preheater, transferring heat to the refrigerant in the inner tube. By changing the length of the preheater (using valves to bypass or connect the different heat exchangers), the hot water inlet temperature and mass flow rate, the vapour quality at the preheater outlet can be set.

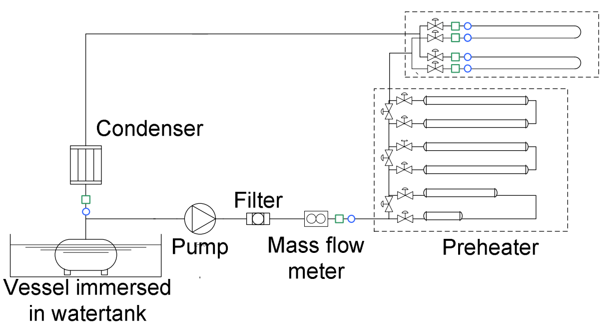

Figure 1 Schematic of the refrigerant loop

After the preheater the refrigerant flows through one of the test sections. Each test section consists of a hairpin with a bend radius R of 11mm and an inner diameter ID of 8mm. The inner tube wall is smooth and the test sections are adiabatic. The section for flow observation consists of a straight tube made out of quartz glass. The flow is filmed with a monochromatic camera (Basler A602f - 640x480 pixels) at 200 frames per second. After the test section, the two phase flow mixture is condensed and subcooled in the condenser, before entering a buffer vessel. This vessel is immersed in an open water bath. The refrigerant saturation temperature is determined by the temperature of the water surrounding the vessel and is set at 15°C for all measurements.

PARAMETER CHOICE

For each set of conditions of mass flow rate and vapour quality at the inlet of the optical test section, a video of 10 seconds is recorded at 200 frames per second. This time is sufficiently long to statistically capture even the slowest phenomena for all the conditions. Therefore, each video stream contains all the data required for flow regime classification. There are 2000 frames, each containing 200 x 300 pixels. This amounts to a dimension of the data vector of 6 million values. This enormous dimension prohibits classification and illustrates the need for choosing characteristic parameters. The entire video stream vector is reduced to a much smaller vector containing fewer parameters. This reduction has to be done with as little loss of useful classification data as possible. Even though algorithms exist to perform this reduction optimally, they are not advised because of the enormous dimension of the data vector. For this reason the characteristic parameters will be chosen manually, based on parameters that have been proven to be effective for classification by other studies. In order to limit the influence of this manual choice on the eventual classification and to reduce the loss of information as much as possible, a sufficiently large amount of parameters is chosen.

The video streams are also manually inspected and assigned to a flow regime. All vectors corresponding to video streams with the same flow regime are said to belong to the same class. It is important to keep in mind that this manual classification is subjective and not entirely accurate. Three different flow regimes were distinguished: slug/plug flow, intermittent flow and annular flow.

Four main items are defined to derive further parameters from. For a fixed pixel position and varying time, an average is made. By doing this for every pixel position, an image is obtained, called the mean image. By taking the standard deviation instead, the standard deviation image is obtained. By taking the average and standard deviation for varying pixel position and fixed time, for all the time instances, two time signals are constructed. These will be called the time variation of the average and the time variation of the standard deviation.

Note that both images still contain 200x300 data points each and both time series still contain 2000 data points each. The dimension will be further reduced by constructing parameters based on these four items.

#	Description
1	maximum of the mean image along the middle column
2	position where this maximum occurs
3	standard deviation of the time variation of the standard deviation
4	mean of the time variation of the standard deviation
5	minimum of the mean image along the middle column
6	position where this minimum occurs
7	standard deviation of the time variation of the average
8	mean of the time variation of the average
9-16	energy fractions of the 3D wavelet packet components of the time variation of the average

17	value of the central autocorrelation peak components of the time variation of the average
18	width of this peak
19	percentage of the total frequency content of the time variation of the average that is below 10 Hz
20	frequency with maximal amplitude
21-28	energy fractions of the 3D wavelet packet components of the time variation of the standard deviation
29	value of the central autocorrelation peak of the time variation of the standard deviation
30	width of this peak
31	degree of symmetry of the mean image
32	mean of the time variation of the degree of symmetry
33	standard deviation of the time variation of the degree of symmetry
34	maximum - minimum of the time variation of the degree of symmetry

Table 1 List of selected parameters

Now every video is summarized in a 34 dimensional data vector. If the choice of parameters is sufficient, the data vectors will contain enough discriminatory information based on the original videos to be able to separate the different flow regimes.

Of course a lot more parameters can be chosen. Furthermore, several of these parameters are most probably strongly correlated. For example, the energy fractions of the mean signal and the standard deviation signal. Strongly correlated parameters increase the dimension of the parameter vector without adding any new information. Before doing the classification, as an intermediate step these parameters will be eliminated from the parameter space. This is done using principal component analysis (PCA) [12]. This allows for parameters to be added without having to be concerned about possible correlation with previously chosen parameters, thus without increasing the dimension unnecessarily.

Another important remark is that the possible spread of all the variables is different. For example, pixel intensity can theoretically vary between 0 and 255, whereas energy fractions can only assume values between 0 and 1. In order to give these parameters the same weight in the classification algorithm, all parameters are linearly scaled so that their maximum and minimum over a large dataset are the same. The maximum is chosen equal to 100, the minimum to -100. The actual values are of no importance, the relative scaling of the parameters is what matters. The dataset used for this scaling has to be representative for all possible situations. This way a test vector not included in the dataset will also be scaled in a sensible way.

In order to classify the different data vectors, techniques from image recognition will be used. Both PCA and linear discriminant analysis (LDA) [13] are widely used for this purpose [14].

PRINCIPAL COMPONENT ANALYSIS (PCA)

PCA centers the dataset and aligns the axes along the directions of maximum spread of the dataset. By removing the axes corresponding to small global spread, a projection is obtained that reduces the dimension of the dataset while conserving as much global spread as possible.

The relative scaling of the axes of the original system will have a strong influence on the axes of the new system.

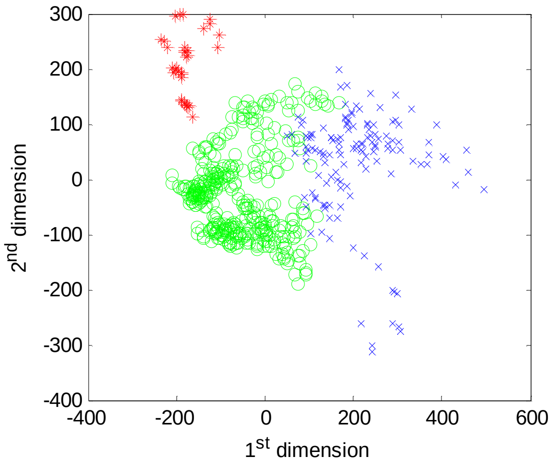

Figure 2 PCA reduction to two dimensions.
o intermittent, × annular, + slug

Figure 2 shows the result of PCA dimension reduction to two dimensions. All the 34-dimensional vectors are projected on the two dimensions corresponding to the maximal global spread. All the vectors of the same class are grouped together in the projected space. This gives confidence in the choice of parameters; they contain enough data from the original video streams to observe a difference between the classes. Without the manual classification though, it would not be clear from Figure 2b where to draw lines to separate the classes, there's no clear distinction between the classes. An unsupervised classification algorithm would not suffice in the 2D PCA space. Furthermore, a line separating the intermittent and the annular classes would have to be quite curved. This makes overfitting likely in case a supervised classification algorithm would be used.

A supervised algorithm can use the knowledge provided by manual classification of a training data set to obtain a new coordinate system that is better able to separate the different classes than PCA. PCA results in maximal conservation of global spread, but it is the spread between the classes which is important.

LINEAR DISCRIMINANT ANALYSIS (LDA)

Linear discriminant analysis finds the directions for which a certain criterion is maximized. For the most basic form of LDA this is the Fisher discriminant. The Fisher discriminant is a fraction where the numerator is a measure for the spread between the classes, and the denominator is a measure for the spread within the classes, in the projected LDA space. In the literature it is proven that the LDA space is at most K-1 dimensional, with K the number of classes. For the case K=3, as studied here, this means the LDA space will be two-dimensional.

The equation for the projection requires the inversion of a matrix which is singular, because of high dimensionality compared to the amount of training data. This is called the small sample size problem, for which several solutions exist. Since the dimensionality of the vectors is limited, calculating the pseudo-inverse suffices [15].

Using the entire dataset as training it will become apparent whether LDA is sufficient to separate the different classes, or whether a more complicated nonlinear function is necessary instead of the linear Fisher discriminant.

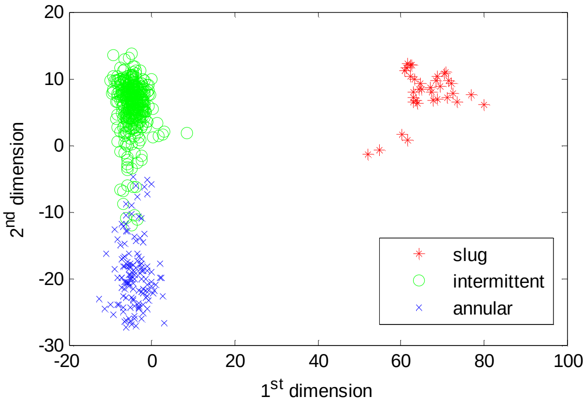

Figure 3 LDA projection using entire dataset as training data. ○ intermittent, × annular, + slug

Figure 3 shows the projection of the data vectors on the LDA space. The abscissa and ordinate have no meaning as they depend on the scaling of the parameters, which is arbitrary. Given that the transition between intermittent and annular flow is gradual, there will never be a clear distinction between these two classes. With this in mind the separation between the classes using LDA is more than adequate, and it is concluded that a linear criterion suffices to separate the classes. Three different clusters can be distinguished.

Starting from N=34 different parameters LDA constructs two new parameters that are linear combinations of the original ones, in such a way that these two new parameters optimally separate the classes.

$$X_1' = \sum_{i=1}^{N} a_i \boldsymbol{x}_i$$

$$X_2' = \sum_{i=1}^{N} b_i \boldsymbol{x}_i$$

For each parameter the value $\sqrt{a_i^2 + b_i^2}$ indicates the contribution of the parameter to the LDA parameters. This value will be referred to as the parameter weight. Variables with quite small values for the contribution can be left out without large changes in the solution, since all parameters are scaled to vary between the same minimum and maximum.

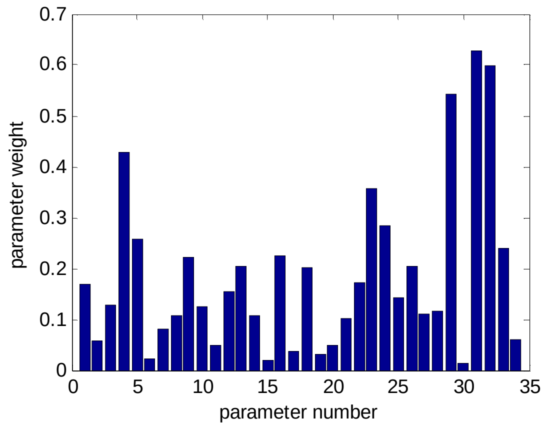

Figure 4 Parameter weights

From Figure 4 it is apparent that parameters 6, 15 and 30 can be omitted as they have very little influence on the LDA parameters.

It is important to keep in mind that the coefficients a_i and b_i depend on the chosen training set. Both directly, since the training set determines the LDA projection**,** as indirectly, since the training set also determines the relative scaling of all data vectors. This implies that the influence of the training set on the final classification will need to be checked.

With the chosen set of parameters a reduction in global spread of 1% using PCA prior to using LDA has a quite detrimental effect on the separation, as shown on Figure 5. This corresponds to a dimension reduction to only 15 parameters. This means that the parameters are strongly correlated. When the global spread can only be reduced by 10^{-6}, the dimension is reduced to 30. In this case there is no visible influence on the LDA projection. This high sensitivity to the reduction in global spread in the prior PCA step is caused by the singularity of the LDA transformation [16].

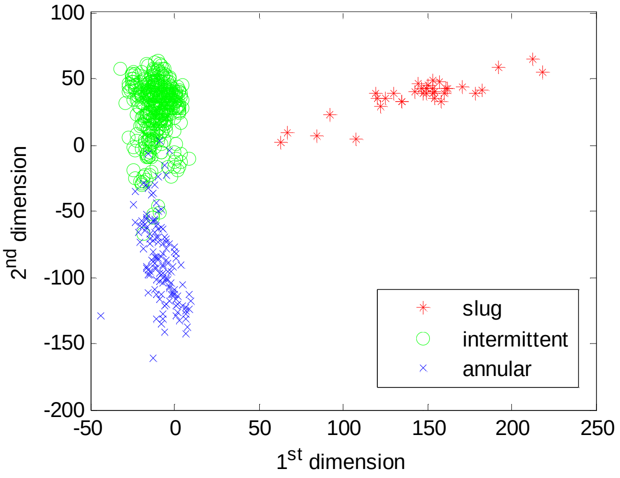

Figure 5 LDA projection after PCA dimension reduction to 99% global spread. O intermittent, × annular, + slug

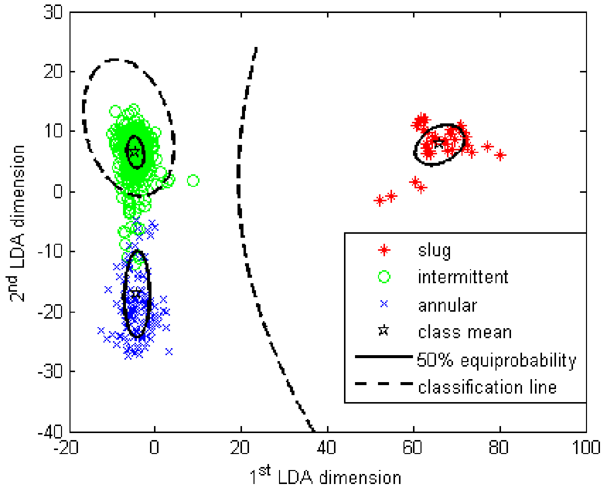

Figure 6 LDA space with EM clustering O intermittent, × annular, + slug

EXPECTATION MAXIMIZATION CLUSTERING (EM)

For the final classification, it is necessary to separate the LDA space in different regions, where each region corresponds to a flow regime. There is a multitude of techniques available for this, both supervised and unsupervised. Many of these techniques have already been applied to two phase flow identification problems, among others support vector machines [7] and neural networks [11] for the supervised algorithms. Unsupervised clustering algorithms like c-means clustering have also been applied [2]. In order to achieve more objective flow regime identification, an unsupervised algorithm is preferred. A supervised algorithm can learn to perform the same subjective classification as was used in the training set. An unsupervised algorithm is not influenced by subjective manual classification of a training set.

Seeing as though several clusters can indeed be identified in the LDA space, an unsupervised clustering algorithm is applicable for the final classification of the data. Since the transition between the different flow regime classes is gradual, each data vector will be assigned a membership grade to each of the flow regime classes. The algorithm that will be used is called expectation maximization [17]. EM finds the multivariate Gaussian distributions that best explain the observed dataset. With a given amount of clusters, EM will identify the different clusters and determine the mean and covariance matrices for each cluster. There is no need to choose any parameters other than the amount of clusters to be identified, which is why EM was selected.

The clusters found by EM are illustrated by the mean and the 50% equiprobability contour. For each point in LDA space the probability of this point being generated by each of clusters can be calculated. The dashed lines represent the points where these probabilities are equal for two classes. These lines effectively separate the clusters. A small amount (6.6%) of vectors which were manually identified as corresponding to intermittent flow is erroneously identified as annular flow. All slug flow and all annular flow data vectors are correctly classified. This shows that the combination of LDA and EM clustering is capable of separating the different flow regimes.

INFLUENCE OF TRAINING DATA SET

Once the scaling and LDA projection has been determined from the training set, test vectors can be scaled and projected on the same space. Then both test vectors and training vectors are clustered by the EM algorithm, and a flow regime is assigned. If the training dataset is representative for all possible test vectors and there is no overfitting, the classification assigned to the test vectors will be accurate.

The subjective influences are limited to the choice of parameters and to the determination of the LDA projection through the manual classification of a training data set. The choice of parameters directly influences the eventual classification because it determines which discriminatory information available in the raw video data is kept and which is discarded. By choosing a large amount of parameters and then reducing the dimension, the influence of this parameter choice is reduced. Because of the simple projection on a hyperplane, as opposed to on more complicated surfaces, the risk of overfitting during the dimension reduction is reduced. This means the influence of noise on the raw data, such as erroneous manual classification and measurement error is reduced.

However, because of the enormous dimension of the original data, the training set will necessarily be so small that removing a few training vectors from the dataset will influence the projection (small sample size problem). Hence, the effect of the specific vectors in the training set on the projection and on the final classification is studied.

The relative scaling of the different parameters is determined by the training dataset and hence influences the projection indirectly. There is also a direct influence, since the projection is the one that optimally separates the training dataset. This implies that it is necessary to verify that the classification of a random test vector does not strongly depend on the specific vectors included in the training data set. Ideally a separate validation dataset would have to be used to validate the algorithm. However due to the time needed to perform each measurement this was not deemed feasible. Hence 20% of the full dataset is randomly selected and used as a test data, the remaining 80% is used as the training data. The LDA projection is determined using the training data. The test data is then projected and classified using the clustering algorithm. The fraction of correctly classified test vectors is then calculated. This process is repeated 30 times, the result is shown in Figure 7.

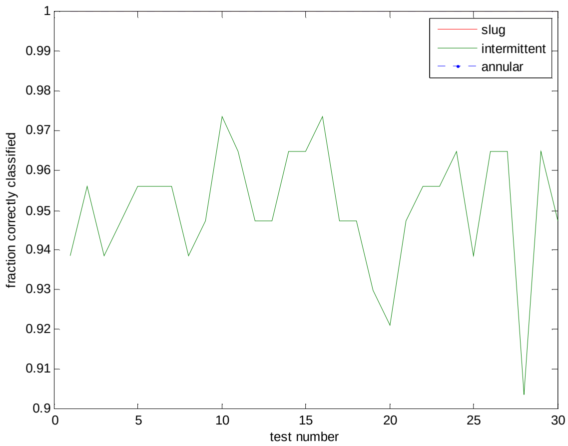

Figure 7 Fraction of correctly classified test vectors

Figure 7 shows that the classification is quite good. On average 95% of all intermittent flow regimes are classified as such, slug flow and annular flow are always correctly identified. The classification does indeed depend on the specific vectors included in the training dataset, but this dependency is weak. With a larger training dataset this dependency would be further reduced. In the worst case 10% of the intermittent flows are incorrectly classified, which considering the gradual transition of intermittent to annular flow is still a very good result.

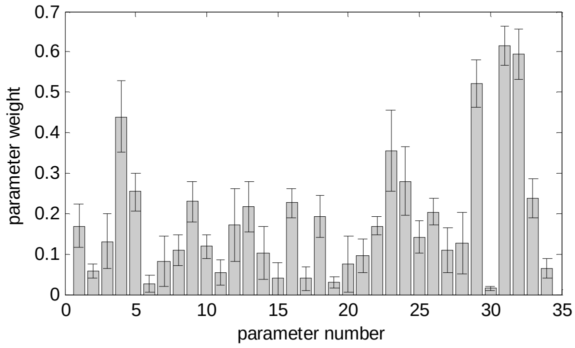

Figure 8 Variation of parameter weights for different training sets

The composition of the training set determines the LDA projection. Hence the parameter weights are examined for each of the tests. The average parameter weights are shown on the bar plot of Figure 8, the error flags correspond to twice the standard deviation. It is apparent that there is a substantial variation in the parameter weights. This corresponds to a variation in the LDA projection. Even though the projection itself varies, from Figure 7 it can be seen that this has no severe impact on the results of the classification. Apparently the separation of the different flow regimes in LDA space is sufficient for the different projections.

Even though there is variation in the parameter weights, several parameters are consistently more important than others. The average degree of symmetry (32) and the degree of symmetry of the average flow (31) are important parameters. On average annular flows are more symmetric than intermittent flows, which are more symmetric than slug flow. Even when not combined with other parameters, these parameters contain valuable information. On the other hand, the position of minimum pixel intensity for the mean image (6) is consistently not useful. This position does not vary much for the different flows and hence is a bad parameter. The width of the autocorrelation peak (30) is another parameter that does not contribute to the classification. It is theoretically possible that a parameter only becomes significant in combination with other parameters. This means that it is not possible to conclude that a single parameter will always be useless based on the parameter weight.

CONCLUSION

An automated flow regime classification algorithm was developed which used a video stream as input data. The video stream is converted into a set of data by considering spatial and temporal information separately. A set of parameters is then calculated based on these data series. These parameters have been manually selected. To reduce the dimension of the parameter space, LDA was used, combined with EM clustering for the final classification. . This allows for an objective identification of the flow regime. The algorithm works well, on average 95% of the classifications for intermittent flow correspond to the visual classification, for slug and annular flow this is 100%.

REFERENCES

[1] C. Tan, *et al.*, "Identification of gas/liquid two-phase flow regime through ERT-based measurement and feature extraction," *Flow Measurement and Instrumentation,* vol. 18, pp. 255-261, Oct-Dec 2007.

[2] H. Caniere, *et al.*, "Probabilistic mapping of adiabatic horizontal two-phase flow by capacitance signal feature clustering," *International Journal of Multiphase Flow,* vol. 35, pp. 650-660, Jul 2009.

[3] L. Liebenberg and J. P. Meyer, "Objective classification of two-phase flow regimes," *Heat Transfer Engineering,* vol. 29, pp. 1-2, 2008.

[4] B. Sun, *et al.*, "Analysis of void fraction of Gas-liquid Two-phase Flow on the Hilbert-Huang Transform," *2008 7th World Congress on Intelligent Control and Automation, Vols 1-23,* pp. 4960-4963
9402, 2008.

[5] J. K. Keska and B. E. Williams, "Experimental comparison of flow pattern detection techniques for air-water mixture flow," *Experimental Thermal and Fluid Science,* vol. 19, pp. 1-12, 1999.

[6] E. van Rooyen, *et al.*, "Probabilistic flow pattern-based heat transfer correlation for condensing intermittent flow of refrigerants in smooth horizontal tubes," *International Journal of Heat and Mass Transfer,* vol. 53, pp. 1446-1460, 2010.

[7] Y. L. Zhou, *et al.*, "Identification Method of Gas-Liquid Two-phase Flow Regime Based on Image Multi-feature Fusion and Support Vector Machine," *Chinese Journal of Chemical Engineering,* vol. 16, pp. 832-840, Dec 2008.

[8] E. Jassim, *et al.*, "Probabilistic determination of two-phase flow regimes in horizontal tubes utilizing an automated image recognition technique," *Experiments in Fluids,* vol. 42, pp. 563-573, 2007.

[9] H. Ding, *et al.*, "Hilbert-Huang transform based signal analysis for the characterization of gas-liquid two-phase flow," *Flow Measurement and Instrumentation,* vol. 18, pp. 37-46, Mar 2007.

[10] A. Mahvash and A. Ross, "Two-phase flow pattern identification using continuous hidden Markov model," *International Journal of Multiphase Flow,* vol. 34, pp. 303-311, Mar 2008.

[11] E. S. Rosa, *et al.*, "Performance comparison of artificial neural networks and expert systems applied to flow pattern identification in vertical ascendant gas-liquid flows," *International Journal of Multiphase Flow,* vol. 36, pp. 738-754, Sep 2010.

[12] M. Refaat, "Principal Component Analysis," in *Data Preparation for Data Mining Using SAS*, ed Burlington: Morgan Kaufmann, 2007, pp. 247-256.

[13] M. T. Brown and L. R. Wicker, "Discriminant Analysis," in *Handbook of Applied Multivariate Statistics and Mathematical Modeling*, E. A. T. Howard and D. B. Steven, Eds., ed San Diego: Academic Press, 2000, pp. 209-235.

[14] L. F. Chen, *et al.*, "A new LDA-based face recognition system which can solve the small sample size problem," *Pattern Recognition,* vol. 33, pp. 1713-1726, Oct 2000.

[15] J. P. Ye and Q. Li, "LDA/QR: an efficient and effective dimension reduction algorithm and its theoretical foundation," *Pattern Recognition,* vol. 37, pp. 851-854, Apr 2004.

[16] H. Yang and J. Y. Yang, "Why can LDA be performed in PCA transformed space?," *Pattern Recognition,* vol. 36, pp. 563-566, Feb 2003.

[17] A. P. Dempster, *et al.*, "Maximum Likelihood from Incomplete Data Via Em Algorithm," *Journal of the Royal Statistical Society Series B-Methodological,* vol. 39, pp. 1-38, 1977.