GMM-based Automatic Defect Recognition Algorithm for Pressure Vessels Defect Detection through ECPT

Xiao Yang∗ Xuegang Huang** Chun Yin∗ Yu-hua Cheng∗ Sara Dadras***

* School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 611731, P. R. China (e-mail: yinchun.86416@163.com, chanyin@uestc.edu.cn).

** Hypervelocity Aerodynamics Institute, China Aerodynamics Research & Development Center, Mianyang 621000, P. R. China.

*** Electrical and Computer Engineering Department, Utah State University, Logan, UT 84321, USA.

Abstract: In order to realize the automatic identification of pressure vessel defects, an improved adaptive defect recognition feature extraction algorithm through ECPT (Eddy current pulsed thermography) is proposed. The proposed feature extraction algorithm consists of five elements: thermal image data segmentation, variable interval search, probability density function modeling, data classification, and reconstructed image acquisition. The combination of data block selection and variable interval search can reduce the double counting. And the KG-EM (Kmeans-GMM-EM) algorithm is proposed to obtain the Gaussian mixture model corresponding to the classification, and thus the corresponding probability is obtained to classify the TTRs (Transient Thermal Response). The reconstructed thermal image is obtained by the classified TTRs. This method can extract the main information of the image accurately and efficiently. Experimental results are provided to demonstrate their effectiveness.

Keywords: Defect recognition, Eddy current pulsed thermography, Transient thermal response, GMM clustering, Nondestructive testing, Defect detection, Pressure Vessels

1. INTRODUCTION

For the space shuttle’s fuel storage, life support, and wind tunnel equipment, the power supply to the pressure vessel plays an important role. However, under high load and complicated use environment, pressure vessels are prone to a series of damages such as micro-cracks, fatigue cracks, and tank corrosion. At the same time, the flexibility of pressure vessel specifications increases the complexity of detecting this important device and identifying defects. In recent years, researchers have studied methods for characterizing pressure vessel characteristics. Lee et al. (2017) presented a method for characterizing the design pressure of a liquefied natural gas cylindrical tank. Al-Gahtani et al. (2014) studied the partial pressure test of a spherical container with a nozzle. Proczka et al. (2013) proposed the effective design and classification criteria for Small Scale Compressed Air Energy Storage (SS-CAES) pressure vessels. Blanc-Vanget et al. (2017) studied the residual detonation pressure of composite cylinders after mechanical shock. Many researchers have studied the effects of defects caused by the continuous use of pressure vessels made of different materials in a wide range of operating environments. van der Burg et al. (1998) refers to the continuous damage analysis of hydrogen attack in pressure vessels. Kim et al. (2011) proposed the effect of composite damage on the fatigue life of natural gas vehicle high pressure vessels. Therefore, detecting potential defects in pressure equipment is very important. Non-destructive testing and evaluation (NDT&E) is an important means of ensuring the safe operation of the production environment and equipment (Maldague (2001), Marinetti et al. (2004), Cheng et al. (2016a)). Huang et al. (2016) proposed some non-destructive testing methods for concrete and fiber cement boards.

Eddy current pulsed thermography (ECPT) has become a non-destructive test with great development potential alternative method (Huang et al. (2016), Schabowicz and Gorzelaczyk (2010)). Pulsed eddy current thermal imaging can quickly and efficiently measure the conductivity, permeability and thermal conductivity distribution of metal components online, which combines the advantages of pulsed eddy current and thermal imaging (Tantichattanon et al. (2007), Francesco et al. (2018)). At present, it has been successfully applied in the nondestructive testing of composite materials, multilayer plate debonding, coagulation, steel, gear and other materials (Yin et al. (2019)). Therefore, ECPT imaging technology is applied to the defect detection of pneumatic equipment. Researchers such as Lim et al. (2016), Liu et al. (2015) have also

* This work was supported by National Basic Research Program of China (Grant No. 61873305) and Sichuan Science and Technology Plan Project (Grant 2019YJ0199, 2018JY0410). (Corresponding author: Chun Yin)
Fig. 1. The experimental schematic diagram made valuable contributions to the processing of infrared camera data, which can be applied to NDT&E. used the normalized fuzzy weighting function to process the records of the experiment. Zhu et al. (2018) proposed an improved defect recognition feature extraction algorithm based on ECPT imaging. The algorithm in this paper uses ECPT to perform non-destructive testing on pressure vessels. The experimental schematic diagram is displayed in Fig. 1.

The algorithm firstly divides the TTRs contained in the thermal image into several parts by thermal image segmentation technology, and finds the low correlation TTRs through the variable interval search method. Specific criteria for variable spacing are given to reduce double counting while retaining the typical TTRs. Secondly the KG-EM algorithm is used to classify the acquired TTRs, and these TTRs are placed into categories that satisfy the highest probability of corresponding GMM models of each type. Thirdly, each type of model is applied to satisfy the maximum value of the probability to find a typical TTR. Finally, a typical TTR can form a matrix to linearly transform the initial image sequence, and then the discriminant features of the infrared image sequence can be extracted by a typical TTR.

2. BACKGROUND: PRESSURE VESSELS

The detection of pressure vessels which show in Fig. 2 is complex. The special fuel pressure vessel used in this experiment is a special equipment for storing and transporting rocket fuel in space launching field. Its operating conditions are harsh, and it has the dangers of leakage, explosion and other accidents. It mainly has the following characteristics.

Fig. 2. Various types of pressure vessels

(1) Big effect: The transfer pipeline is mainly responsible for the mission of transporting fuel from tank trucks to special fuel storage tanks, and the refill pipeline is the life channel for transporting energy to rockets and other spacecraft. Special fuel pressure vessels are one of the most important ground equipment for storing fuel.

(2) Variety specifications: Material types are mainly divided into carbon steel, low alloy steel and a small amount of stainless steel.

(3) Special working conditions: medium such as N2O4 and dimethyl hydrazine is transported, which has stress corrosion and fatigue corrosion conditions, corrosion perforation, fluid scour, bearing corrugated perforations and corrosion cracks are likely to occur due to structural vibration damage.

(4) Destructive consequences of defects: Accidents not only endanger the safety of personnel and equipment, but also lead to delays and failures in test tasks.

In this paper, aiming at solving the above problems and exploring the application of infrared detection in scientific research test pressure vessel defects inspection.

3. METHOD STATEMENT

Due to the depth of the defects and the different shapes, different thermal excitation regions have different temperature change rates, and the spatial temperature response of all these samples is recorded by the infrared camera as an infrared image sequence. These thermal responses cannot be directly identified by the infrared sensor and can be viewed as several different feature regions with different thermal response characteristics, which will help us extract the corresponding infrared reconstructed image. The general ICA method has a relatively high calculation cost. To improve the processing speed, Huang et al. (2018) proposed a K-means-based defect recognition and segmentation algorithm, and defined TTRs (transient thermal responses).

Fig. 3. Transformation selection of data processing methods

In the above algorithm, the Kmeans algorithm is used to classify the obtained TTRs. In addition, the fuzzy c-means (FCM) algorithm is also a widely used clustering algorithm. The above algorithm is used to divide the sample into a certain group according to the similarity of distance metrics. The algorithm regards each category as a separate data set for clustering operation. This clustering method will make the samples in different classes. Fig. 3 clears that when the data overlaps, the cluster clusters within the class will also overlap, and the points at the class boundaries are easily misclassified. In this case, a
valid clustering algorithm is needed to distinguish the data of the overlapping regions, and the probability is more abundant than the distance and easier to compare. Therefore, if a data point is located in the middle of two overlapping clusters, we can simply define its category, the probability is more abundant than the distance and easier to compare.

3.1 KG-EM Algorithm

In this paper, the KG-EM algorithm is first used to initially cluster the sampled sample data $O(\cdot, a_i), i = 1, 2, ..., G$ using K-Means. In this algorithm, $G$ is the number of TTRs contained in $O$. The original sample set $O(\cdot, a_1)$ is represented as $O(\cdot, a_1) = \{a_1, a_2, \ldots, a_i, \ldots, a_G\}$, $i = 1, 2, ..., G$, and $a_k, k \in \{1, 2, \ldots, K\}$ represents the Gaussian mixture component coefficient of the sample, where $i, k(1 \leq i \leq G, 1 \leq k \leq K)$. For each class of the initial classification results obtained by the K-means method, a Gaussian distribution is assumed, and the probability density function of the following formula is satisfied:

$$f_k(a_i | \phi_k, \Upsilon_k) = \frac{1}{(2\pi)^{\frac{N}{2}}}e^{-\frac{1}{2}(a_i - \phi_k)^T \Psi_k^{-1}(a_i - \phi_k)}.$$ 

Specific steps as follows:

1. $K$ vectors $O(\cdot, a_1'), ..., O(\cdot, a_K')$ denoted as $Cen_1', ..., Cen_K'$ and as the initial mean vector as the cluster center, which selected from $O(\cdot, a_1)$ randomly. Update the cluster center iteratively in order to satisfy the stop condition. Then, class $K$ sample data is obtained. The cluster center $Cen_1'^{new}, ..., Cen_K'^{new}$ finally obtained as the initial value $\phi_k = Cen_k^{new}, k = 1, 2, ..., K$ of the GMM-enhanced clustering mean vector.

2. Install all $p(x_i | \Lambda'^{old}) = \sum_{k=1}^{K} x_{ik} f_k(a_i | \phi_k, \Upsilon_k), k = 1, 2, ..., G$ which is a function of Gaussian mixture probability density. The initial value of $\phi_k$ is $\phi_k = Cen_k'^{new}, k = 1, 2, ..., K$.

3. EM-step

Find the $i$th term $O(\cdot, a_i)$ is the posterior probability $v_{ik} = p(k | a_i, \Lambda'^{old}) = \frac{p(x_i, f_k(a_i | \phi_k, \Upsilon_k))}{\sum_{j=1}^{K} x_{ij} f_j(a_i | \phi_j, \Upsilon_j)}, i = 1, 2, ..., z, k = 1, 2, ..., K$ from the $k$th Gaussian distribution. Get $\Lambda'^{new} = \arg \max S(\Lambda, \Lambda'^{old}) \log S(\Lambda, \Lambda'^{old}) = E \left[ \log p(O, X | \Lambda) | O, \Lambda'^{old} \right], S(\Lambda, \Lambda'^{old}) = \sum_{k=1}^{K} \sum_{i=1}^{G} p(k | a_i, \Lambda'^{old}), x_{ik}'^{new}$

that is, use the following formula to update the parameters to get $\Lambda'^{new}$.

Figure up the new mixing factor: $x_{ik}'^{new} = \frac{\gamma}{\sum_{i=1}^{G} v_{ik}}$

Figure up the new mean vector: $\phi_{ik}'^{new} = \frac{\sum_{i=1}^{G} v_{ik} x_{ik}}{\sum_{i=1}^{G} v_{ik}}$

Figure up the new covariance matrix: $\Upsilon_{ik}'^{new} = \frac{\sum_{i=1}^{G} v_{ik} (a_i - \phi_k)(a_i - \phi_k)^T}{\sum_{i=1}^{G} v_{ik}}$

(4) Among them, the gradient method is used to solve the extreme value $S(\Lambda, \Lambda'^{old})$, the condition, and the equation $\begin{cases} \frac{\partial S}{\partial \pi_i} = \lambda \frac{\partial S}{\partial x_i} \\ x_1 + x_2 + \ldots + x_K = 1 \end{cases}$ can be listed as $\begin{cases} \sum_{i=1}^{N} p(\pi_i | a_i', \Lambda'^{old}) \sum_{i=1}^{K} p(K | a_i', \Lambda'^{old}) \\ x_1 + x_2 + \ldots + x_K = 1 \end{cases}$. Be-

cause $\sum_{k=1}^{K} \sum_{i=1}^{N} p(k | a_i, \Lambda'^{old}) = G$ can obtain $\lambda = G$, the revaluation formula of the mixture coefficient is $x_k = \frac{1}{G} \sum_{k=1}^{G} v_{ik}$, where $v_{ik} = p(k | a_i, \Lambda'^{old})$, for $S(\Lambda, \Lambda'^{old})$.

(5) When $S'^{new} - S'^{old} \leq \eta$ is satisfied, outputting the parameters of GMM , else, take $x_{ik} = x_{ik}'^{new}, \phi_{ik} = \phi_{ik}'^{new}, \Upsilon_{ik} = \Upsilon_{ik}'^{new}$ and carry on with steps (5).

(6) The GMM is determined by the model parameters. Then the cluster marks of each sample $a_i$ are determined by the following formula: $\lambda_i = \arg \max v_{ik}, \lambda = k, i \in \{1, 2, ..., K\}$. Then $a_i$ is drawn enter the corresponding cluster $C_{\lambda_i} = C_{\lambda_i}, \cup \{a_i\}$ to get the cluster partition $C = \{C_1, C_2, ..., C_K, ..., C_K\}$. GMM (Reynolds et al. (2000)) can effectively capture differences in temperature-based features. Thereby obtaining an effective classification of TTRs. Since the GMM needs to give the initial mean vector and the initial covariance and needs to solve the iterative parameters repeatedly, the algorithm complexity is high. Steps (1) use the distance metric to make the faster K-means method to find the cluster center. Step (2) constructs the GMM and uses the cluster center vector as the mean value of the mean vector. This increases the speed of the algorithm without losing accuracy. (3) and (4) use the EM algorithm to determine the parameters of the model. (5) is used to judge whether the iteration stops. (6) uses the data corresponding model probability to classify accordingly. Through the KG-EM algorithm, the TTRs data sets of all the temperature variation features sampled are constructed into each corresponding GMM, and the probability that each data satisfies each feature is obtained, and the data set has a probability feature.

4. ALGORITHM DESCRIPTIONS

The classification result of the TTR data set can be obtained by section 3. In this part we will develop a feature extraction algorithm method with KG-EM algorithm for defect recognition to automatically obtain the defect part temperature variation curve and the defect reconstruction image. The algorithm is arranged as follows: Step 1 gives the basic notation of the algorithm. Steps 2-5 to obtain the classification data set. Step 6 uses the KG-EM algorithm to classify the TTRs. Step 7 select the typical TTR for the probability. Step 8 obtain infrared reconstructed image.
Step 1 Notation: $S_{M \times N \times R}$ is the infrared video stream. $S(i, j, k)$ contains the thermal transient response for each pixel in the sequence of thermal images. $K$ is the number of classes. PCC is calculated as $PCC_{A, B} = \frac{D(A, B)}{\sqrt{var(A) \times var(B)}}$.

Step 2 $MP = \max_{m=1, \ldots, M; n=1, \ldots, N; r=1, \ldots, R} [S(m, n, r)]$ the ordinate, abscissa and coordinate of $MP$ are $I_{MP}$, $J_{MP}$ and $T_{MP}$. The PCCs of $S(I_{MP}, L_{MP}, :)$ and $S(I_{MP}, J_{MP}, :)$, $j = 1, \ldots, J_{MP} - 1, J_{MP} + 1, \ldots, N$) are calculated until the $PCC < R_{CL}$. Then, the PCC and $S(I_{MP}, L_{MP}, :)$ of the vector $S(MP, J_{MP}, :)$ are larger than $R_{CL}$ and are recorded as the column interval value $CL$.

Step 3 Design time series $\sum(k), k = 1, 2, 3, \ldots, K, K = 1, 2, 3, \ldots, M * N$ is the time of $ith$ TTR which has the peak value. TTR is divided into $K + 1$ data block by comparing $\sum(k), k = 1, 2, 3, \ldots, M * N$ and $\Gamma_{pk}, i = 1, 2, \ldots, M * N$ . The TTR of the data block of the $kth$ row, the $mth$ row, and then the column is recorded as $S^k(m, n, i)$.

Step 4 Calculate $P_n = \max_{m=1, \ldots, M; n=1, \ldots, N; r=1, \ldots, R} [S^k(m, n, r)]$, get $(I_{nk}, J_{nk}, T_{nk})$ as coordinates of $P_n$.
Calculate the PCC of $S^k(i_{nk}, J_{nk}, :)$, $i = 1, 2, \ldots, M$ constantly until $PCC < R_{FR}^k$, $k = 1, 2, \ldots, K$. $RL^k_n$ is the row interval value of the $nth$ column of the $kth$ data block. Index that the PCC of vector $S^k(i_{nk}, :)$, $i = 1, 2, \ldots, M$ is greater than the amount of $R_{FR}^k$ by use $RL^k_n$.

Step 5 $ee$ is a threshold. Choose $O(\cdot, o_l)$ as the starting point of the loop calculation function. In Fig. 4 shows the detailed calculation process:

1. $S^k(i, j, :)$ and $O(\cdot, o_l)$ of PCC, where $O(\cdot; o_l)$ indicates that the PCC of TTR and $O(\cdot; o_{l-1})$ are smaller than the threshold $EE$.
2. If $PCC < ee$, then, let $l = l + 1$ , $O(\cdot, o_l) = S^k(i, j, :) \ (\text{save new features})$. Otherwise ( i.e., $PCC \geq EE$), let $i = i + RL^k_n$, if $k$ or horizontal coordinate $n$ in $S^k(i, j, :)$ changes, $RL^k_n$ should change. Further, the PCC of the next TTR is calculated using $O(\cdot, o_l)$.
3. If $i > M$, set $i = i - M$. If the line number exceeds the total line number, change to the $j = j + CL$ column.
4. If $j > N$, complete the specific calculation process.

Step 6 We know $O(\cdot, o_l), l = 1, 2, \ldots, G$ contains representative temperature profiles for all representative features in the image sequence through the previous steps. This step uses the KG-EM algorithm (see Section 3 for specific steps) to get the probability that each TTR meets each type of GMM to get the corresponding cluster division $C = \{C_1, C_2, \ldots, C_K\}$.

Step 7 Choose the $Ze^k = \max w_{lk}, k = 1, 2, \ldots, K$ final representative TTRs which as the representation TTR of $Kth$ classification in the cluster $C = \{C_1, C_2, \ldots, C_K\}$. The TTR with $Ze^k$ is contained into $A_l(\cdot, k), k = 1, 2, \ldots, K$.

Step 8 A new matrix $P(x, y)_{v;b} = a \times b = N \times M$ is constructed from a frame image vector and a row vector as a matrix. From $U = A_l * P$ get access to a 2D image matrix $U$, where $A_l$ is a $K \times d$-dimensional pseudo-inverse matrix of the matrix $A_l$. The $U$ is further divided into rows to obtained infrared reconstructed images of size $N \times M$ with a amount of $K$.

5. EXPERIMENTAL RESULTS
In this experiment, the pressure vessel samples provided by the China Aerodynamics Research Institute were selected. Sample A is a circular device for a pressure vessel, as shown in Fig. 5. The width of the defect is 2mm, and the defect of sample A is on the subsurface and cannot be directly observed from the surface. And the sample is very small on a huge pressure vessel and difficult to observe.

In the experiment, the ECPT was used to detect defects by eddy currents in the material, that is, the sample was subjected to eddy current induction while the surface temperature was recorded. The experimental schematic diagram is shown in Fig. 6.

Comparison of classification results In this comparative experiment, the \( O_{l} \), \( l = 1, 2, \ldots, 213 \) obtained by the same sampling uses different clustering methods for TTRs classification. And set \( K \) to 3. From Table 1(1) and (2), it can be seen that the KMeans and FCM algorithms corresponding to the same data set using distance as the metric have repeated divisions. Table 1(1) shows the repeated division due to the equal distance between different cores of some data. It can be seen from the point P1P2 in the Table 1(2) that the corresponding three types of membership degrees in the overlapping classification data portion are \( \phi_1 = \phi_2 = \phi_3 = 0.333 \) to cause repeated division. At the same time, due to the inaccurate division of overlapping regions, the correct classification will be affected. For example, in the middle point of the circle in Table 1(2), the corresponding three types of membership degrees of P3 and P4 are very close. For the circled part of Table 1(2), the corresponding clear division is obtained by KG-EM as shown in Table 1(3). The two TTRs curves of the midpoint of the circle are extracted separately. The temperature change rates of the two TTRs curves shown in Table 1(3) are different, which proves that the circled area can be divided into two categories. This shows the correctness of the KG-EM division.

### Table 2. Experimental parameter setting

<table>
<thead>
<tr>
<th>Sample time: ( \text{REC}_{A} \cdot 0.92 )</th>
<th>( K=3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{REFR}, \text{REFR}<em>{A}^{1} = 0.91, \text{REFR}</em>{A}^{2} = 0.95, \text{REFR}_{A}^{3} = 0.95 )</td>
<td></td>
</tr>
</tbody>
</table>

**Sample A** In this experiment, the heating time is set to 1s, and the infrared camera is used for 12s. Taking sample A as an example, the obtained infrared video stream \( S_{512 \times 640 \times 273} \) has a total of 273 images, and there are 327680 pixel points per image. According to Step 2 - Step 5, data block selection is first performed, and pixel points are divided into three data blocks. Secondly, for these three parts, we set the parameters shown in Table 2 to perform variable step search, and then sample 213 pixel points containing typical TTR features are obtained which show in table 4.

### Table 4. Video stream data classification result

<table>
<thead>
<tr>
<th>Data type</th>
<th>number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original TTRs</td>
<td>327680</td>
</tr>
<tr>
<td>Sampling TTRs</td>
<td>213</td>
</tr>
<tr>
<td>Classification TTRs</td>
<td>( C_{A}^{1} = 26, C_{A}^{2} = 175, C_{A}^{3} = 12 )</td>
</tr>
</tbody>
</table>

According to Step 6, the TTRs are classified by the KG-EM algorithm, the specific parameters are shown in table 3. Use Step 7 to select three TTRs that satisfy the highest probability of GMM to form a linear matrix \( A_{1}(; k), k = 1, 2, 3 \) with dimensions \( 273 \times 3 \). Obtain \( U^{A} = A_{i}' \cdot P_{A} \) and recalculate \( U^{A} \) by row values to form a two-dimensional image with the original image size of \( 512 \times 640 \). That is, 3 reconstructed infrared images are obtained. The image and the corresponding TTRs curves are shown in the Fig. 7.

![Fig. 7. Algorithm result for Sample A](image)

(a) Subsurface defect  
(b) Background  
(c) Excitation coil  
(d) TTRs comparison

The TTR curves classified as shown in Fig. 7 can be observed that the TTR of different classifications has different differences in temperature rise rate and temperature decrease rate, and the type of expression region in the reconstructed image can be judged according to the difference. The type of region is (a) Subsurface defect (b) Background (c) Excitation coil as shown in Fig. 7.

### 6. CONCLUSION

In this study, a GMM-based automatic defect recognition algorithm is proposed for defect recognition in eddy current pulse thermal imaging. This method can extract the main features of the thermal image sequence. At the same time, the main features are used to reconstruct the infrared image. Defect detection can be performed on both surface defects and subsurface defects. The experimental results demonstrate the effectiveness of the method.

### REFERENCES


Table 1. Comparison of classification results

<table>
<thead>
<tr>
<th>Parameter type</th>
<th>Parameter value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMM model</td>
<td>$p_A(x</td>
</tr>
<tr>
<td>Mean vector</td>
<td>$\mu_1^1 = [23.3386, 23.3523, ..., 1 \times 273]$, $\mu_2^2 = [22.6149, 22.6168, ..., 1 \times 273]$, $\mu_3^3 = [22.1764, 22.1688, ..., 1 \times 273]$</td>
</tr>
<tr>
<td>Covariance</td>
<td>$\Sigma_1^1 = 0.0202 0.0180 ...$, $\Sigma_2^2 = 0.0101 0.0094 ...$, $\Sigma_3^3 = 0.0289 0.0289 ...$</td>
</tr>
</tbody>
</table>

Table 3. GMM parameter results for sample A

<table>
<thead>
<tr>
<th>Parameter type</th>
<th>Parameter value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter type</td>
<td>Parameter value</td>
</tr>
<tr>
<td>Parameter type</td>
<td>Parameter value</td>
</tr>
</tbody>
</table>


