附件 5C：译文

指导教师评定成绩
(五级制)：

指导教师签字：

基于高斯混合模型的控制图模式识别

School of Mechatronics Engineering and Automation, Shanghai University, Shanghai 200072, P.R. China; bState Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710054, P.R. China

摘要： 控制图中显示的异常模式可以与制造过程的变动明确地联系起来。快速、准确的控制图模式识别（CCPR）能够显著地缩小查找异常的原因范围，更快地解决异常问题，因此研究 CCPR 是完全必要的。本文利用多个高斯混合模型（GMM）模型构建了一个 CCPR 模型，提出了基于 GMM 的 CCPR 模型。将生产的统计特征和小波能量特征作为输入变量输入模型，使得改善后的多个 CCPR 模型比单一的 CCPR 模型训练更加简单和高效，同时模型效果也更加好，这使得质量程师和操作人员能够在生产车间更加方便地使用改进之后的模型。并且该模型采用动态计划，当条件变化时，可以很好地应用到新的控制图。实验结果表明模型在当前 CCPs 的检测和识别方面效果良好，并有效地适应新的 CCPs。更重要的是，模型因为其计算的高效性和良好的识别效果为在线检测提供一种可能。最后本文提出了一个基于 GMM 统计过程控制（SPC）识别系统的开发方案。

关键词：质量控制；模式识别；控制图；统计模型；统计过程控制；控制图模式识别；高斯混合模型；特征抽取；自适应识别

1. 简介

统计过程控制（SPC）已经被广泛地应用在监控和控制工业部件生产过程，用来监控部件的生产率和质量情况。在实施 SPC 过程中最核心的部分是控制图的异常变量展示的测量。通常在 SPC 中，一个过程在只有随机因素影响时会假设其是自然或者正常的。常见考虑的原因是由于其内在的自然或者正常过程。特殊原因被定义为对正常过程的异常冲击，这种冲击需要尽可能快地被标识和清除。当控制图中显示异常就需要确定并清除异常原因。进一步讲，对观察结果进行分析，例如，控制图显示的正常和不正常模式，提供了很多过程信息，该信息可以通过合适的分析来推断出造成这种现象背后的原因。因此，控制图模式（CCPR）识别成了 SPC 的关键任务。CCPR 可以结合工程知识极大地缩小异常原因的范围。这种极大地缩小异常诊断范围的方式给予处于失控状态的过程以必要的修正和调整使之

回归受控。然而，控制图再出现异常时不提供任何和原因相关的信息，只根据当前的观测结果监控制造过程。

多年来，研究人员提出像 zone tests 、run rules 等很多规则来补充对控制图的分析，以提高效果。这些规则来协助质量人员检测控制图上异常模式（CCP）。使用这些规则存在的主要问题有，应用所有规则会造成的过量的假警报（Guh2005）；同时，它往往依赖于技巧和分析的经验来确定是否存在异常的 CCP。此外，为了构建新的 CCPs 需要添加新的规则，这样会使规则集变的很大，从而导致难以维护和管理。因此，CCP 在 zone tests or run rules 方面有很多困难，因为这涉及到模式识别、自适应学习方面。

由于过程监控中数据量大，手工检查效率低，不一致，现有的基于人工智能的方法为实现过程监控和故障诊断自动化提供了一种有效途径。人工神经网络（ANNs）具有良好的实时噪声容忍度，不需要对监测数据的统计分布进行假设。这一重要特性使人工神经网络成为有前途和有效的工具，可以用来改善制造质量控制应用中的数据分析。在过去的十年中，不同的结构人工神经网络和学习算法在质量控制中得到了广泛的应用。概率神经网络（Plummer 1993），学习向量化神经网络 (LVQ)，自适应谐振理论（ART）(Hwarng and Chong 1995)，径向基础函数（RBF）（Cook and Chiu 1998），多层感知器（MLP）神经网络（(Chang and Ho 1999 ）等。已经用来识别基础异常的模式如：移位，趋势，周期，混合。Guh（2004）使用遗传算法来优化 CCPR 中的 ANN 误差前馈。Guh（2005）更进一步提出了混合学习模型，整合了 BPN 和决策树，为 CCP 时设计了一个高效的识别系统。wang（2007）针对离散过程提出一个包括 ANN 新的框架能够识别 6 种过程信号。他们的结果表明，通过小波滤波去噪可以有效地提高用于识别过程模式和 ART 神经网络识别性能。Jiang（2009）提出了一种神经网络数值拟合模型来识别 CCPs。在这个模型中，使用数值拟合方法来估计参数和模式的特定类型，这是不同于常规的人工神经网络的分类方法。近年来，集成人工神经网络方法已被用于诊断监测制造过程(Yu and Xi 2009, Yu et al. 2009).

使用 ANNs 存在一个缺点，即人工神经网络的过程信息是不可见的，使用者无法理解内部机理，不了解产生这样结果的原因。另外，ANN 网络结构一般比较复杂，训练过程耗时长。为了解决这些问题，有人提出了基于特征的分类方法(Hassan et al. 2003, Assaleh and Al-Assaf 2005, Gauri and Chakraborty 2007, Yu et al. 2007)，实验结果表明，识别器的结构可以可见，并且显著减少了训练时间。

近年来，其他的识别器（如，决策树，(Wang et al. 2008)；c-means， (Wang and Kuo 2007)；支持向量机(SVM) (Ranaee et al. 2010)；统计相关系数方法，(Yang and Yang 2005)等）已经被应用到 CCPR。虽然基于这些分类模型的识别器表现出良好的性能，但其结构一般都是固定的，一旦建成使用，由于复杂的动态工作条件，时常产生新的异常模式，可能会限制他们在现实生产中的使用。

根据以上的分析可以看出，快速准确地识别关键控制点是非常必要的，尤其是在实际的制造过程。此外，新的 CCPS 在线识别仍然是需要解决的一个重要问题。类比其他相近的研

究(Hassan et al. 2003, Guh 2005, Gauri and Chakraborty 2007, etc.), 本研究还考虑了以下八种有意义的模式：正常状态、上下转换、上下趋势、周期、混合和系统模式。

本研究中，在独立离散过程中使用了基于高斯混合分布（GMM）的模型。GMM 一个是对估计过程观测很有效果的概率密度函数（PDF）。GMM 是基于无监督学习的模型，模型会根据输入数据的性质进行拟合，学习复杂的分布。模型的训练完全是数据驱动的。与之前使用单一的识别器识别所有的 CCP 不同，针对在制造过程中发生异常，我们构建了 6 个 GMM 模型。这样的模型构造的方案在应用到实际制造过程中很有吸引力。事实上，只使用一个 CCP 的知识来构建一个 GMM，因此比所有六个点的训练数据训练的 GMM 更好的泛化性能。此外，基于 GMMs 的分类模型是不固定的，可以根据实际应用的要求改变。例如，它可以为新的 CCP 添加新的高斯分布，改变 CCP 只需训练一个 GMM 即可。根据我们的发现，很少有论文论述这种建模方案对提高制造过程适用性的重要性。更重要的是，现有的大多数研究往往只考虑在预先定义（已知）的模式下的识别性能，但没有考虑到自适应能力相对于在制造过程产生的新的（未知）的模式情况。因此，本文的研究目标如下：

（1）基于时间和时频域改进 GMM，减少控制图模式识别的建模时间并提高其识别性能。

（2）设计一个基于多个 GMMs 的 CCPR 模型。不仅是结合当前数据的模型，而且要适合可能出现的新模式，具有较强的泛化能力。

（3）解决该模型的在线识别性能，包括过程监视和识别。

论文的其余部分安排如下。第 2 节描述了 GMM 和相关算法。3 部分提出了一种基于 GMM 模型的新分类。模拟过程用于说明第 4 节中所提出方法的性能。第 5 节给出结论并进行延伸讨论。

2．高斯混合模型

本文，提出了八个关键的控制点（即正常，向上/向下移动，向上/向下的趋势、周期、系统和混合是在 3.1 节中讨论），基于这些关键控制点建立 GMM 的识别器。我们先建立 6 个 GMMs 分类器，然后识别前 6 个 CCPs。最后新添加的两个关键控制点作为检测的自适应学习能力，提出基于 GMM 模型的新 CCPs 国际。有关 GMM 的详细信息如下所示。

设 x 是一个 d 维的随机变量， x 代表一个特定的输出。即 x 服从一个有限个分布相加的混合分布，其概率密度函数（pdf）可以写成有限个已知分布的加权的和。比如，每个分布都是高斯分布，x 服从高斯混合分布（GMM）。

$$p(x|\phi) = \sum_{m=1}^{M} \pi_m p(x|\theta_m), \tag{1}$$

该等式中，$\pi_m \in (0,\ 1)(\forall m\ = 1,2,\dots,M)$服从条件$\sum_{m=1}^{M} \pi_m = 1$. 对于高斯混合每个部分（component）的密度函数$p(x|\theta_m)$是正态分布的, 每个部分按照自己的参数$\theta_m(\mu,\ S_m)$，贡献均值μ_m和协方差矩阵S_m。我们可以将这些参数封装到起来，以向量的形式表达：$\emptyset = \left(\pi_1,\dots,\pi_m; S_1\ \dots, S_m\right)$。

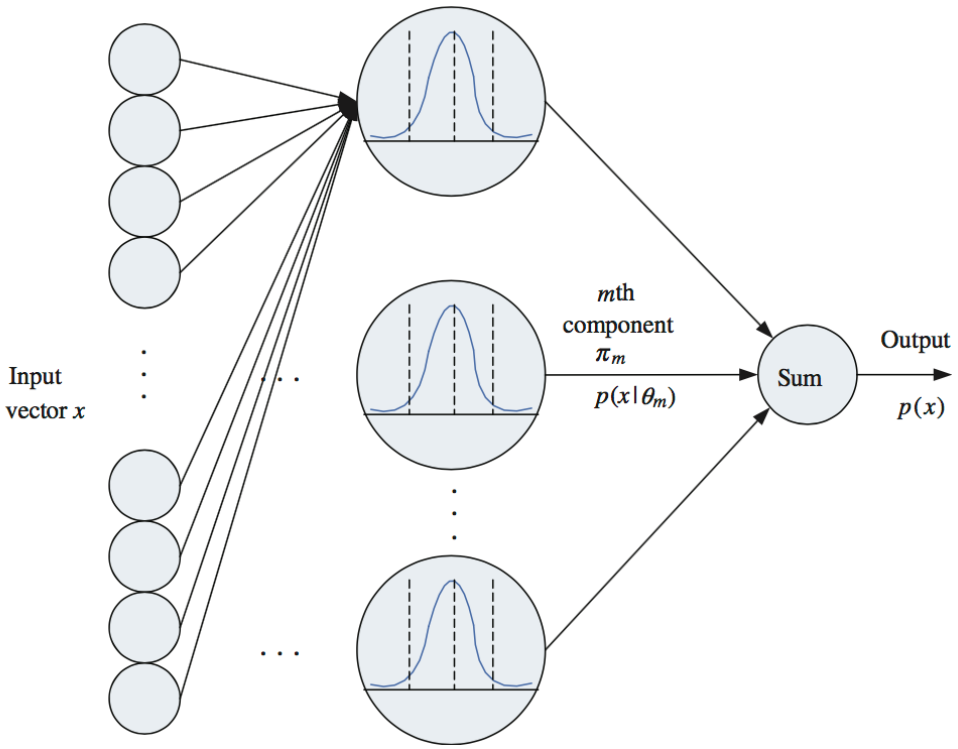

图 1 高斯混合模型

在参数估计时，我们假设训练集 $X = \left(x^{(1)}, \ldots, x^{(n)}\right)$是$n$个来源于独立同分布的随机变量$X$，学习的目的是找到$M$个部分优化其中参数。其极大似然函数为：

$$\log(p(X|\phi)) = \log \prod_{i=1}^{n} p(x^{(i)}|\phi). \tag{2}$$

通常的做法是用期望最大化算法（EM）来优化目标函数，得到混合参数向量。EM 是一种强大的统计工具算法，用于求解涉及可观测变量和隐藏变量问题的最大似然解。它通过迭代，不断的寻找局部最优解。EM 分为 E-step 和 M-step，其中 E-step 和 M-step 被不断的迭代进行。EM 算法在高斯混合模型使用的很广泛，其收敛性已经得到了很好的证明。(Dempster et al. 1977)。

但标准的 EM 算法有一个弱点：它需要预先确定 M（即分布的数量），以便将参数优化到一个较优的局部最优解。为了克服这一困难，目前已经提出了许多评价标准来估计 GMM 中适当的分布个数。大多数实际的模型选择方法都是基于最大化以下类型的准则：

$$J(M, \phi(M)) = \log p(X; \phi(M)) - P(M), \tag{3}$$

其中$p(X:\emptyset(M))$是给定数据集的对数似然，该部分可以使用 EM 算法来使之最大化。但是，添加更多的分布数肯定是可以增加对数似然，为了达到均衡需要引入惩罚项来惩罚分布的复杂度。模型的复杂度该如何选择已经有不少研究了，有很多选择的准则比如：赤池信息准则（AIC）（赤池 1974），最小描述长度（MDL）（rissansen 1987），贝叶斯推理的标准（BIC）（Schwarz 1978），在这项研究中，我们使用了 Figueiredo 和 Jain（2002）提出的 FJ 规范：

$$FJ = L(\theta, X) = \log P(X|\theta) - \frac{N}{2}\sum_{m=1}^{M}\log\left(\frac{n\pi_m}{12}\right) - \frac{m}{2}\log\frac{n}{12} - \frac{m(N+1)}{2}, \tag{4}$$

$$\hat{\theta} = \arg\max_{\theta} L(\theta, X), \tag{5}$$

其中 N 是指定每个分布的数量，n 是训练样本的总数。Figueiredo 和 Jain 的实验结果表明，该标准比其他标准如 MDL，BIC 的效果要更好一些。一般来说，为 GMM 选择一个好的评价指标是很困难。一个有效的方法是根据先前的经验来从 AIC，BIC，FJ 中选择。我们之前的实验结果表明，FJ 标准的效果优于其他标准（如 AIC 和 BIC），因此选择 FJ 标准作为本研究的模型的评价标准。

3. 基于 GMM 的控制图模式识别

本部分主要对离散制造过程和独立同分布的 CCPs 使用 GMM 模型进行了概述。使用的方法包括模拟训练测试 GMMs 模型，设置影响模型性能的关键参数，使用 GMM 模型作为 CCPs 识别器提高识别性能。

3.1 数据集获取

理想情况下，最好可以从现实世界中获取制造过程的多元数据。但是由于需要大量的数据，所以本文借鉴了蒙特卡洛法来生成一些数据。模拟的方式可以使得在独立离散的生产过程中发现未知模式、潜在问题成为可能，为上述需求提供支撑。

在本文中，考虑了八种典型的离散制造过程的 CCPs，包括：即正常模式，向上/向下移动，向上/向下的趋势，周期，混合系统。这些模式用 CCP1，CCP2，…，CCP8 来表示。生成八个 CCPs 的表达式为：

$$x(t) = n(t) \sim N(0, 1), \tag{6}$$

$x(t) = n(t) \sim N(0,1)$是一个模拟第 t 次的样本，服从$N(0，1)$分布的均值为 0，方差为 1.

CCP2/CCP3：

$$x(t) = n(t) \pm \mu(t - t_d)\xi, \tag{7}$$

ε是变化的幅度，在$(1.5\sigma \le \varepsilon \le 2.5\sigma)$区间内，$t_d$是发生在一个时间窗口的时间改变点，$\mu(t - t_d)$是一个阶跃函数。

$$\mu(t - t_d) = \begin{cases} 0, & t < t_d, \\ 1 & t \ge t_d. \end{cases} \tag{8}$$

CCP4/CCP5:

$$x(t) = n(t) \pm \mu(t - t_d) \times s \times t, \tag{9}$$

s是变化的趋势的区间在$(0.05\sigma \le s \le 0.15\sigma)$区间内

CCP6：

$$x(t) = n(t) + \mu(t - t_d) \times a \times \sin(2\pi t/\Omega), \tag{10}$$

a 是一个周期的振幅$(1.5\sigma \le a \le 3\sigma)$，Ω是周期$\Omega = 12$

CCP7：

$$x(t) = n(t) + \mu(t - t_d) \times (-1)^t \times \eta, \tag{11}$$

η是振幅$(1.0\sigma \le \eta \le 2\sigma)$

CCP8：

$$x(t) = n(t) + \mu(t - t_d) \times (-1)^{\varepsilon} \times \delta, \tag{12}$$

δ是振幅$(1.5\sigma \le \delta \le 2.5\sigma)$。为了方便使用，设过程中的数据服从$N(0，1)$分布，均值为 0，标准差为 1。

在本研究中，模拟前 6 个模式的数据用于模型的训练和测试。除去正常的模式的数据，有每五个由正常模式转向异常的样本数据，同时包括造成异常的原因。在时间窗口的一半左右随机地选择变更点（参见图 2（b）-（d））。为了检验制造过程中出现的新的异常模式时模型的适应性能。从其他两个模式中选取样本，即 ccp7（系统）和 ccp8（混合物），产生新的关键控制点（见图 2（g）和（h））。

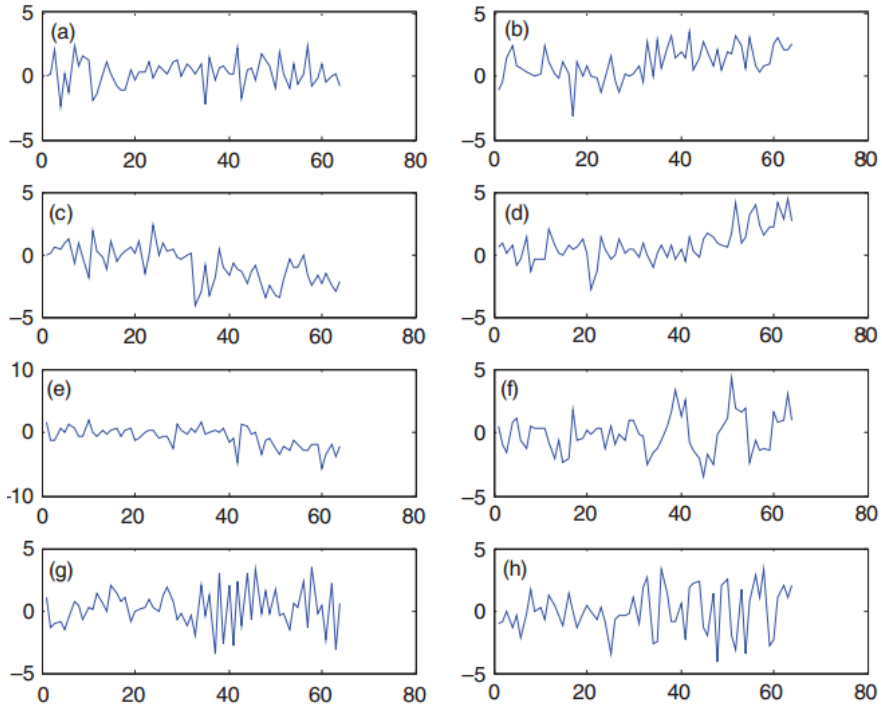

图 2 八种控制图模式

译文原文出处：Yu, J. (2012). Gaussian mixture models-based control chart pattern recognition. International Journal of Production Research, 50(23), 6746 - 6762.

International Journal of Production Research

ISSN: 0020-7543 (Print) 1366-588X (Online) Journal homepage: http://www.tandfonline.com/loi/tprs20

Gaussian mixture models-based control chart pattern recognition

Jianbo Yu

To cite this article: Jianbo Yu (2012) Gaussian mixture models-based control chart pattern recognition, International Journal of Production Research, 50:23, 6746-6762, DOI: 10.1080/00207543.2011.623724

To link to this article: https://doi.org/10.1080/00207543.2011.623724

International Journal of Production Research
Vol. 50, No. 23, 1 December 2012, 6746–6762

Taylor & Francis
Taylor & Francis Group

Gaussian mixture models-based control chart pattern recognition

Jianbo Yu[ab]*

[a]*School of Mechatronics Engineering and Automation, Shanghai University, Shanghai 200072, P.R. China;* [b]*State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710054, P.R. China*

(*Received 11 October 2010; final version received 7 September 2011*)

Abnormal patterns exhibited in control charts can be associated with certain assignable causes for process variation. Hence, accurate and fast control chart pattern recognition (CCPR) is essential for significantly narrowing down the scope of possible causes that must be investigated, and speeds up the troubleshooting process. This study proposes a Gaussian mixture models (GMM)-based CCPR model that employs a collection of several GMMs constructed for CCPR. By using statistical features and wavelet energy features as the input features, the proposed CCPR model provides a more simple and effective training procedure and better generalisation performance than using a single CCPR recogniser, and hence is easier to be used by quality engineers and operators. Furthermore, the proposed model is capable of adapting novel control chart patterns (CCPs) by applying a dynamic modelling scheme. The experimental results indicate that the GMM-based CCPR model shows good detection and recognition performance for current CCPs and adapts further novel CCPs effectively. Moreover, the proposed model provides a promising way for the on-line recognition of CCPs because of its efficient computation and good pattern recognition performance. Analysis from this study provides guidelines for developing GMM-based statistical process control (SPC) recognition systems.

Keywords: quality control; pattern recognition; control charts; statistical methods; statistical process control; control chart pattern recognition; Gaussian mixture models; feature extraction; adaptive recognition

1. Introduction

Statistical process control (SPC) has been widely used to monitor and control manufacturing processes to maintain the quality and productivity of industrial components. The most essential aspect of implementing SPC is control charting (e.g., X-bar, EWMA CUSUM, etc.) to reveal abnormal variations of monitored measurements. Generally, in SPC, a process is assumed natural or normal if only random causes are affecting its operation. Common causes are considered to be due to the inherent nature of the normal process. Assignable causes are defined as abnormal shocks to the processes, which should be identified and eliminated as soon as possible. When an abnormal variation is signaled by the control chart, one has to determine the assignable causes and eliminate them. Further, analysis of the observations, i.e. normal and abnormal patterns presented by the control chart, provides important process information, which can be associated with a particular set of assignable causes if suitable process knowledge is available for analysis. Therefore, control chart pattern recognition (CCPR) is a critical task in SPC for the coupling with engineering knowledge of the process to greatly narrow down the set of possible causes. This significantly minimises diagnostic search efforts and helps one to take the necessary corrections and adjustments to bring an out-of-control process back to the in-control state. However, control charts do not provide any pattern-related information when the process is in the out-of-control state, because they consider only the current observations to monitor manufacturing processes.

Over the years, numerous supplementary rules such as zone tests or run rules (Western Electric 1958, Nelson 1984, Duncan 1986, Grant and Leavenworth 1996) have been developed to analyse control charts. These rules were developed to assist quality practitioners in detecting unnatural control chart patterns (CCPs). The primary problems with applying run rules are that the application of all the available rules can yield an excess of false alarms (Guh 2005); also, it often relies on the skills and experience of the analyst to determine if an unnatural CCP exists. Additionally, new rules need to be constructed for novel CCPs, which often results in a large rule set and thus results

*Email: jianboyu@shu.edu.cn

ISSN 0020–7543 print/ISSN 1366–588X online

http://dx.doi.org/10.1080/00207543.2011.623724
http://www.tandfonline.com

in difficulty in maintaining and managing these rules. Therefore, the interpretation of CCPs still remains difficult for zone tests or run rules because this involves pattern recognition and adaptive leaning aspects.

Because the huge volume of data in process monitoring makes manual checking extremely inefficient and inconsistent, current approaches based on artificial intelligence provide an effective way to automate process monitoring and fault diagnosis. Artificial neural networks (ANNs) have excellent noise tolerance in real time, requiring no hypothesis on the statistical distribution of monitored measurements. This important feature makes ANNs promising and effective tools that can be used to improve data analysis in manufacturing quality control applications. In the last decade, various ANNs with different structures and learning algorithms have widely been adopted in quality control. Probabilistic ANN (Plummer 1993), learning vector quantisation networks (LVQ) (Pham and Oztemel 1994), adaptive resonance theory networks (ART) (Hwarng and Chong 1995), radial basis function (RBF) (Cook and Chiu 1998), multilayer perceptron (MLP) ANN (Chang and Ho 1999), etc. have been used to identify basic abnormal patterns such as shift, trend, cycle, mixture patterns, etc. Guh (2004) used a genetic algorithm to optimise a feedforward ANN for CCPR. Guh (2005) further proposed a hybrid-learning-based model integrating back propagation networks (BPN) and decision trees as an effective identification system of CCPs. Wang *et al*. (2007) proposed a novel framework involving ANN models for the recognition of the six main types of process signals (i.e. normal condition, variance dispersion, positive mean shift, negative mean shift, positive mean shift and variance dispersion, and negative mean shift and variance dispersion) in discrete processes. Their results showed that denoising through a wavelet filter is effective for improving the identification performance of LVQ and ART ANNs used for identifying process patterns. Jiang *et al*. (2009) proposed a BPN-numerical fitting model to recognise CCPs. In this model, the numerical fitting method estimates the parameters and specific types of patterns, which is different from regular ANN-based CCPR methods. Recently, the ANN ensemble approach has been used for monitoring and diagnosis in manufacturing processes (Yu and Xi 2009a,b, Yu *et al*. 2009).

One disadvantage of ANNs is that the process information of ANNs is implicit and virtually inaccessible to users. In addition, the network structure is generally complex and the corresponding training process is time-consuming. Regarding this, feature-based CCPR methods have been proposed (Hassan *et al*. 2003, Assaleh and Al-Assaf 2005, Gauri and Chakraborty 2007, Yu *et al*. 2007) to solve these issues, and the experimental results show that feature extraction makes the structure of recognisers explicit and reduces the training time significantly.

Recently, other recognisers (e.g., decision tree (Wang *et al*. 2008), c-means (Wang and Kuo 2007), support vector machine (SVM) (Ranaee *et al*. 2010), the statistical correlation coefficient method (Yang and Yang 2005), etc.) have been used to implement CCPR. Although these recognisers-based CCPR models show good performance, their structures will generally be fixed once they are constructed and used, which could limit their real-world applications where novel abnormal patterns often occur due to complicated and dynamic working conditions.

According to the above analysis it can be seen that fast and accurately recognising CCPs are required, especially in real-world manufacturing processes. Moreover, the online recognition of novel CCPs is still an important issue that needs to be solved. Similar to other research (Hassan *et al*. 2003, Guh 2005, Gauri and Chakraborty 2007, etc.), this research also considers several meaningful patterns comprising the following eight types of process signals: normal condition, upward and downward shift, upward and downward trend, cycle, mixture and systematic patterns.

In this study, a Gaussian mixture model (GMM)-based model is developed to recognise CCPs in independent and discrete processes. GMM is capable of estimating the probability density function (PDF) of process observations. GMM is based on unsupervised learning to organise itself according to the nature of the input data with a complicated distribution, which means that the training is entirely data-driven. Different to previous modelling approaches using a single recogniser to identify all CCPs, we construct GMM for each CCP occurring in manufacturing processes, and thus six GMMs corresponding to the considered six CCPs (i.e. normal, upward and downward shift, upward and downward trend and cycle) are constructed and then combined to recognise CCPs in manufacturing processes. Such a model construction scheme is appealing in real-world manufacturing processes. Actually, only the knowledge of each CCP is required to construct each GMM, and thus the GMM has better generalisation performance than the GMM trained by the training data with all six CCPs. Moreover, the whole CCPR model with six GMMs is not fixed and can be changed according to the requirements of real-world applications. For example, it can add new GMMs for novel CCPs, and retrain only one GMM for changed CCP. To the best of our knowledge, few papers have addressed the importance of such a modelling scheme to improve applicability in manufacturing processes. More importantly, most existing research tends to consider only the performance of pattern recognition with respect to predefined (known) patterns, but does not take into account the

adaptive capability with respect to novel (unknown) patterns occurring in manufacturing processes. Hence, the three goals of this study are as follows:

(1) address the benefit of feature extraction based on the time and time-frequency domain with decreasing modelling time of GMMs and improving their recognition performance;
(2) design a CCPR model with multiple GMMs that not only incorporates current patterns, but also adapts future novel patterns; and
(3) address the online performance of the proposed model, which includes process monitoring and recognition.

The rest of the paper is organised as follows. Section 2 describes the GMM and related algorithms. Section 3 proposes a novel GMM-based CCPR model. Simulated processes are used to illustrate the performance of the proposed approach in Section 4. Section 5 provides the conclusions and an extensive discussion.

2. Gaussian mixture model

In this study, eight CCPs (i.e. normal, upward/downward shift, upward/downward trend, cycle, systematic and mixture to be discussed in Section 3.1) are generated and then the GMM is used as the recogniser of these CCPs. Six GMMs are constructed and then combined to recognise the first six CCPs. The last two CCPs are used as novel CCPs to test the adaptive learning capability of the proposed GMM-based CCPR model. Detailed information concerning the GMM is presented as follows.

Let $X = [X_1, \ldots, X_d]$ be a d-dimensional random variable, with $x = [x_1, \ldots, x_d]$ representing one particular outcome of X. It is said that X follows a finite mixture distribution when its probability density function $p(x)$ can be written as a finite weighted sum of known densities. In cases where each component is the Gaussian density, X follows a Gaussian mixture ϕ. A GMM $p(x|\phi)$ is the weighted sum of $M > 1$ components $p(x|\theta_{\mathrm{m}})$, as shown in Figure 1:

$$p(x|\phi) = \sum_{m=1}^{M} \pi_m p(x|\theta_m), \tag{1}$$

where $\pi_m \in (0, 1)(\forall m = 1, 2, \ldots, M)$ are the mixing proportions subject to $\sum_{m=1}^{M} \pi_m = 1$. For the Gaussian mixtures, each component density $p(x|\theta_m)$ is a normal probability distribution, where each component is denoted by the

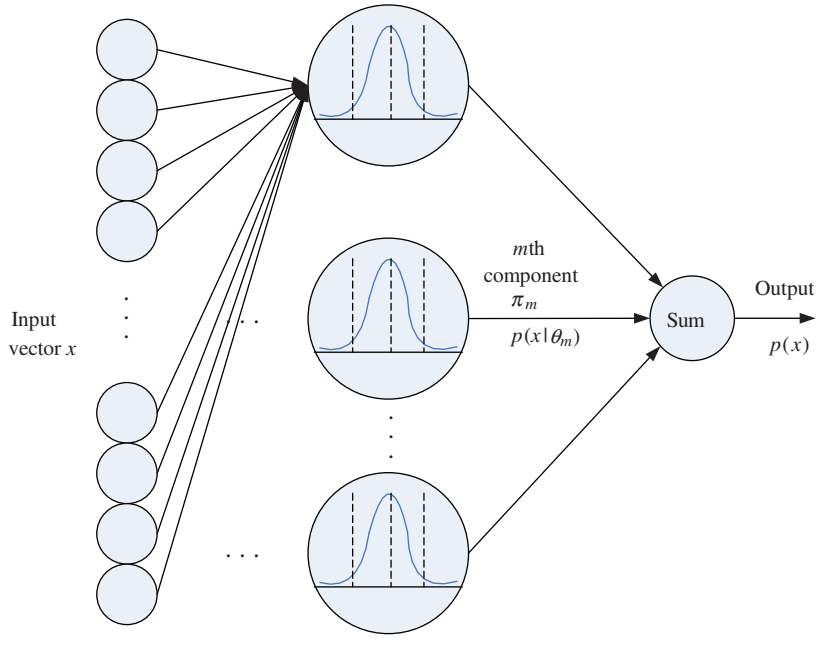

Figure 1. The Gaussian mixture model.

parameters $\theta_m = (\mu_m, S_m)$, the mean vector μ_m and the covariance matrix S_m. We encapsulate these parameters into a parameter vector to obtain $\phi = (\pi_1, \ldots, \pi_M; \theta_1, \ldots, \theta_M)$. For the estimation problem we assume a training set $X = \{x^{(1)}, \ldots, x^{(n)}\}$ with n independent and identically distributed samples of the random variable x. Learning aims at finding the number of components M and the optimum vector $\phi^* = (\pi_1^*, \ldots, \pi_M^*; \theta_1^*, \ldots, \theta_M^*)$ that maximises the likelihood function

$$\log(p(X|\phi)) = \log \prod_{i=1}^{n} p(x^{(i)}|\phi). \tag{2}$$

The usual choice for obtaining the optimum vector ϕ^* of the mixture parameters is the expectation maximisation (EM) algorithm (Dempster *et al*. 1977). EM is a powerful statistical tool for finding maximum likelihood solutions to problems involving observed and hidden variables. EM is an iterative procedure to find local maxima of $\log(p(X|\phi))$, where an E-step and M-step are implemented iteratively. For the case of Gaussian mixtures, the convergence behaviour of EM has been well studied (Dempster *et al*. 1977).

Standard EM for mixture learning has a weakness that also affects the EM algorithm introduced above: it requires knowledge of M (i.e. the number of components) to reach a good local optimum. To overcome this difficulty, many deterministic criteria have been proposed to estimate the appropriate number of components in GMM. Most of the practical model selection techniques are based on maximising the following type of criterion:

$$J(M, \phi(M)) = \log p(X; \phi(M)) - P(M), \tag{3}$$

where $\log\ p(X; \phi(M))$ is the log-likelihood for the given data set X. This part can be maximised using the EM algorithm. However, introducing more components always increases the log-likelihood. The balance is achieved by introducing $P(M)$ that penalises the complex solutions. Some examples of such a model selection criterion are the Akaike information criterion (AIC) (Akaike 1974), the minimum description length (MDL) (Rissansen 1987), the Bayesian inference criterion (BIC) (Schwarz 1978), etc. In this study, we use the FJ criterion proposed by Figueiredo and Jain (2002) for our model:

$$FJ = L(\theta, X) = \log P(X|\theta) - \frac{N}{2} \sum_{m=1}^{M} \log\left(\frac{n\pi_m}{12}\right) - \frac{m}{2} \log \frac{n}{12} - \frac{m(N+1)}{2}, \tag{4}$$

$$\hat{\theta} = \arg\ \max_{\theta} L(\theta, X), \tag{5}$$

where N is the number of parameters specifying each component and n is the total number of training samples. Figueiredo and Jain's experimental results showed that this criterion is better than other criteria such as MDL, BIC, etc. Generally, it is difficult to select one good model selection index for GMM. An effective approach is to implement a prior experiment to select one from AIC, BIC, FJ, etc. Our prior experimental results show that FJ is similar to or better than other criteria (i.e. AIC and BIC) and thus is used as the model selection criterion in this study.

3. GMM-based control chart pattern recognition

This section provides a practical overview for using GMMs with respect to identifying CCPs in independent and discrete manufacturing processes. The methodology used consists of simulation data set development for training and testing GMMs, the setting of key parameters that influence the generalisation performance of the GMM, and GMM is used as an identifier of CCPs to carry out the experiments.

3.1 *Data set development*

Ideally, various CCP data should be collected from real-world manufacturing processes. Due to the need for large numbers of example patterns for the recogniser's training and testing, a Monte-Carlo simulation approach is used to generate the required data sets of CCPs. Simulation provides a platform from which investigations into potential problems associated with unnatural patterns in an independent and discrete process can begin.

In this study, eight typical CCPs presenting in discrete manufacturing processes are considered, namely normal pattern, upward/downward shift, upward/downward trend, cycle, systematic and mixture, which are denoted CCP1, CCP2, ..., CCP8, respectively. The expressions used for the generation of the eight CCPs are as follows:

CCP1: Normal condition,

$$x(t) = n(t) \sim N(0, 1), \tag{6}$$

where $x(t)$ is the sample value at time t, and $N(0, 1)$ is the normal distribution with mean 0 and variance 1.

CCP2/CCP3: Upward/downward shift,

$$x(t) = n(t) \pm \mu(t - t_d)\xi, \tag{7}$$

where ξ is the magnitude of the shifts ($1.5\sigma \leq \xi \leq 2.5\sigma$), t_d is the change point occurring in the time point of the time window, and $\mu(t - t_d)$ is a unit step function:

$$\mu(t - t_d) = \begin{cases} 0, & t < t_d, \\ 1 & t \geq t_d. \end{cases} \tag{8}$$

CCP4/CCP5: Upward/downward trend,

$$x(t) = n(t) \pm \mu(t - t_d) \times s \times t, \tag{9}$$

where s is the trend slope ($0.05\sigma \leq s \leq 0.15\sigma$).

CCP6: Cycle,

$$x(t) = n(t) + \mu(t - t_d) \times a \times \sin(2\pi t/\Omega), \tag{10}$$

where a is the amplitude of the cycle ($1.5\sigma \leq a \leq 3.0\sigma$) and Ω is the period of the cycle ($\Omega = 12$).

CCP7: Systematic,

$$x(t) = n(t) + \mu(t - t_d) \times (-1)^t \times \eta, \tag{11}$$

where η is the magnitude ($1.0\sigma \leq \eta \leq 2\sigma$).

CCP8: Mixture,

$$x(t) = n(t) + \mu(t - t_d) \times (-1)^{\varepsilon} \times \delta, \tag{12}$$

where δ is the magnitude ($1.5\sigma \leq \delta \leq 2.5\sigma$). For ease of application, let the process follow the standard normal distribution with a zero mean and one standard deviation, namely $x(t) \in N(0, 1)$.

In this study, simulation samples of the first six patterns were generated as training and testing data of the GMMs. Except for the normal pattern (i.e. CCP1), each sample in the other five patterns starts from the normal condition, and enters the abnormal condition after the change points resulting from various assign causes. The change points are randomly chosen around half of the time window (see Figures 2(b)–(d)). In order to test the adaptive performance of the proposed model for novel abnormal patterns occurring in manufacturing processes, simulation samples from the other two patterns, i.e. CCP7 (systematic) and CCP8 (mixture), are generated as novel CCPs (see Figures 2(g) and (h)).

3.2 *Input features for GMM*

In general, GMM finds it difficult to describe a probability density distribution of high dimension and sparse data. In addition, the number of parameters in a mixture model with k local Gaussian components increases by $2k(m + 1)$ when a new observation is added to the m observations. Furthermore, the number of parameters increases by $m + m^2$ when an additional Gaussian component is added to a mixture model. Reduction of the data dimension alleviates these difficulties due to the reduction in the number of parameters to be determined in GMM. Thus, raw data with 64 sample points cannot be used directly as inputs of GMM.

The selection of the input features in the training set significantly affects the performance of identifiers. The input feature vector must be able to strengthen the pattern feature of the data set. In this study, instead of inputting

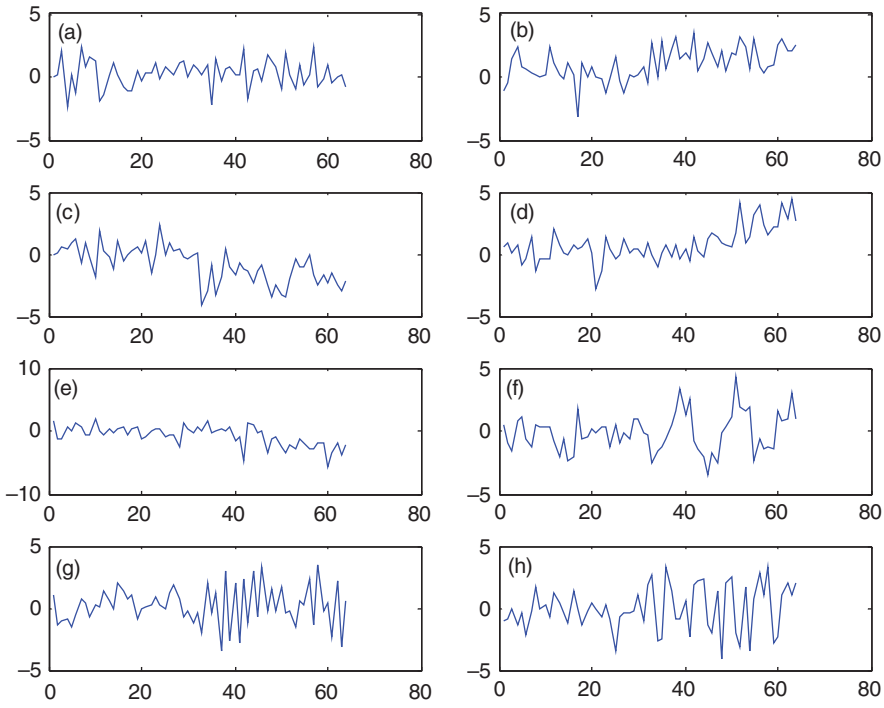

Figure 2. Eight categories of control chart patterns: (a) normal condition, (b) upward shift, (c) downward shift, (d) upward trend, (e) downward trend, (f) cycle, (g) systematic, and (h) mixture.

raw data into a recogniser, which is the common method of generating the feature vector, the statistical features to be extracted from the raw data are used as input vectors of GMMs.

The statistical features can strengthen the pattern feature so as to improve the performance of a recogniser. Smith (1994) used raw data, statistical features (i.e. standard and range) or integrating them as an input vector in BPNs to recognise the means shifts of discrete and independent processes. The results showed that the BPN integrating raw data and statistical features as input feature vectors improves the performance of ANNs. Hassan *et al.* (2003) conducted an experimental study using BPNs to identify six types of basic SPC patterns, where the performances of two BPN recognisers using statistical features and raw data as input feature, respectively, were compared. The results indicated that the BPN using statistical features as input vectors has better performance than the other BPN using raw data as input vectors. Ranaee *et al.* (2010) used shape and statistical features as inputs of SVM for CCPR. A particle swarm optimisation model was used to select the effective feature subset to improve the recognition performance of SVM. Yu and Xi (2009b) proposed a statistical-feature-based ANN ensemble model to recognise the out-of-control sources of abnormal signals in multivariate processes. The experimental results showed that the statistical feature is capable of reducing the dimension of the input feature and training time cost while maintaining good generalisation performance.

In this study, the mean, standard deviation (STD), kurtosis, skewness, autocorrelation, root-mean-square (RMS), and peak-to-peak (P-P) extracted from raw data are used as input features of the GMM model. These features are defined as follows (x_i is the sample and N is the number of samples):

$$\text{Mean} = \frac{1}{N}\sum_{i=1}^{N} x_i, \tag{13}$$

$$\text{STD} = \left(\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2\right)^{1/2}, \tag{14}$$

$$\mathrm{RMS} = \left(\frac{1}{\mathrm{N}}\sum_{i=1}^{N} x_i^2\right)^{1/2}, \tag{15}$$

$$\mathrm{Skewness} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^3/\sigma^3, \tag{16}$$

$$\mathrm{Kurtosis} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^4/\sigma^4, \tag{17}$$

$$\mathrm{Autocorrelation} = \left(\frac{1}{N-1}\sum_{i=1}^{N-1}(x_i x_{i+1})\right), \tag{18}$$

$$\mathrm{P-P} = \max(x_i) - \min(x_i). \tag{19}$$

The majority of the above features are commonly used in statistics applications. The mean square value is the 'average power' of the input vector. Kurtosis is a measure of how outlier-prone a distribution is. The skewness is a measure of the asymmetry of the data around the sample mean. Autocorrelation measures the dependence of a data at one instant in time with another data at another instant in time.

Wavelet transform is a very powerful tool in time-frequency domain analysis. The major advantage of wavelet transform is the presentation of signals in time-frequency distribution diagrams with multi-resolution (Burrus *et al.* 1998). The discrete wavelet transform of the function $f(n)$ with wavelet $\psi(n)$ and scaling $\phi(n)$ functions is given as follows:

$$f(n) = \frac{1}{\sqrt{M}} W_\varphi(j_0, k)\phi_{j_0,k}(n) + \frac{1}{\sqrt{M}}\sum_{j=j_0}^{\infty}\sum_{k} W_\psi(j,k)\psi_{j,k}(n), \tag{20}$$

where j_0 is an arbitrary starting scale and $W_\varphi(j_0, k)$ and $W_\psi(j,k)$ are real-valued 'approximation' and 'detail' expansion coefficients. $f(n)$, $\phi_{j_0,k}(n)$ and $\psi_{j,k}(n)$ are functions of the discrete variable $n = 0, 1, 2, \ldots, M-1$. The first sum in Equation (20) uses scaling functions to provide an approximation of $f(n)$ at scale j_0. For each larger scale $j > j_0$ in the second sum, a finer resolution function – a sum of wavelets – is added to the 'approximation' to provide 'details'.

The wavelet energy can represent the characteristics of sample signals from processes (Assaleh and Al-assaf 2005), and thus is used as the input features of CCPR models. The wavelet energy can be calculated as follows:

$$E_j = \begin{cases} W_\varphi(0,0), & j = 1, \\ \frac{\sum_{k=2^{j-2}+1}^{2^{j-1}} |W_\psi(j,k)|^2}{2^{j-2}}, & j = 2, 3, \ldots, 7. \end{cases} \tag{21}$$

In this study, we use Daubechies's (1988) system of wavelets, because it has the advantages of orthogonality, compact support in the time domain and computational simplicity. A specific Daubechies wavelet 'db5' in five levels is used to generate the wavelet energy. Because the wavelet energy of the input signals often concentrates on the first several levels of wavelet transformation and too large a dimension of the input features reduces the performance of GMM, the first three level wavelet energies are chosen as three input features of GMMs. Finally, 10 features (i.e. seven statistical features and three wavelet energy features (WE1–WE3)) are generated as the input features of GMMs to implement CCPR.

In order to observe the effectiveness of the used features for separating the CCPs, the box plots and the feature space plots of the 10 features for the six classes generated by this process are presented in Figure 3. It can be observed from the figure that the data distributions of the features related to the different CCPs are located close to each other and these features separate the six CCPs relatively well. However, using only one feature or several features cannot separate the six CCPs effectively. An effective approach is to fuse all these features to implement CCPR.

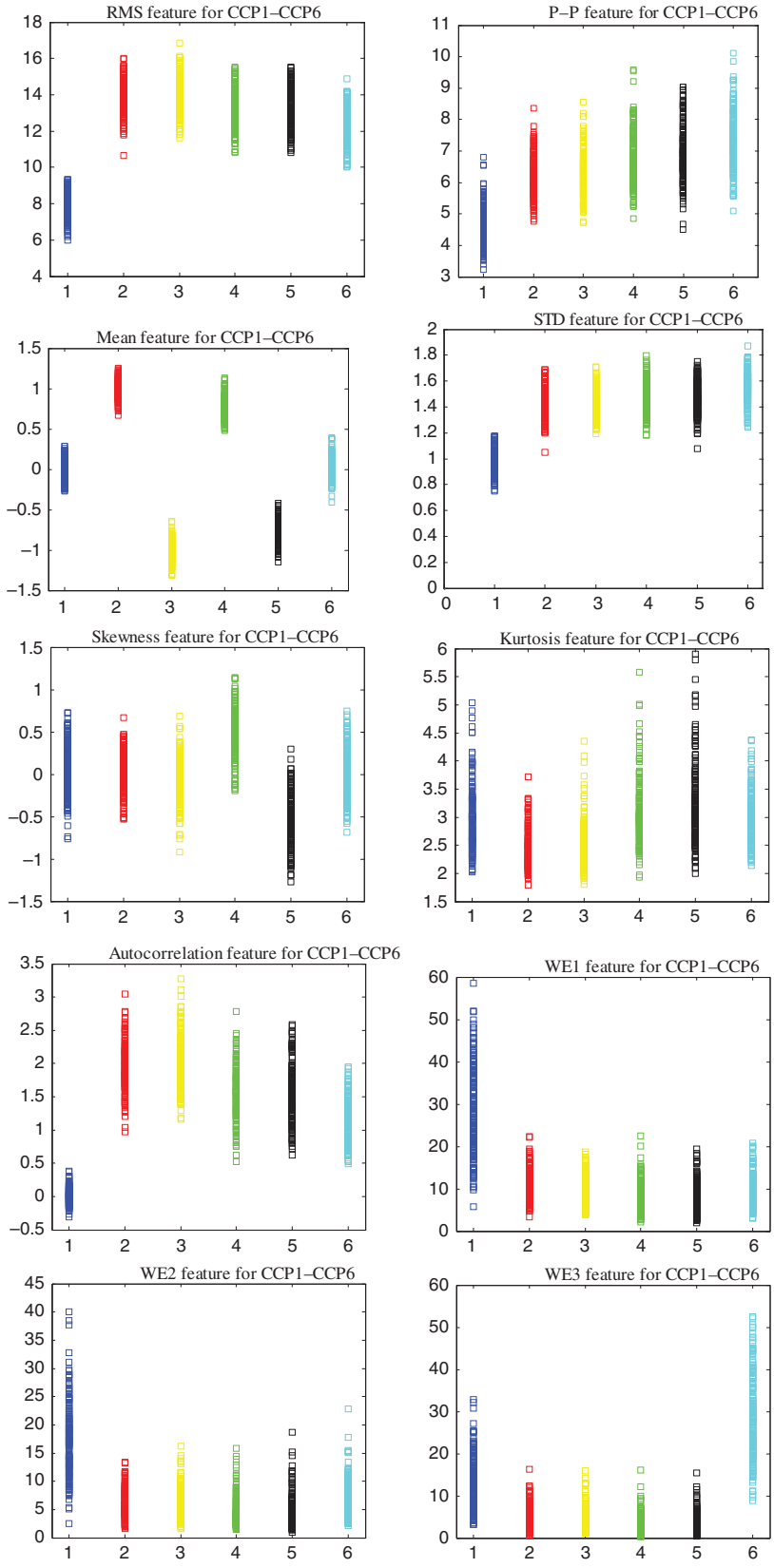

Figure 3. Box plots of the 10 features used for CCP1–CC6 (1=CCP1, 2=CCP2, . . . , CCP6=6).

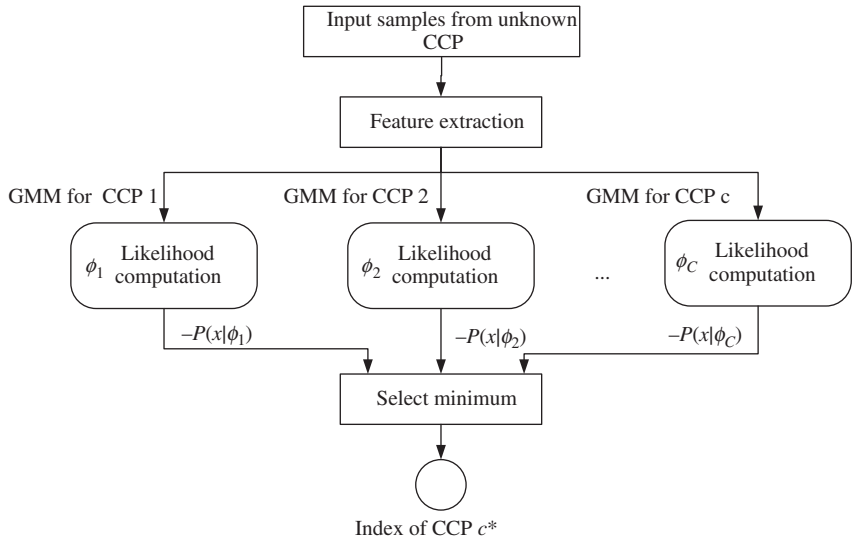

Figure 4. System diagram of the procedure for control chart pattern recognition.

3.3 *GMM-based CCPR model*

Once GMM is created based on the training data, it is then used to recognise CCPs. A quantification criterion is needed to determine which GMM recognises a new input. For each new input, GMM provides $P(x)$ (see Equation (1)), the unconditional probability density (log likelihood), which indicates how the input follows the probability distribution of the GMM trained by the data set with one CCP. Thus, the log likelihood probability can be a good quantification criterion for recognising CCPs. In general, since the log likelihood values are smaller than zero, to improve intelligibility, the negative log likelihood probability (NLLP) is used as the quantification index in this study:

$$NLLP = -\log P(x|\phi). \tag{22}$$

The proposed approach consists of two phases: model training and pattern recognition. The following two steps are involved in the design of the CCPR model.

Step I: The first step is to build the GMM λ_c for each CCP c. In other words, we must estimate the model parameters $\phi = (\pi_1, \ldots, \pi_M; \theta_1, \ldots, \theta_M)$ of each GMM that optimise the likelihood (or maximise $P(x|\phi_c)$) of the training set that only includes samples with the cth CCP using the EM algorithm. A total of six independent GMMs are constructed for the six CCPs (i.e. CCP1–CCP6).

Step II: The procedure for recognising CCPs is presented in Figure 4. As shown, the features of the new input sample x are extracted and then input to all the trained GMMs in Step I. This is then followed by calculation of the model likelihoods (i.e. NLLPs) of all GMMs, $P(x|\phi_c)$, $1 \le c \le C$, which is the probability of the input vector x, given by the GMM c. Finally, C NLLPs are obtained from the C GMMs, and the GMM c^* with the lowest NLLP is considered to be the best candidate for representing the current CCP, i.e.

$$c^* = \arg\min_{1 \le c \le C}[-\log(P(x|\phi_c))]. \tag{23}$$

4. Experiments and analysis of the results

To illustrate the performance of the proposed model, we apply it to CCPR in independent and discrete manufacturing processes that exhibit various process patterns, and then compare it with other recogniser-based CCPR models. Moreover, extensive studies such as adaptation recognition and online CCPR of the GMM-based models are further implemented to illustrate its good performance and potential real-world applications.

Table 1. Recognition rates (%) for 10 runs of the GMM-based CCPR model.

Run	CCP1	CCP2	CCP3	CCP4	CCP5	CCP6	AVG
1	99	97	93	91	91	100	95.17
2	100	91	89	94	90	100	94.00
3	100	91	94	93	91	100	94.83
4	100	91	89	90	90	100	93.33
5	100	95	90	95	95	100	95.83
6	99	92	89	90	95	100	94.17
7	99	89	91	89	93	99	93.33
8	100	94	90	90	90	100	94.00
9	100	89	91	91	93	100	94.00
10	100	92	92	90	91	100	94.17
Total AVG							94.28

4.1 *Experimental setup*

For each CCP category, 200 samples are generated as the training data set to train each GMM (see the detailed construction scheme of GMMs in Section 3). The EM algorithm is run for certain cycles to train each GMM for each CCP and the final error, which is equivalent to the total log-likelihood value, is recorded as the evaluation standard of the trained GMM. The number of Gaussian components is set to 2–6, and the FJ index is used to choose the best GMM with the largest FJ value as the CCPR model. The termination for the EM is as follows. (1) Terminate when the total number of generations specified by the user is reached (1000). (2) Terminate when the log-likelihood value of the GMM cannot be improved in two consecutive iterations. Termination criterion 2 is often met in the following experiments, which means that the convergence of the EM is sufficiently fast not to have to implement 1000 iterations.

4.2 *Control chart pattern recognition*

GMM is trained using the data with only one CCP type, and then a total of six GMMs are constructed and combined to recognise CCPs. In this study, the six GMMs are denoted GMM1, GMM2, ..., GMM6 for the corresponding six types of CCPs, respectively. For each type of CCP, 100 samples are generated as the testing data set to evaluate the CCPR performance of GMMs. We implemented such testing 10 times to obtain the average recognition rate. Table 1 presents the results of the 10 runs. It can be seen that GMMs recognised 98.87% of samples for all six CCPs successfully. In particular, GMMs show perfect performance for recognising normal and cycle patterns because they recognised all samples of CCP1 and CCP6. Intuitively, most classification errors result from misclassifying a shift pattern (or trend pattern) as a trend pattern (or shift pattern) due to the similarity of their pattern characteristics.

In order to illustrate the detailed recognition information of the GMM-based model for the six CCPs, the recognition result (i.e. the fourth row of Table 1) with the least accuracy in the 10 tests is presented in Table 2. Clearly, most classification errors result from misidentifying a shift (or trend) as a trend (or shift).

The BPN, RBF and PNN models are used widely to implement CCPR (e.g., Plummer (1993), Cook and Chiu (1998), Hassan *et al*. (2003) and Guh (2005)). Thus, we further compared GMM with BPN, RBF and PNN to further demonstrate its good CCPR performance. The parameters of BPN, RBF and PNN were set as follows.

BPN: The network structure has 10 input nodes, one hidden layer with 15 hidden nodes, and six output nodes. The umber of iterations is 1000. The learning rate and momentum factor were set to 0.1 and 0.4, respectively. The initial weights were randomly set between [−0.01, +0.01]. The hyper-tangent (tansig) and sigmoid (logsig) functions were used as the activation functions of the hidden and output layer, respectively. Prior experimental results showed that the use of activation functions (i.e. tansig and logsig) for the hidden and output layers does not affect the performance of BPN for CCPR significantly.

RBF: The network construction function 'newrb()' from the ANN toolbox of the Matlab system (The MathWorks Company 2007) is used to train the RBF network. Several key parameters for the 'newrb()' function are set as

Table 2. Recognition of the six predefined CCPs by the GMM-based CCPR model.

	Normal	Upward shift	Downward shift	Upward trend	Downward trend	Cycle
Normal	100	0	0	0	0	0
Upward shift	0	91	0	9	0	0
Downward shift	0	0	89	0	11	0
Upward trend	0	10	0	90	0	0
Downward trend	0	0	10	0	90	0
Cycle	0	0	0	0	0	100

Table 3. Performance comparison of BPN, RBF, PNN and GMM.

	PPR model			
	BPN	RBF	PNN	GMM
Recognition accuracy (%)	93.58	80.9	83.47	94.28

Table 4. *P*-values for BPN, RBF and PNN.

BPN	RBF	PNN
0.0485	1.9709e–016	1.4809e–016

follows: spread of radial basis functions (2.0), maximum number of neurons (1000), and number of neurons to add between displays (25). For detailed information on 'newrb()', the interested reader can refer to the literature (The MathWorks Company 2007). A prior experiment was implemented to obtain the above parameter values to train a relatively good RBF model.

PNN: The network construction function 'newpnn()' from the ANN toolbox of the Matlab system (The MathWorks Company 2007) is used to train the PNN network. The key parameter for this training function is set as follows: spread of radial basis functions (2.0). For detailed information on 'newrb()', the interested reader can refer to the literature (The MathWorks Company 2007).

The testing dataset used by GMMs is to test the recognition performance of BPN, RBF and PNN, and 10 runs are implemented and their mean recognition rate obtained. As shown in Table 3, GMM demonstrates a better recognition performance than that of BPN, RBF and PNN. This comparison further demonstrates that the GMM-based CCPR model with an unsupervised learning scheme has performance comparable to that of ANN with a supervised learning scheme. The greatest difference between our model and BPN, RBF and PNN is that only one GMM is constructed for one type of CCP and then all trained GMMs are combined to recognise all CCPs. Thus, our recognition model is not fixed and can dynamically add new GMMs for novel CCPs without reconstructing the original recognition model, which is very important for improving the engineering applicability in real-world manufacturing processes.

Apart from the recognition rate comparison, a paired *t*-test (Montgomery *et al*. 2001) was performed and the results are listed in Table 4. For all the paired *t*-tests, GMM is used as the control group, while the other CCPR models are the treatment groups. The alpha level was chosen as 0.05. The *P*-values represent the probability for the null hypothesis that the accuracies of the two CCPR models are equal is false. It can be observed from Table 4 that small *P*-values were obtained, indicating that the CCPR models are not equal for the testing dataset used.

4.3 *Adaptive capability analysis for novel patterns*

Due to the complicated working conditions in some manufacturing processes, novel (unexpected) CCPs could occur dynamically and randomly, and need to be recognised online. The adaptive capability of the proposed model with respect to novel patterns is demonstrated in this section. Before implementation of the adaptation of the GMM-based CCPR model, a confidence limit η_{NLLP} for each trained GMM is needed to determine whether or not the novel CCPs occur in the manufacturing processes. Note that the latent variables in GMM do not follow a Gaussian distribution, thus the confidence limit η_{NLLP} for each CCP cannot be determined directly from a particular

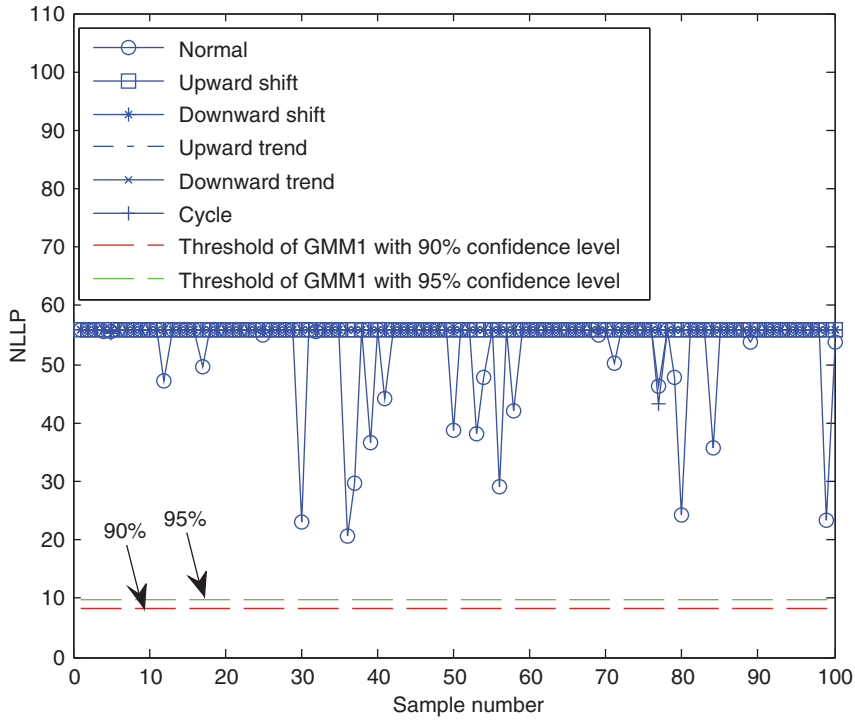

Figure 5. NLLP outputs of the six trained GMMs for the novel CCP CCP7 (i.e. systematic).

approximate distribution. An alternative approach to defining the specific operating region of GMM is to use kernel density estimation (KDE) (Martin and Morris 1996). Here, a univariate kernel estimator can be defined by

$$\widehat{f}(x,h) = \frac{1}{nh}\sum_{i=1}^{n} K\Big(\frac{x - x_i}{h}\Big), \tag{24}$$

where x is the data point under consideration, x_i is an observation value from the data set, h is the bandwidth matrix (called the smoothing parameter), and K is a kernel function that satisfies the following conditions:

$$K(x) > 0 \text{ and } \int_{R^m} K(x)\mathrm{d}x = 1. \tag{25}$$

There are certain kernel functions used for K. However, in practice, the Gaussian kernel function is often used.

When using the density- estimation- based approach to define the confidence limit for each GMM corresponding to each CCP, the density function of NLLP is calculated based on the training data of each CCP. In this study, the function 'ksdensity' from the statistics toolbox of the Matlab system software is used to compute the density estimate using the kernel-smoothing method, where the default parameters of the Matlab system software are used. The setup of the smoothing parameter h can be obtained from the literature (Jongnum *et al.* 2004), and thus $h = (4/(3 * N))^{\wedge}(1/5) * \sigma$ (σ is the estimated standard deviation of the variable) is used to calculate the threshold of each GMM in this study. Depending on the confidence level, the threshold η_{NLLP} can be calculated to define the confidence limit for the NLLP of each GMM. In this study, a confidence level of 95% (or 90%) is used, and η_{NLLP} is obtained for each GMM (i.e. the NLLP outputs of the GMM for the 95% (or 90%) case should be below η_{NLLP}). An important issue for monitoring models is to reduce false alarms, which result in the costly stopping of the process. It should be noted that a balance between the false alarm rate and the missing alarm rate should be obtained based on the requirements of real-world applications.

In particular, 100 samples for each novel pattern (i.e. systematic (i.e. CCP7) and mixture (i.e. CCP8)) are generated to evaluate the adaptive capability of the GMM-based CCPR model. Figures 5 and 6 show the outputs of the six trained GMMs for the two novel CCPs. The thresholds for GMM1 and GMM6 are set up as in Figures 5 and 6, respectively, because GMM1 and GMM6 exhibit the smallest NLLPs. As shown, the NLLPs of all six

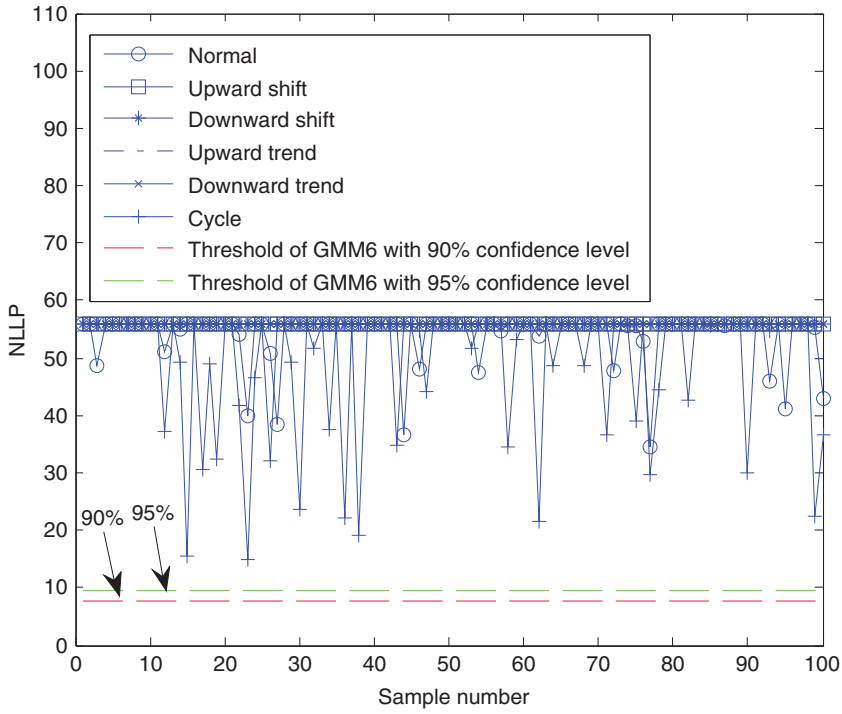

Figure 6. NLLP outputs of the six trained GMMs for the novel CCP CCP8 (i.e. mixture).

Table 5. Recognition of the two novel CCPs CCP7 and CCP8 by the GMM-based CCPR model.

	Normal	Upward shift	Downward shift	Upward trend	Downward trend	Cycle	Novel pattern
CCP7 (systematic)	2	0	0	0	0	0	98
CCP8 (mixture)	10	0	0	2	0	14	74

GMMs exceed the 90% and 95% confidence limits, which indicates that the CCPR model with six GMMs detected the two novel CCPs successfully.

According to the above experimental results, the GMM-based CCPR model is capable of detecting the novel patterns occurring in manufacturing processes. In order to recognise the novel CCPs (i.e. CCP7 and CCP8), two new GMMs should be constructed and then added to the CCPR model. Similar to the training procedure for the six trained GMMs, 200 samples of the novel patterns CCP7 and CCP8 are generated to construct two new GMMs, i.e. GMM7 and GMM8, respectively. GMM7 and GMM8 are then added to the GMM-based CCPR model and all eight GMMs are combined to recognise CCPs. The 100 testing samples for each novel pattern CCP7 and CCP8 are then input into the new GMM-based CCPR model; the recognition results are presented in Table 5. It can be observed that the new GMM-based CCPR model can recognise the novel CCP (CCP7) successfully. Because the mixture pattern is similar to the normal and cycle patterns, the recognition performance of the new GMM-based CCPR model is relatively poor.

This experimental result illustrates that the proposed model is capable of adapting those novel patterns occurring in manufacturing processes. Once the continual outputs of the trained GMMs exceed the corresponding threshold η_{NLLP} by a certain confidence bound, it can be determined that a novel pattern occurs in the manufacturing process, and a new GMM should be constructed and then added to the CCPR model to recognise it. Thus, the proposed GMM-based CCPR model is not fixed and can help on-line process engineers to discover new CCPs and construct an on-line adaptive learning-based CCPR model.

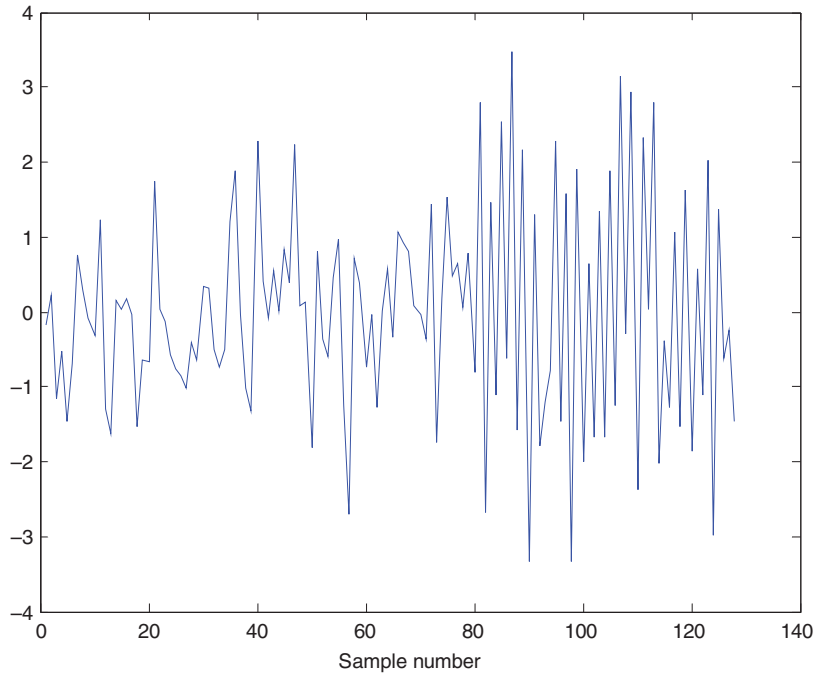

Figure 7. Simulated process for novel CCP CCP7 (i.e. systematic pattern).

4.4 *Extensive study for online monitoring and recognition*

This section describes an extensive study that was implemented to further illustrate the online monitoring and recognition performance of the GMM-based CCPR model. The manufacturing process will start in an in-control state, and then abnormal patterns will occur gradually in the process. In this study, abnormal patterns CCP7 (i.e. systematic) and CCP8 (i.e. mixture) are considered. To generate a NLLP chart for each process, moving window vectors with 64 samples over time series were generated. The simulation process begins with the first in-control 77 points (i.e. the first 14 in-control moving window vectors), with the systematic (or mixture) pattern occurring at point 78. The simulation data series for CCPs CCP7 and CCP8 are presented in Figures 7 and 8, respectively. After feature extraction, the two simulation data sets are input into the eight trained GMMs, and Figures 9 and 10 present the recognition results for the two processes.

It can be observed from Figures 9 and 10 that GMM1 has the smallest outputs (i.e. NLLP) compared with those of the other seven GMMs for the first 25 moving window vectors for both abnormal patterns, which indicates that the proposed model does not trigger false alarms for the first in-control 77 points and then detects the out-of-control signal at point 88 (the first abnormal point occurring in this process is at point 78), i.e. the out-of-control ARL of the proposed model is 10 for both processes.

From Figure 9, after the first 10 abnormal points enter into the moving window, the NLLPs of GMM7 corresponding to a systematic pattern are smaller than those of other GMMs, and decrease gradually as more abnormal points enter into the moving window, which means that the GMM-based model detected and recognised CCP7 with more abnormal points entering into the moving window. From Figure 10, after the first 10 abnormal points enter into the moving window, the NLLPs of GMM6 corresponding to a cycle pattern are smaller than those of other GMMs, which further demonstrates that the mixture pattern can be misclassified as a cycle pattern by GMMs. However, the proposed CCPR model recognises the novel mixture pattern at moving window point 51 with more abnormal points entering into the moving window. This experimental result further shows that the proposed model is capable of detecting and recognising novel abnormal patterns (e.g., the CCP changes from normal to systematic).

5. Conclusions and future study

We propose a novel GMM-based model to identify typical CCPs in independent and discrete manufacturing processes. The salient feature of the model is the combination of several GMMs to implement CCPR, where each

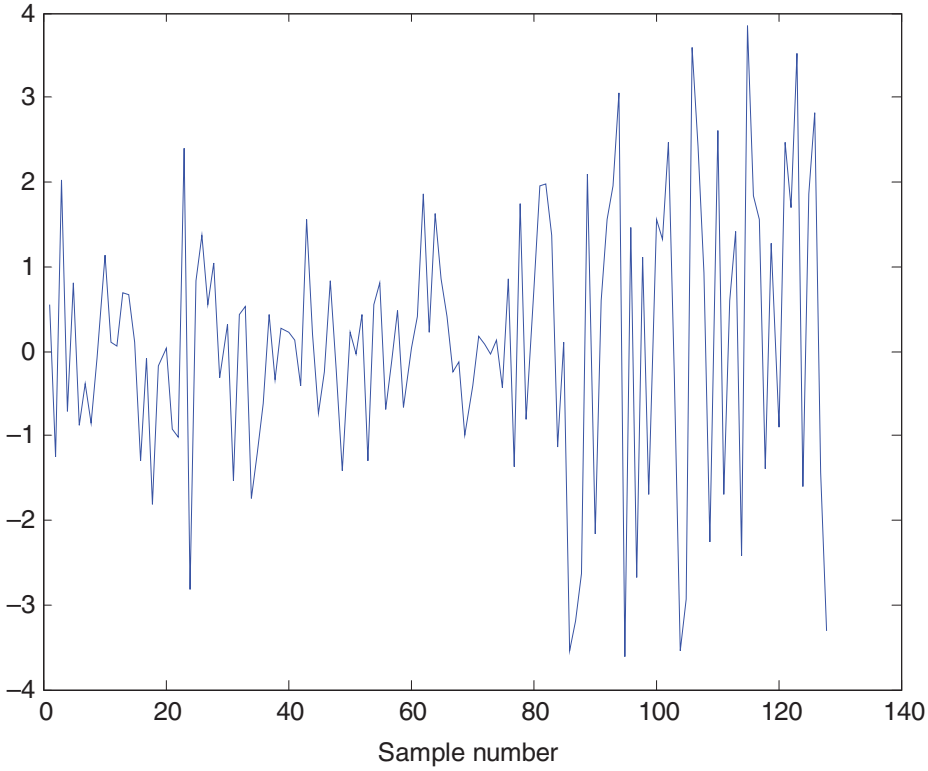

Figure 8. Simulated process for novel CCP CCP8 (i.e. mixture pattern).

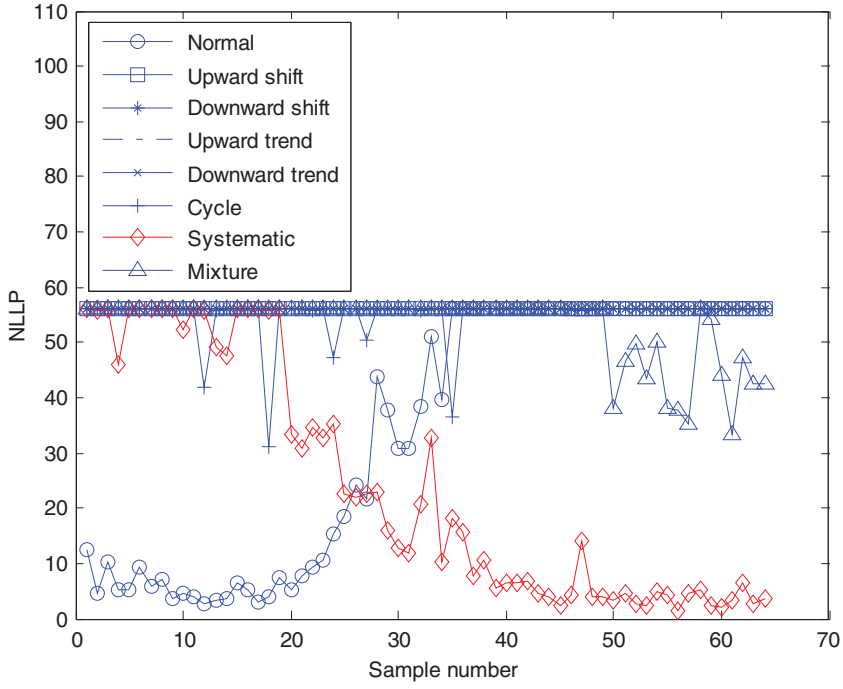

Figure 9. NLLP outputs of the GMM-based CCPR model for novel CCP CCP7 (i.e. systematic pattern).

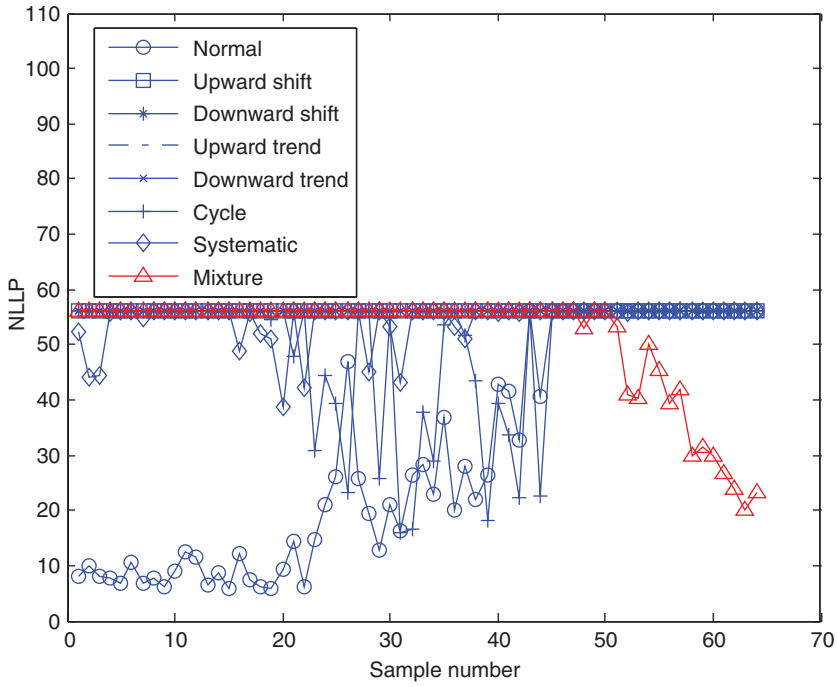

Figure 10. NLLP outputs of the GMM-based CCPR model for novel CCP CCP8 (i.e. mixture pattern).

GMM is constructed for just one type of CCP. The experimental results indicate that the GMM-based model is very effective in identifying CCPs and is superior to other single recognisers (e.g., BPN, RBF, and PNN).

In addition, the performance of the GMM-based CCPR model is further evaluated according to the online learning capability with respect to unexpected novel CCPs. The experimental result demonstrates that the GMM-based model not only incorporates similar patterns (stability), but also readily adapts novel patterns (plasticity). In comparison with CCPR models with a fixed structure, the GMM-based CCPR model is capable of updating itself according to changes in manufacturing process states, and thus provides a more promising way to adapt unexpected CCPs. Furthermore, the GMM-based model shows potential application for the online detection and recognition of CCPs. According to the empirical analysis and evaluation results, we can draw the conclusion that the GMM-based CCPR model is an effective tool for quality control. We are currently extending this work to monitor and diagnose out-of-control signals of multivariate process control (Yu and Xi 2009a,b).

Acknowledgements

This work is supported by the National Science Foundation of China (No. 71001060), the Open Fund of the State Key Laboratory for Manufacturing Systems Engineering (No. 2010008) and the Research Fund for the Doctoral Program of Higher Education of China (No. 20103108120010).

References

Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19 (6), 716–723.

Assaleh, K. and Al-assaf, Y., 2005. Feature extraction and analysis for classifying causable patterns in control charts. *Computer & Industrial Engineering*, 49 (1), 168–181.

Burrus, C., Gopinath, R., and Guo, H., 1998. *Introduction to wavelets and wavelet transforms*. Engewood Cliffs, NJ: Prentice Hall.

Chang, S.I. and Ho, E.S., 1999. A two-stage neural network approach for process variance change detection and classification. *International Journal of Production Research*, 37 (7), 1581–1599.

Cook, D.F. and Chiu, C.C., 1998. Using radial basis function neural networks to recognize shifts in correlated manufacturing process parameters. *IIE Transactions*, 30 (3), 227–234.

Daubechies, I., 1988. Orthonormal bases of compact supported wavelets. *Communications on Pure and Applied Mathematics*, 41 (7), 909–996.

Dempster, A., Laird, N., and Rubin, D., 1977. Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 30 (B), 1–38.

Duncan, A.J., 1986. *Quality control and industrial statistics*. 5th ed. Illinois: Irwin Book Company.

Figueiredo, M.A.T. and Jain, A.K., 2002. Unsupervised learning of finite mixture models. *IEEE Transactions on Pattern Analysis and Artificial Intelligence*, 24 (3), 381–396.

Gauri, S.K. and Chakraborty, S., 2007. A study on the various features for effective control chart pattern recognition. *International Journal of Advanced Manufacturing Technology*, 34 (3–4), 385–398.

Grant, E.L. and Leavenworth, R.S., 1996. *Statistical quality control*. 7th ed. New York: McGraw-Hill.

Guh, R.S., 2004. Optimizing feedforward neural networks for control chart pattern recognition through genetic algorithms. *International Journal of Pattern Recognition and Artificial Intelligence*, 18 (2), 75–99.

Guh, R.S., 2005. A hybrid learning-based model for on-line detection and analysis of control chart patterns. *Computers & Industrial Engineering*, 49 (1), 35–62.

Hassan, A., *et al*., 2003. Improved SPC chart pattern recognition using statistical features. *International Journal of Production Research*, 41 (7), 1587–1603.

Hwarng, H.B. and Chong, C.W., 1995. Detecting process non-randoms through a fast and cumulative learning ART-based pattern recognizer. *International Journal of Production Research*, 33 (7), 1817–1833.

Jiang, P., Liu, D., and Zeng, Z., 2009. Recognizing control chart patterns with neural network and numerical fitting. *Journal of Intelligent Manufacturing*, 223 (6), 623–640.

Jongnum, L., Yoo, C.K., and Lee, I.B., 2004. Statistical monitoring of dynamic process based dynamic independent component analysis. *Chemical Engineering Science*, 59 (11), 2995–3006.

Martin, E.B. and Morris, A.J., 1996. Non-parametric confidence bounds for process performance monitoring charts. *Journal Process Control*, 6 (6), 349–358.

Montgomery, D.C., Runger, G.C., and Hubele, N.F., 2001. *Engineering statistics*. 2nd ed. New York: Wiley.

Nelson, L.S., 1984. The Shewhart control chart-tests for special causes. *Journal of Quality Technology*, 16 (4), 237–239.

Pham, D.T. and Oztemel, E., 1994. Control chart pattern recognition using learning vector quantization networks. *International Journal of Production Research*, 32 (3), 721–729.

Plummer, J., 1993. Tighter process control with neural networks. *AI Expert*, 8 (10), 49–55.

Ranaee, V., Ebrahimzadeh, A., and Ghaderi, R., 2010. Application of the PSO–SVM model for recognition of control chart patterns. *ISA Transactions*, 49 (4), 577–586.

Rissansen, J., 1987. Stochastic complexity. *Journal of the Royal Statistical Society, Series B (Methodological)*, 59 (4), 223–239.

Schwarz, G., 1978. Estimating the dimension of a model. *Annals of Statistics*, 6 (2), 461–464.

Smith, A.E., 1994. X-bar and R control chart interpretation using neural computing. *International Journal of Production Research*, 32 (2), 309–320.

The MathWorks Company, 2007. *Neural network toolbox*. Math Works Inc.

Wang, C.H. and Kuo, W., 2007. Identification of control patterns using wavelet filtering and robust fuzzy clustering. *Journal of Intelligent Manufacturing*, 18 (10), 343–350.

Wang, C.H., Kuo, W., and Qi, H., 2007. An integrated approach for process monitoring using wavelet analysis and competitive neural network. *International Journal of Production Research*, 45 (1), 227–244.

Wang, C.H., *et al*., 2008. Decision tree based control chart pattern recognition. *International Journal of Production Research*, 46 (17), 4889–4901.

Western Electric Company, 1958. *Statistical quality control handbook*. Indianapolis: Western Electric.

Yang, J.H. and Yang, M.S., 2005. A control chart pattern recognition system using a statistical correlation coefficient method. *Computers & Industrial Engineering*, 48 (2), 205–221.

Yu, J.B. and Xi, L.F., 2009a. A neural network ensemble-based model for on-line monitoring and diagnosis of out-of-control signals in multivariate manufacturing processes. *Expert Systems with Applications*, 36 (1), 909–921.

Yu, J.B. and Xi, L.F., 2009b. A hybrid learning-based model for on-line monitoring and diagnosis of out-of-control signals in multivariate manufacturing processes. *International Journal of Production Research*, 47 (15), 4077–4108.

Yu, J.B., Xi, L.F., and Wu, B., 2007. A neural network ensemble approach for the recognition of SPC chart patterns. *In: Proceedings of the third international conference on natural computation*, 2, 575–579.

Yu, J.B., Xi, L.F., and Zhou, X., 2009. Identifying source(s) of out-of-control signals in multivariate manufacturing processes using selective neural network ensemble. *Engineering Applications of Artificial Intelligence*, 22 (1), 141–152.