A Survey of Fault Diagnosis and Fault-Tolerant Techniques Part I: Fault Diagnosis with Model-Based and Signal-Based Approaches

Zhiwei Gao, Senior Member, IEEE, Carlo Cecati, Fellow IEEE, and Steven X. Ding

Abstract—With the ever increase of complexity and expense of industrial systems, there is less tolerance for performance degradation, productivity decrease and safety hazards, which greatly stimulates to detect and identify any kinds of potential abnormalities and faults as early as possible, and implement real-time fault-tolerant operation for minimizing performance degradation and avoiding dangerous situations. During the last four decades, fruitful results were reported about fault diagnosis and fault-tolerant control methods and their applications in a variety of engineering systems. The three-part survey paper aims to give a comprehensive review for real-time fault diagnosis and fault tolerant control with particular attention on the results reported in the last decade. In the first part review, fault diagnosis approaches and their applications are reviewed comprehensively from model-based and signal-based perspectives, respectively.

Index Terms—Analytical redundancy, model-based fault diagnosis, signal-based fault diagnosis, real-time monitoring, fault tolerance

I. INTRODUCTION

As is known, many engineering systems, such as aero engines, vehicle dynamics, chemical processes, manufacturing systems, power network, electric machines, wind energy conversion systems, and industrial electronic equipment and so forth, are safety-critical systems. There is an ever increasing demand on reliability and safety of industrial systems subjected to potential process abnormalities and component faults. As a result, it is paramount to detect and identify any kinds of potential abnormalities and faults as early as possible, and implement fault-tolerant operation for minimizing performance degradation and avoiding dangerous situations.

A fault is defined as an unpermitted deviation of at least one characteristic property or parameter of the system from the acceptable/usual/standard condition [1]. Examples of such malfunctions are the blocking of an actuator, the loss of a sensor (e.g., a sensor gets stuck at a particular value or has a variation in the sensor scalar factor), or the disconnection of a system component. Therefore, the faults are often classified as actuator faults, sensor faults and plant faults (or called component faults or parameter faults), which either interrupt the control action from the controller on the plant, or produce substantial measurement errors, or directly change the dynamic input/output properties of the system, leading to system performance degradation, and even the damage and collapse of the whole system. In order to improve the reliability of a system concerned, fault diagnosis is usually employed to monitor, locate and identify the faults by using the concept of redundancy, either hardware redundancy or software redundancy (or called analytical redundancy). The basic idea of the hardware redundancy is to use identical components with the same input signal so that the duplicated output signals can be compared leading to diagnostic decision by a variety of methods such as limit checking, and majority voting etc. The hardware redundancy is reliable, but expensive and increasing weights and occupying more space. It is necessary for key components to equip with the redundant duplicate, but would not be applicable if the hardware redundancy is applied to the whole system due to the cost or the difficulty for physical installing when the space and/or weight are strictly constrained. With the mature of modern control theory, the analytical redundancy technique has become the mainstream of the fault diagnosis research since the 1980s, whose schematic diagram can be depicted by Fig.1. For a controlled system subjected to actuator fault $f_a$, process/component fault $f_c$, and sensor fault $f_s$, the input $u$ and output $y$ are used to construct a fault diagnosis algorithm, which is employed to check the consistency of the feature information of the real-time process carried by the input and output data against the pre-knowledge on a healthy system, and a diagnostic decision is then made by using diagnostic
logics. Compared with hardware redundancy methods, analytical redundancy diagnostic methods are more cost effective, but more challenging due to environmental noises, inevitable modeling error, and the complexity of the system dynamics and control structure.

The schematic of fault-tolerant control is depicted by Fig. 2, which is shown that fault tolerant control is integrated with fault diagnosis in essence. Real-time fault diagnosis can detect whether the system is faulty, and tell where the fault occurs and how severe the malfunction is. Based on the valuable information, the supervision system can thus take appropriate fault-tolerant actions such as off-setting the faulty signals by actuator/sensor signal compensation, tuning or reconfiguring the controller, and even replacing faulty components by redundant duplicates, so that the adverse effects from faults are accommodated or removed.

During the last four decades, fruitful results have been reported on fault diagnosis methods, fault-tolerant control techniques and their applications in various industrial processes and systems. A number of survey papers were written, for instance [2-38], which are depicted by Table 1, categorized in terms of years and methods/applications.

### TABLE 1

<table>
<thead>
<tr>
<th>Years</th>
<th>Fault Diagnosis Methods</th>
<th>Fault Diagnosis Applications</th>
<th>Fault Tolerance Methods</th>
<th>Fault Tolerance Applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>1970s</td>
<td>[2]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1980s</td>
<td>[3,4]</td>
<td></td>
<td>[27-29]</td>
<td></td>
</tr>
<tr>
<td>1990s</td>
<td>[5-8]</td>
<td>[18]</td>
<td>[30-32]</td>
<td>[36]</td>
</tr>
<tr>
<td>2000s</td>
<td>[9-14], [33]</td>
<td>[19, 20], [22-24]</td>
<td>[34]</td>
<td></td>
</tr>
<tr>
<td>2010s</td>
<td>[15-17], [35]</td>
<td>[21], [25,26]</td>
<td>[33], [35]</td>
<td>[37, 38]</td>
</tr>
</tbody>
</table>

Specifically, in 1976, Willsky presented the key concepts of analytical redundancy for model-based fault detection and diagnosis in the early survey paper [2]. More comprehensive model-based fault diagnosis methods such as parity space approaches, observer-based methods and parameter estimation techniques are reviewed by [3-9]. A three-part survey paper [10-12] on fault diagnosis was presented in 2003, respectively from the viewpoint of quantitative model-based methods, qualitative model-based methods and process history based methods. In [13], a structured and comprehensive overview of the research on anomaly detection is provided, which is referred to the problem of finding patterns in data that do not conform to expected behavior, and has an extensive use in a wide variety of applications such as intrusion detection for cyber-security, military surveillance for enemy activities, as well as fault detection in safety critical systems. In [14-16], comprehensive fault diagnostic methods were reviewed respectively from the data-driven perspective. In [17], a short review on fault detection in sensor networks was provided. With respect to fault diagnosis methods for various processes/systems applications, a couple of survey papers were addressed for mining equipment [18], electric motors [19-21], building systems (such as heating, ventilating, air-conditioning and refrigeration) [22, 23], machinery system [24, 25], and swarm systems (consisting of multiple intelligent interconnected nodes and possessing swarm capability) [26], respectively.

For fault-tolerant control, an early review paper was presented by [27] in 1991, which introduced the basic concepts of fault-tolerant control and analyzed the applicability of artificial intelligence (e.g., neural network and expert systems) to fault-tolerant control systems. In 1997, an overview of fault-tolerant control was given from the system...
development view [28]. In the same year, a comprehensive review was contributed by [29], which presented the key issues of the fault-tolerant control systems and outlined the state of the art in this field. Reconfigurable fault-tolerant control systems are reviewed extensively respectively by [30-32]. Some results on fault-tolerant control for nonlinear systems were reviewed by [33]. Along with fault diagnosis, brief reviews on data-driven fault-tolerant control and model-based fault-tolerant reconfiguration were presented by [34, 35] respectively. From the viewpoint of industrial applications, fault tolerance techniques were reviewed for electric drive systems [36] and power electronics systems [37, 38], respectively.

The three-part survey paper aims to give a comprehensive overview for real-time fault diagnosis and fault tolerant control with particular attention on the results reported in the last decade. Generally, fault diagnosis methods can be categorized into model-based methods, signal-based methods, knowledge-based methods, hybrid methods (the combination methods of at least two methods) and active fault diagnosis methods. In the first-part survey paper, fault diagnosis techniques will be reviewed from the model-based and signal-based perspectives, and the knowledge-based, hybrid, and active fault diagnosis techniques will be reviewed in the second-part survey paper. The first-part and second-part survey papers aim to review the existing fault diagnosis methods and applications within a framework by using the up-to-date references.

The rest of this paper is organized as follows. Following the introduction session, model-based fault diagnosis techniques are reviewed in Section II. Signal-based fault diagnosis is reviewed in Section III. The paper is ended by Section IV with conclusions.

II. MODEL-BASED FAULT DIAGNOSIS METHODS

Model-based fault diagnosis was originated by Beard [39] in 1971 in order to replace hardware redundancy by analytical redundancy, and comprehensive results were documented in some well-written books (i.e., see [40, 41]). In model-based methods, the models of the industrial processes or practical systems are required to be available, which can be obtained by using either physical principles or systems identification techniques. Based on the model, fault diagnosis algorithms are developed to monitor the consistency between the measured outputs of the practical systems and the model predicted outputs. In this section, model-based fault diagnosis methods are reviewed following the four categories: deterministic fault diagnosis methods, stochastic fault diagnosis methods, fault diagnosis for discrete-events and hybrid systems, and fault diagnosis for networked and distributed systems, which are classified in terms of types of the models used.

A. Deterministic fault diagnosis methods

Observer plays a key role in model-based fault diagnosis for monitored systems/processes characterized by deterministic models. The schematic diagram of the observer-based fault diagnosis is depicted by Fig. 3, which includes fault detection, fault isolation and fault identification (or called fault reconstruction or fault estimation).

![Observer-Based Fault Diagnosis](image)

Fig. 3. Scheme of model-based fault diagnosis

For simplicity, the model of the process in Figure 3 is assumed to be linearized state-space model, which is described by the following:

\[
\begin{aligned}
\dot{x}(k+1) &= (A + \Delta A)x(k) + (B + \Delta B)u(k) + B_d d(k) \\
y(k) &= (C + \Delta C)x(k) + D_d f_a(k) + D_o \omega(k)
\end{aligned}
\]

where

- \( x(k) \) is the system state,
- \( u(k) \) is the control input,
- \( y(k) \) is the measured output,
- \( d(k) \) is the system state, control input, measured output, unexpected actuator fault, component/parameter fault, sensor fault, process disturbance and measurement noises, respectively.
- \( A, B, C, D, \Delta A, \Delta B, \Delta C \) are known parameter matrices, and \( \Delta A, \Delta B, \Delta C \) are unknown modelling parameter errors.

An observer-based fault detection filter is given in the following form:

\[
\begin{aligned}
\hat{x}(k+1) &= A\hat{x}(k) + B\hat{r}(k) + K\hat{r}(k) \\
\hat{y}(k) &= C\hat{x}(k)
\end{aligned}
\]

where \( \hat{x}(k) \) and \( \hat{y}(k) \) are the estimates of the state and output, respectively; \( \hat{r}(k) \) is the residual signal and \( K \) is the observer gain to be designed. Let \( \epsilon(k) = x(k) - \hat{x}(k) \), the frequency-domain residual signal can be described by

\[
r(z) = G_{\Delta}(z)\epsilon(z) + G_f(z)\hat{f}(z)
\]

where

\[
\begin{aligned}
G_{\Delta}(z) &= C(\delta l - A + \Delta C)^{-1}B_{\Delta} + D_{\Delta} \\
G_f(z) &= C(\delta l - A + \Delta C)^{-1}B_f + D_f \\
B_{\Delta} &= (\Delta A - K\Delta C)\Delta B \quad B_f = -KD_{\omega})
\end{aligned}
\]
\[ \begin{align*}
B_f &= (B_a \quad B_c \quad -KD_3) \\
\hat{D}_a &= (\Delta C \quad 0 \quad 0 \quad D_\omega) \\
\hat{D}_f &= (0 \quad 0 \quad D_\omega)
\end{align*} \]

\[ \begin{align*}
\hat{a}(z) &= (x^T(z) \quad v^T(z) \quad d^T(z) \quad \omega^T(z))^T \\
\hat{f}(z) &= (f_a^T(z) \quad f_c^T(z) \quad f_\omega^T(z))^T.
\end{align*} \]

It is indicated from (3) that the residual signal is subjected to both fault signals and disturbance signals (including modelling errors, process disturbances and measurement noises). In order that the residual signal is sensitive to faults, but robustness against disturbances, the observer gain can be designed by solving an optimization problem over a specific frequency range:

\[
\text{minimize} \left( \frac{\|G_a(z)\|}{\|G_f(z)\|} \right)
\]

In order to solve (4), the parametric eigenstructure assignment approach for fault diagnosis was initialized by [42] and further revisited in [43], in which the observer gain \( K \) is formulated as the function of the eigenvalues and eigenvectors, therefore seeking an optimal \( K \) is transformed to the problem of finding optimal eigenvalues and eigenvectors. Recently, eigenstructure assignment based fault diagnosis approaches have been applied to vehicles [44], gas turbine engines [45], spacecraft [46] and wind turbine systems [47]. Alternatively, the multi-object optimization problem described by (4) can be reformulated by linear matrix inequality (LMI), which has been a popular method for fault diagnosis research and applications owing to its wide applicability to a variety of dynamic systems. Recent development of the LMI-based fault diagnosis can be found for various systems such as Lipschitz nonlinear systems [48], TS fuzzy nonlinear systems [49, 50], time-delay systems [51], switching systems [52], and application to structure damage detection [53], and shaft crack detection [54] etc.

A bank of observer-based residuals is generally required in order to realize fault isolation. A nature idea is to make a single residual is sensitive to the fault concerned but robustness against other faults, disturbances and modelling errors, which is called structure residual fault isolation [4]. Alternative fault isolation logic is to make each residual signal sensitive to all but one fault, and robustness against modelling errors and disturbances, which is called generalized residual fault isolation [5]. Recent results on robust fault isolation are developed for nonlinear systems [55,56], and various applications such as for aircraft engine [57], robotic manipulators [58], and lithium-ion batteries [59]. The unknown input observer, proposed by [60], is another fault isolation tool by decoupling input disturbance, modelling errors and other faults in the corresponding residuals. Recently, the unknown input observer based fault isolation techniques are extended to nonlinear systems [61, 62] and applied to aircraft systems [63], inductor motors [64] and waste water treatment plant [65].

Fault identification (or called fault reconstruction/fault estimation) is to determine the type, size and shape of the fault concerned, which is vital information for fault tolerant operation. Advanced observer techniques such as proportional and integral (PI) observers [66, 67], proportional multiple-integral (PMI) observers [68-70], adaptive observer [71-73], sliding mode observers [74, 75], and descriptor observers [76, 77] are usually utilized for fault estimation/reconstruction. The essence of the advanced observers is to construct an augmented system by introducing the concerned fault as an additional state and the extended state vector is thereafter estimated, leading to the estimates of the concerned fault signal together with original system states. Therefore, the advanced observers are also called simultaneous state and fault observers. The above advanced observer techniques are in an advantage position either for reconstructing slow-varying additive faults (PI, and PMI observers), slow-varying parameter faults (adaptive observers), actuator faults with sinusoidal waveforms (sliding mode observers), and high-frequency sensor faults (descriptor system approaches). Actually the above observer techniques may be integrated or combined in order to solve engineering-oriented problems. For instance in [78], integral observer, sliding observers and adaptive observers are combined to reconstruct sensor faults for satellite control systems. In [79], PI observer and descriptor observer techniques are integrated to estimate parameter faults for an aero engine system.

Another well-known model-based fault diagnosis is parity relation approach, which was developed in the early of 1980s [80, 81]. The parity relation approach is to generate residuals (parity vector) which is employed to check the consistency between the model and process outputs. The parity relation approach can be applied to either time-domain state-space model or frequency-domain input-output model, which is well revisited by the books [40,41,82]. Recently, the parity relation method is extended for fault diagnosis for more complex models such as TS fuzzy nonlinear systems [83] and fuzzy tree models [84], and applied to various industrial systems such as aircraft control surface actuators [85] and electromechanical brake systems [86].

Stable factorization approach is frequency-domain fault diagnosis method, which was initiated in 1987 by [87] and further extended by [88] in 1990. The basic idea is to generate a residual, based on the stable coprime factorization of the transfer function matrix of the monitored system, which is made sensitive to the fault, but robustness against disturbances by selecting an optimal weighting factor. Recent developments of stable factorization approach can be found in [89] for nonlinear systems, and [90, 91] for applications in auto-balancing two-wheeled cart and thermal process, respectively.

It is worthy to point out that the parity relation method and stable factorial approach both have some kind of connections with observers. For instance, the parity relation approach is equivalent to the use of a dead-beat observer, and the coprime factorization realization includes the design of observer gain (together with state-feedback gain).

B. Stochastic Fault diagnosis methods

In parallel with the development of the fault diagnosis for deterministic systems, stochastic approaches were also developed for fault diagnosis in the early 1970s. A general
fault detection and diagnosis procedure was first proposed in [92] by using residuals (or innovations) generated by Kalman filters with similar structure to observers, where the faults were diagnosed by statistic testing on whiteness, mean and covariance of the residuals. A variety of statistical tools, such as generalized likelihoods [93], $\chi^2$ testing [94], cumulative sum algorithms [95] and multiple hypothesis test [96], were further developed for testing Kalman-filter based residuals in order to check the likelihood that a particular fault occurs. Further researches have led to a couple of modified Kalman filter techniques for fault diagnosis, such as extended Kalman filters, unscented Kalman filters, adaptive Kalman filters, and augmented state Kalman filters. Unlike the conventional Kalman filters, the extended Kalman filter (EKF) can be used to diagnose the faults in a nonlinear industrial process [97]. The unscented Kalman filter (UKF), depending on a more accurate stochastic approximation, i.e., unscented transform, can better capture the true mean and covariance leading to better diagnosis performance [98, 99]. Adaptive Kalman filters can be employed to tune process noise covariance matrix, or measurement noise covariance matrix in order to obtain satisfactory fault diagnosis [100,101]. The augmented state Kalman filters can be utilized to simultaneously estimate system states and fault signals [102]. Recent application examples of Kalman filter-based fault diagnosis can be found in [103-105] respectively for combustion engines, electronic systems under mechanical shock, and permanent-magnet synchronous motors.

Another important stochastic fault diagnosis method is parameter estimation on the basis of system identification techniques (e.g., least-square error and its derived methods), which was initialized by [106]. In this approach, the faults are assumed to be reflected in system parameters, and only the model structure is needed to be known. The basic idea of the detection method is to identify the parameters of the actual process on-line which are compared with the reference parameters obtained initially under healthy conditions. The parameter estimation based fault diagnosis methods are very straightforward if the model parameters have an explicit mapping with the physical coefficients. This method was well reviewed in the early survey papers [3,9] and book [107]. Recent development of this approach can be found in [108-110].

Motivated by combustion processes, paper making systems and chemical processes, the monitored system outputs can be described by probability density functions. For this class of stochastic systems, fault diagnosis was first addressed by [111], where the probability density function outputs are approximated by using B-spline expansion techniques, and random noises/ errors can be non-Gaussian. In order to improve robustness against measurement noises, modelling errors and process disturbances, an integration of descriptor estimator and parametric eigenstructure assignment was utilized to detect faults in [112]. Recent development of fault diagnosis for nonlinear systems can be found in [113]. In addition, fault diagnosis methods were also developed for other classes of stochastic systems such as stochastic processes with Brownian motions [114, 115] and Markovian jumps [116].

C. Fault diagnosis for discrete-events and hybrid systems

In industrial processes, the signals of some dynamic systems switch from one value to another rather than changing their values continuously. This kind of systems is called discrete-event systems. Fault diagnosis of discrete-event systems was initialized by [117] in 1990s, and the underlying theory of fault diagnosis for discrete-event systems was proposed. The basic event-driven fault diagnosis problem is to perform model-based inferring at run-time to determine whether a given unobservable fault event has occurred or not in the past by using sequences of observable events [118]. According to the model used, the fault diagnosis methods for discrete-event system can be roughly classified into automata based method and Petri net based method. In order to overcome the complexity of the task, the automata-based fault diagnosis method has evolved into decentralized method [119], symbolic method [120] and the combination of decentralized and symbolic methods [121]. On the other hand, Petri net has intrinsically distributed nature where the notions of state and action are local, which has been an asset to reduce the computational complexity of solving fault diagnosis problems [122]. Nevertheless, improved results were developed for avoiding complexity by either applying integer linear programming to Petri nets [123] or using partially observed Petri nets [124]. Recently, event-based approaches were applied to fault isolations for continuous dynamic processes where a high-level discrete-event system fault diagnosers was employed to improve the robustness of the fault diagnosis against large environment disturbances [125] or isolate abrupt parameter faults [126].

Some complex industrial systems are both driven by time-based continuous dynamics and event-driven discrete dynamics, which are called hybrid systems, emerging from complex mechatronic systems, manufacturing systems, complex chemical processes, aerospace engineering systems, automotive engine control and embedded control systems. Monitoring and fault diagnosis for hybrid systems entails challenges due to the fact that the continuous dynamics and discrete event changes are mutually dependent and interacted. Hybrid automata are the most common models to represent hybrid systems, which can be utilized to design fault diagnosis algorithms to detect and isolate faults [127, 128]. Bond graph has become a powerful model to be used for fault diagnosis due to its capability of modelling complex systems in a unified way, and the ease for obtaining analytical redundancy relation from the causalities on the graph. Recent results on Bond graph based fault diagnosis and their applications for hybrid systems can be found in [129-132].

D. Fault diagnosis for networked and distributed systems

The rapid developments in network technologies have much stimulated the real-time control and monitoring via communication channels, that is called networked control and
monitoring, which have valuable advantages such as cost effectiveness, less weight and power requirements, easier for installation and maintenance as well as resources sharing [133]. It is noted that the introduction of limited-capacity network cables or wireless sensors into control and monitoring loops has unavoidably brought some unanticipated problems such as random communication delays, data dropout, and scheduling confusion, which make the network based monitoring and fault diagnosis more challenging compared with conventional point-to-point control and monitoring systems. Therefore, in the network based fault diagnosis, the residual or fault estimation error should be not only robust against modelling errors, process disturbance, and measurement noises, but also robust against transmission delays, data dropouts, and incomplete measurements caused by the limit capacity of communication channels [134]. Recently, a variety of fault diagnosis techniques have been developed for various networked systems. For instance, fault detection filters were developed in [135] for systems subjected to communication delays and missing data, where the network status is assumed to vary in a Markovian fashion. In [136], least-square filters and Kalman filters were integrated for fault detection, isolation and estimation for network sensing systems. In [137], a networked based fault diagnosis technique was addressed for nonlinear systems, which was assessed by an experimental test system with the use of IEEE 802.15.4 wireless sensor networks. In addition, it is also of interest to detect the anomaly of a communication network itself, which may affect the performance of network control systems. In order to monitor the guaranteed quality of service (QoS) of the router and the whole topology, sliding mode observer techniques were employed in [138, 139] for anomaly detection in the transmission control protocol (TCP). Very recently, model-based detection and monitoring approaches were addressed in [140, 141] for monitoring potential intermittent connections or faulty nodes for controller area networks (CANs).

Complex industrial systems can be modelled as an interconnection of subsystems, and each subsystem has a decision marker (intelligent agent) which might have access to local measurements, subsystem models, local estimators/controllers, and constrained communication channels between the agent and its neighbors [142]. This kind of decentralized or distributed structure has become the main stream in complex industrial processes owing to its less use of network resources, cost effectiveness and convenience for expansion. On the other hand, real-time monitoring and fault diagnosis for distributed systems is much challenging due to the constrained information redundancy and limited communication capacity. A general idea of the distributed fault diagnosis is to design local estimators or fault detection filters by intelligent agents according to the local sensing and computing resources, and a consensus strategy is utilized to ensure the whole detection or estimation performance of all the agents in the network. Recent developments on distributed fault detection [143] and distributed fault estimation [144] are developed respectively by using adaptive thresholds and sliding modes techniques to improve the robustness against noises and modelling errors. Moreover, applications on unmanned airships [145] and power networks [146] are reported as well.

III. SIGNAL-BASED FAULT DIAGNOSIS METHODS

Signal-based methods utilize measured signals rather than explicit input-output models for fault diagnosis. The faults in the process are reflected in the measured signals, whose features are extracted and a diagnostic decision is then made based on the symptom analysis and prior knowledge on the symptoms of the healthy systems. Signal based fault diagnosis methods have a widely application in real-time monitoring and diagnosis for induction motors, power converters and mechanical components in a system. A schematic diagram of signal-based fault diagnosis is depicted by Fig. 4.

![Fig. 4. Scheme of signal-based fault diagnosis](image)

The feature signals to be extracted for symptom (or pattern) analysis can be either time-domain (e.g., mean, trends, standard deviation, phases, slope, and magnitudes such as peak and root mean square) or frequency-domain (e.g., spectrum). Therefore, signal-based fault diagnosis methods can be thus classified into time-domain signal based approach, frequency-domain signal based approach and time-frequency signal based method.

A. Time-domain signal based methods

For a continuous dynamical process to be monitored, it is natural to extract time-domain features for fault diagnosis. For instance in [147], by analyzing the changes of the measured root-mean-square current characteristics between healthy conditions and the situations under single/dual transistor short circuit or open circuit, a fault diagnosis method was developed for power converters of switched reluctance motors. In [148], the absolute value of the derivative of the Park’s vector phase angle was used as a fault indicator, which was employed for diagnosing multiple open-circuit faults in two converters of permanent magnet synchronous generators (PMSG) drives for wind turbine applications. By observing the slope of the induction current over time, a fault diagnosis method was addressed in [149] for open and short circuits switch fault diagnosis in non-isolated DC-DC converters, and the field programmable gate array (FPGA) digital target was then used for real-time experimental implementation. In [150], by using the measured motor phase currents and their corresponding
reference signals, a real-time algorithm was developed for detecting and locating multiple power switch open circuit faults in inverted-fed AC motor drives. In [151], it was shown that, under balanced supply voltage, the phase angle, the magnitude of the negative and zero-sequence currents can be considered as reliable indicators of stator faults in the induction motors. In [152], a statistical method for the detection of sensor abrupt faults in aircraft control systems was presented, where the covariance of the sensing signals was used for feature extraction. Recently, a time-domain signal-based diagnostic algorithm was developed in [153] for monitoring of gear faults, by combining fast dynamic time warping (Fast DTW) as well as correlated kurtosis (CK) techniques. The fast DTW algorithm was employed to extract the periodic impulse excitations caused from the faulty gear tooth, and the extracted signal was then resampled for subsequent diagnostic analysis using the CK technique. Taking advantages of the periodicity of the geared faults, the CK algorithm can identify the position of the local gear fault in the gearbox.

Different from the approaches for fault detection and diagnosis using features of the measured signal in one-dimension domain, a two-dimension signal-based method was proposed in [154], where the vibration signal was translated into an image (two dimensions), and the local features were then extracted from the image using scale invariant feature transform (SIFT) for fault detection and isolation under a pattern classification framework. Very recently, a two-dimension approach was reported in [155] for fault diagnosis of induction motors, where time-domain vibration signals acquired from the operating motor were firstly converted into two-dimension gray-scale images, and the discriminating texture features were then extracted from these images utilizing local binary patterns (LBP) technique. The extracted texture features were finally used for fault diagnosis with the aid of a classifier. It is noted that, when converting signals into images, the added noise acts as illumination variation. As both the SIFT technique and the LBP operator have illumination invariance capability to some extent, the proposed fault diagnosis methods in [154, 155] have robustness even in a high level of background noises.

B. Frequency-domain signal based methods

Frequency-domain signal based method is to detect changes or faults by using spectrum analysis tool such as discrete Fourier transformation (DFT). One of the most powerful frequency-domain methods for diagnosing motor faults is motor-current signature analysis (MCSA), which utilizes the spectral analysis of the stator current to sense rotor faults associated with broken rotor bars and mechanical balance. Without requiring access to the motor, the MCSA approach has received much attention, which was well reviewed in [19, 20]. Recent development of current based spectrum signature analysis for fault diagnosis can be found in [156, 157].

Vibration signal analysis is a common method for condition monitoring and diagnosis for mechanical equipment such as gear box, as machine sound indicates a lot about working condition of the machine. In [158], an acoustic fault detection method was addressed for gear box on the basis of the improved frequency domain blind de-convolution flow. Recently in [159], Fourier spectrum and the demodulated spectra of amplitude envelope were employed to detect and locate multiple gear faults in planetary gearboxes.

C. Time-Frequency signal based methods

For machines under an unloaded condition, or unbalanced supply voltages, varying load, or load torque oscillations, the measured signals are generally transient and dynamic under the concerned time section. Therefore, analysis of the stationary quantities in some cases finds difficult to monitor or detect faults via either a pure time-domain or frequency-domain method. Due to the time-varying frequency spectrum of the transient signals, suitable time-frequency decomposition tools are needed for real-time monitoring and fault diagnosis. Time-frequency analysis can identify the signal frequency components, and reveal their time variant features, which has been an effective tool for monitoring and fault diagnosis by extracting feature information contained in non-stationary signals [25].

Various time-frequency analysis methods have been proposed and applied to machinery fault diagnosis. Among the time-frequency methods, short-time Fourier transform (STFT), wavelet transforms (WT), Hilbert-Huang transform (HHT), and Wigner-Ville distribution (WVD) are most common used approaches. For instance, STFT method allows determining signal frequency contents of local sections as the signal changes in time, which has been widely applied to detect both stator and rotor faults in inductor motors [160]. However, the STFT method suffers the high computational cost if it is required to obtain a good resolution. As a linear decomposition, WT based method can provide a good resolution in time for high-frequency components of a signal and a good resolution in frequency for low-frequency components, which has demonstrated the effectiveness for tracking fault frequency components among some cases find difficult to monitor or detect faults via either a pure time-domain or frequency-domain method. For instance, the selection of a suitable window size in STFT is required, but it is generally not a known priori. The type of the basic wavelet function in WT has a direct effect on the effectiveness in identifying transient elements hidden within a dynamic signal. However, on the basis of the instantaneous frequencies resulting from the intrinsic-mode functions of the signal being analyzed, HHT method is not constrained by the uncertain limitations with respect to the time and frequency resolutions suffered by some time-frequency techniques (e.g., STFT and WT), which has shown quite interesting performance in terms of fault severity evaluation [163]. WVD method features a relatively low computational cost and high resolution, as the entire signal is utilized to obtain the energy at each time-frequency bin, which has been successfully applied to the fault diagnosis along with current analysis [164] or vibration analysis [165]. A significant defect of the
conventional WVD method is the appearance of the cross terms in the distribution of artifacts, which hinders the application of WVD methods. Very recently, via combining advanced notch FIR filters and the conventional WVD method, an improved WVD based fault diagnosis algorithm was proposed in [166], which can effectively minimize the cross terms and provide seamless high-resolution time-frequency diagrams enabling the diagnosis of rotor asymmetries and eccentricities in induction machines directly connected to the grid even in the worst cases. In [167], a self-adaptive WVD method, based on local mean decomposition, was addressed, which can evidently remove the cross-terms of WVD to improve the performance of the defect diagnosis.

IV. CONCLUSION

In the first-part survey paper, fault diagnosis techniques and their applications have been comprehensively reviewed from model-based and signal-based perspectives, respectively. Specifically, model-based fault diagnosis is reviewed following the categories of fault diagnosis approaches for deterministic systems, stochastic fault diagnosis methods, discrete-event and hybrid system diagnosis approaches, and networked and distributed system diagnosis techniques, respectively. Meanwhile, signal-based fault diagnosis is surveyed following the classifications of time-domain, frequency-domain, and time-frequency-domain approaches, respectively. The overview on knowledge-based fault diagnosis, hybrid and active fault diagnosis is to be carried out in the second-part review paper, which will complete the whole overview on the fault diagnosis techniques and their applications.

REFERENCES


Zhiwei Gao (SM’08) received the B.Eng. degree in electric engineering and automation and M.Eng. and Ph.D. degrees in systems engineering from Tianjin University, Tianjin, China, in 1987, 1993, and 1996, respectively. From 1987 to 1990, he was with Tianjin Electric Drive and Design Institute as an Assistant Engineer. From 1996 to 1998, he was with the Department of Mathematics, Nankai University, as a Postdoctoral Researcher. In 1998, he joined the School of Electric Engineering and Automation and received a professorship in control science and engineering in 2001. Before joining the Faculty of Engineering and Environment at the University of Northumbria in 2011, he held lecturing and research positions with the City University of Hong Kong, University of Manchester Institute of Science and Technology, University of Duisburg-Essen, University of Manchester, University of Leicester, University of Liverpool, and Newcastle University. His research interests include data-driven modelling, estimation and filtering, fault diagnosis, fault-tolerant control, intelligent optimisation, large-scale systems, singular systems, distribution estimation and control, renewable energy systems, power electronics and electrical vehicles, bioinformatics and healthcare systems.

Dr. Gao is presently the associate editor of the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, and IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY.

Carlo Cecati (M’90-SM’03-F’06) is a Professor of industrial electronics and drives and the coordinator of the Ph.D. course on renewable energy and sustainable building at DISIM-University of L’Aquila, L’Aquila, Italy. He is also Chief International Academic Adviser at Harbin Institute of Technology, Harbin, China. His research and technical interests cover several aspects of power electronics, distributed generation, and smart grids. In these areas he has published more than 130 journal and conference papers. Since nineties, he has been an active member of IEEE-IES; currently he is a Senior AdCom Member and the Editor in Chief of the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS. Prof. Cecati has been a co-recipient of the 2012 and of the 2013 Best Paper Awards from the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS and of the 2012 Best Paper Award from the IEEE INDUSTRIAL ELECTRONICS MAGAZINE.

Steven X. Ding received the Ph.D. degree in electrical engineering from the Gerhard-Mercator University of Duisburg, Germany, in 1992. From 1992 to 1994, he was an R&D Engineer at Rheinmetall GmbH. From 1995 to 2001, he was a Professor of control engineering at the University of Applied Science Lausitz in Senftenberg, Germany, and served as Vice President of this university during 1998–2000. Since 2001, he has been a Professor of control engineering and the head of the Institute for Automatic Control and Complex Systems (AKS) at the University of Duisburg-Essen, Germany. His research interests are model-based and data-driven fault diagnosis, fault-tolerant systems and their application in industry with a focus on automotive systems, mechatronic and chemical processes.