

Contextualized Early Failure Characterization of Cantilever Snap Assemblies.

Juan Rojas
School of Software
Sun Yat Sen University
Guangzhou, Guangdong, 510006, China

Kensuke Harada, Hiromu Onda, Natsuki
Yamanobe, Eiichi Yoshida, and Kazuyuki
Nagata.
Intelligent Sys. Research Institute, AIST
Tsukuba, Ibaraki, 305-8568, Japan

Abstract—Failure detection and correction is essential in robust systems. In robotics, failure detection has focused on traditional parts assembly, tool breakage, and threaded fastener assembly. However, not much work has focused on classifying failure into various sub-modes. This is an important step in order to provide accurate failure recovery. Our work implemented a contextualized failure characterization scheme for cantilever snap assemblies. A rule based approach was used through which assemblies whose trajectories deviated from the normal approach trajectory were identified in the beginning of the task. We not only identified failure but also the failure type that occurred. The method identified exemplars that characterized salient features for specific deviations from the initial approach trajectory in the assembly task. A contact-state map was generated through sampling the contact space during training. Contextualized statistical measures were used to classify trials during the testing phase. Our work classified failure deviations with 88% accuracy. According to the statistic measures used, varying success was experienced in correlating failure deviation modes. Each case was analyzed using gaussian statistics and one and two standard deviations. Cases with trajectory deviations in one direction had {75%, 92%} accuracy, cases with deviations in two directions had {61%, 94%} accuracy, and cases with deviations in three directions had {69%, 100%} accuracy. Our work provides further insights into the early failure characterization of complex geometrical parts which will serve to implement failure recovery techniques in the face of significant and unexpected errors.

I. INTRODUCTION

Failure detection and correction are essential to produce flexible, adaptable, efficient and robust autonomous robotic systems. Failure detection early on focused on detecting abrupt changes [14]. Numerous filters and rule-based methods emerged. Overtime, statistical and machine learning methods were implemented to detect failure and failure modes [10], [11], [4]. Failure detection methods can be divided into model or model-free approaches. The former use theoretical system designs to identify failure, while the latter accumulate experimental data for both outcomes. Model-free approaches can be computationally expensive and damaging to robot equipment given that failure experiments need to be carried out. Recent work has focused on implementing a technique that requires relatively few trials to learn failure classification [10].

In robotics, failure detection has traditionally focused in parts assembly [5], tool breakage [3], [6], and threaded fastener assembly [1]. Recent work has used support vector machines and principal component analysis to classify

successful or failed assemblies upon completion [10], [3]. A similar work tried to continuously detect the task and identify early failure by combining relevance vector machines with a Markov chain model [11]. Less work has focused on the harder problem of failure mode characterization. In this work, we define failure characterization (FC) as, not only detecting failure, but what kind of failure (*e.g.* failure mode) occurred. In [4], a Bayesian nonparametric time series was used to monitor fault detection. While the work stated that it could identify failure mode detection the authors did not execute such experiments.

Our work will perform FC on cantilever snap parts as they are amply used but also have complex geometries and elastic snaps that lead to intricate force signatures. The complexity is such that assembly planning methods like contact-state graphs [9] have yet to be used in this context. In general, failure characterization schemes have been built on top of state estimation techniques. Some recent work has dealt with state estimation of intricate assemblies. In [12], a non-parametric system is used to encode FT signatures based on relative-change according to an increasingly abstract taxonomy and Bayesian methods to reason about the task. In [4], a sticky-Dirichlet Process Hidden Markov Model performed continuous estimation of humanlabeled nominal executions and identified whenever the trajectory deviated from the routine trajectory. With respect to FC, [13] studied cases in which the trajectory of an assembly task deviated from a nominal trajectory in 1 of 3 directions. The study focused on the initial approach trajectory that leads to the first contact between mating parts. They correlated salient FT features with trajectory deviations in a single direction (from here on referred to as “deviations” and should not be confused with statistical standard deviations) and studied if features for individual deviations could be identified when multiple deviations were linearly superposed simultaneously. They also performed failure-mode correlation. That is, after classifying failure modes in multiple directions, they correlated the data to validate if those deviations actually took place. That work performed FC with modest success. The accuracy of failure and failure modes detection diminished with increased number of deviation directions.

Our current work, seeks to build a contact-state *map* to identify failure modes. Such a map is generated by populating sample points with FC exemplars. The sample

points represent contact points when deviations in 1-3 directions are inserted in the nominal approach trajectory. The exemplars in this work, as in [13] hold salient force-moment data. However, we do not wish to linearly superpose the data, instead we try to group exemplars based on Gaussian properties of mean and standard deviation. For this reason, our exemplars first consider deviation directions but then they also consider the number of simultaneous deviations that are enacted (1-3 are analyzed) as this fact significantly changes the possible groupings of exemplars by statistical parameters. Finally, within a limited contact area we sample the space and characterize failure modes through further statistical analysis of the exemplars.

Our FC method extracts key FT information through the Relative Change-Based Hierarchical Taxonomy (RCBHT) [12] to produce exemplars. The RCBHT yields four abstraction levels of increasing intuitiveness and one additional layer for state reasoning. Each layer consists of labels and quantitative data from which salient features are detected. After system exemplars were identified, off-line training was used to average exemplar values and upper and lower boundaries at different contact points with different deviations. These measures helped identify: (i) failure detection and (ii) failure mode identification. Furthermore, we conducted a correlation study to verify the reliability of the data.

Training consisted of 114 trials in which deviations were applied in three separate directions and combinations of these three in world coordinates. The testing phase used 36 new trials. Testing results showed that our scheme detected failure with 88% accuracy and failure modes with an average of 76% accuracy. Depending on the number of standard deviations chosen for boundary values, failure modes could be correlated between 68% of the time to 95% of the time. We discovered that the contextual identification of exemplars based on constrained trajectory motion can help to identify failure and failure modes, although with some limitations. Our work provides a simple and flexible approach for failure characterization of cantilever snap assemblies.

The rest of the paper is organized as follows: the experimental set-up is presented in Sec. II, the RCBHT system under which FC is implemented is described in Sec. III, the FC actual scheme is presented in Sec. IV, Experiments and Results are detailed in Sec. V, the discussion is present in Sec. VI, and the conclusion to close in Sec. VII.

II. EXPERIMENTAL SETUP

NX-HIRO, a 6 DoF dual-arm anthropomorph robot was simulated in the OpenHRP 3.0 environment [7]. Male and female 4-snap cantilever camera parts were used. The male part was rigidly mounted on the robot's wrist, while the female snap was rigidly fixed to the ground as in Fig. I. For this work we consider the Pivot Approach strategy [12] that, for a successful assembly, completes four states: the Approach, Rotation, Insertion, and Mating states. The Approach state drives the male part along a smooth trajectory until it contacts the female part at an angle at the docking

pivot as in Fig. I. Our FC scheme inserts deviations from the nominal trajectory in the Approach state to generate and analyze failure signals early in the assembly task. FC analysis starts during the Approach state and may continue into the Rotation state. The small deviations are characteristic of what a human adult would make when trying to enact a snap assembly but narrowly misses the mark.

III. THE RELATIVE CHANGE-BASED HIERARCHICAL TAXONOMY

The RCBHT yields state representations by hierarchically abstracting snap assembly force-torque (FT) data in increasingly intuitive ways. The latter is composed of four increasingly abstract layers that encode relative-change in the task's force signatures and one monitoring reasoning layer. The system's taxonomy is built on the premise that relative-change patterns can be classified through a small set of categoric labels and aided by contextual information. The RCBHT analyzes FT signatures from all force axes independently and contextualizes the state according to automata state participation. Previously, the RCBHT's monitor only discerned a task's success or failure at the end of the task, but for FC it has been extended to monitor from the Approach state. In this section, the RCBHT first two layers are explained, while the last two are omitted since they are not used or needed in this work (for more details on the taxonomy see [12]). The FC method introduced in Sec. IV runs as a state-monitor on top of the RCBHT.

A. Primitive Layer

The first layer, the Primitives, partitions FT data into linear data segments and classifies them according to gradient magnitude. Linear regression along with a correlation measure are used to segment data when a minimum correlation threshold is flagged. Gradient classification has three main subgroups: positive, negative, and constant gradients. Both the positive and negative sets are subdivided into 4 ranges: small, medium, large, and very large. Contact phenomena is characterized by abrupt changes in force signals almost approximating an (positive or negative) impulse, it is such impulses that are characteristic of snap motions. Negative and positive gradient ranges are labelled as: "sneg, mneg, bneg, nimp" and "spos, mpos, bpos, and pimp" respectively. Constant gradients are those whose change is trivial. They are classified as such if their gradient magnitude is lower than the absolute value of a calibrated threshold and are labeled as "const". In order to generalize the parameter thresholds for this layer, an optimization routine devised in [12] uses contextual information derived from the assembly strategy to determine where the largest gradients and the near-constant gradients are located according to task state and axis. For details see [12]. Filtering is executed for this and every layer of the taxonomy. Filtering merges extracted behaviors based on three different criteria: (i) a time-duration context, (ii) repeated behaviors, and an (iii) amplitude value context. The time-duration context seeks to eliminate behaviors whose duration is so small that they become negligible. The repeated

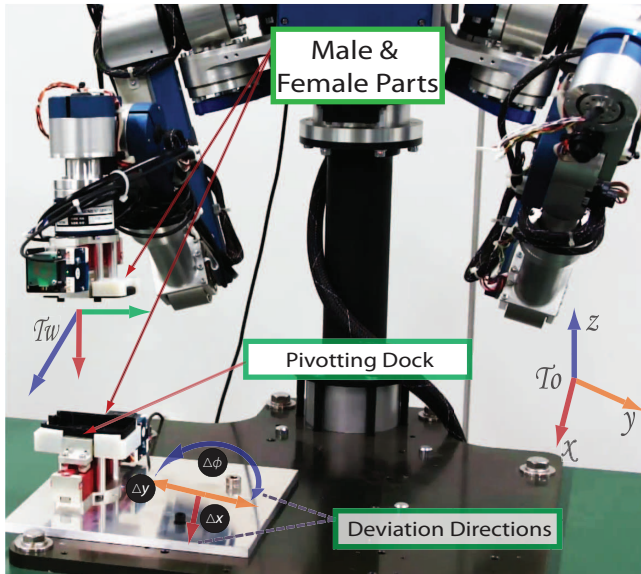


Fig. 1. The NX-HIRO dual-armed humanoid robot assembles male-female snap parts.

behavior principles merge repeated behaviors and clean noisy signals. The amplitude criteria examines adjacent signals magnitudes. If one is much larger than the other (except for impulse signals), the lower amplitude signal is merged with the larger one. Each layer runs a filtering cycle 2-3 times to reduce the label number to a most representative number. There are other details concerning filtering which are not necessary for this discussion but can be found in [12].

B. Composites Layer

The next layer analyzes ordered-pair Primitives sequences to create “motion compositions” (MC’s). By studying patterns in the ordered-pairs, seven sets of higher-level abstractions are extracted. These actions represent force-torque behaviors labeled as: adjustments, ‘a’, contacts ‘c’, increases, ‘i’, decreases, ‘d’, constants, ‘k’, and unstable motions, ‘u’. Adjustments are motions where a positive/negative gradient is followed by a negative/positive gradient respectively. They represent a small rattle motion between male and female snap parts. Additionally, if two positive or negative gradients succeed each other, it points to actions in which the force or torque is increasing or decreasing respectively. With respect to our previous statements, we treat impulse gradients (pimp, or nimp) independently. We acknowledge that when a positive impulse is followed by a negative one or vice-versa a contact action between male and female parts is likely. Additionally, if two positive or negative impulses succeed each other, it may lead to unstable behavior. Besides the assigned label, each MC possess quantitative data such as: average magnitude values, maximum signal values, average amplitude values, and starting and ending times for each of the primitives

Filtering is executed after the motion compositions creation. Filtering merges motion compositions based on three different criteria: (i) a time-duration context, (ii) repeated

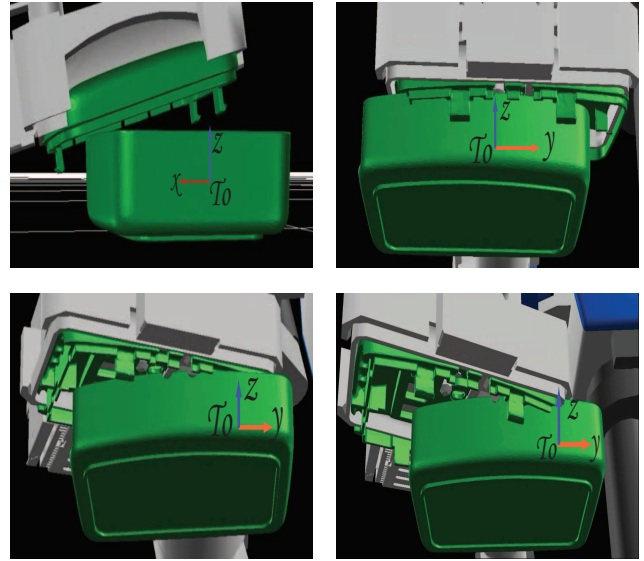


Fig. 2. Examples of Deviations in Nominal Approach Trajectories in the x-dir, y-dir, Yaw-dir, and xyYaw-dir.

MCs, and an (iii) amplitude value context. Filtering seeks to maintain meaningful representations while filtering noisy compositions.

The MC layer is key to the FC exemplar identification. The MC’s quantitative data is general enough that we do not tamper with noisy signal dynamics, but not abstract enough that we lose salient features’ detail. All of our exemplars consist of either key motion composition average magnitudes or average amplitude values.

IV. FAILURE CHARACTERIZATION SCHEME

Our FC method is a novel model-free rule based approach that uses the RCBHT. The latter, facilitates the identification of failure modes by look at the distortion of MCs in specific FT axes, within a given automata state. In particular, we can compute MC average magnitudes or amplitudes and durations, all of which encode characteristics of either successful or failed tasks. Before presenting the FC scheme, a set of fundamental principles for FC is presented.

A. Failure Characterization Requirements

FC systems consist of at least three components [8]:

- (a) General Failure Identification: the system identifies if failure is present. The system should be robust against falsepositive situations.
- (b) Failure Mode Identification: the system also identifies what aspect of the task failed.
- (c) Diagnostics: the system provides quantitative data (when, how much?) about failure.

This information can then be used by a failure recovery mode and attempt to return the task’s state to normal operation.

B. Failure Sources

Current state-of-the-art robots still introduce error into tasks. Manipulator failure sources include: upper bounds on

payloads and speeds; absolute position accuracy (depends on temperature, load, speed, and workspace position), and repeatability error. Similarly, mechanical parts introduce error due to differences in design within the permitted tolerance range and uncertainty in pose of both robot effectors and parts [2].

C. Failure Scheme

The FC Scheme is divided into three stages: (i) exemplar identification; (ii) contact-state map and statistical parameter generation, and (iii) testing. In this work, six exemplars $f_j \in F$ were designed to discern up to three *simultaneous* deviations $\delta_i \in D$. Mean \bar{x} and standard deviations σ of FT data are encoded as exemplars are derived to identify success or failure modes. During training, a contact-state map is generated by pairing trajectory deviations $\prod_{i=1}^m \Delta_i$ with exemplar values $\prod_{j=1}^n F_j$, such that there is a correspondence between possible deviation combinations and corresponding exemplars as stipulated in Fig. 3 leading to : $\prod_{i=1}^m \Delta_i \Leftrightarrow \prod_{j=1}^n F_j$. Deviations are limited (for simplification) up to three directions. They can be enacted individually or simultaneously. By testing individual deviations first and simultaneous deviations later, a map of contact points can be generated in which a set of deviations that lead to contact points can be paired with a corresponding set of exemplars. In effect, we are sampling a contact-space. In this way, failure direction and magnitude can be indirectly represented. Both of these data are important to effect precise failure correction methods to the assembly task.

1) *Exemplar Identification*: Exemplar identification assumes a constrained assembly that allows to generate similar-patterned signals across trials and facilitate encoding. For the case of early FC, we introduce limited trajectory deviations in the Approach state of the assembly task in different directions and magnitudes. Deviations are introduced in the: x-direction (xDir), y-direction (yDir), and rotation about the z-direction (YawDir): $\langle +x, \pm y \text{ and } \pm \phi \rangle$. Noise is inserted within limited bounds and in 1-, 2-, or 3-simultaneous directions. Compared to success signals, failure signals reveal clear differences across FT signal patterns. Selected exemplars can be used to discern two aspects of the failed assembly: (i) the present deviation magnitude, and (ii) whether the deviations were introduced in 1-, 2-, or 3-simultaneous directions.

Exemplars were identified empirically. The RCBHT's MC level was determined to have the best granularity to capture change in salient features. Likewise, a number of criteria was used to identify the presence of salient features: (i) a specific force axis; (ii) a specific PA automata state; (iii) a limited automata state duration, and (iv) the analysis of the MC amplitude or magnitude value within the FT axis-state-duration criteria. Furthermore, we contextualize exemplars by detecting if a narrow band of average FT values can characterize deviations in 1-, 2-, or 3- directions. That is, consider the deviation & exemplar pair: (xDir, MyR). We analyze if, when xDir is mixed with yDir or YawDir, whether or not, the exemplar's mean average value plus up to two standard

Dir	Dev's	Force Axis	PA State	Parameter (Avg Val)	State Duration	Notation
xDir	1:3	My	Rotation	Magnitude	50%	MyR
yDir	1	Mz	Rotation	Amplitude	100%	MzR1
	2:3	"	"	"	"	MzR23
Roll	1	Fx	Approach	Magnitude	-33%	FxA1
	2	"	"	"	"	FxA2
	3	"	"	"	"	FxA3

Fig. 3. Exemplar Identification: Six exemplars were identified for trajectory deviations in three directions. MyR identifies xDir deviations (works for 1 to 3 simultaneous deviations); MzR1 and MzR23 identifies yDir (the former for cases when 1 deviation is enacted, the latter for 2 to 3 simultaneous deviations); FxA identifies YawDir deviations (three exemplars used to identify deviation the three different deviation sub-groups). Note that " represents a repetition of the same data on the row above.

deviations can capture all FT data or not. If yes, one exemplar is used, if not, we analyze if we need 2 or 3 exemplars. The process repeats by considering only 2 deviations. If it works we assign the exemplar to that deviation pair, if not we try the next pair, if not 3 exemplars must be used for the three deviations. This categorization, is referred to as "Deviation sub-groups". The contextualization is more powerful as it identifies the minimum exemplar number that represents salient features in simultaneously enacted deviations. The results are shown in Fig. 3. Exemplars use local wrist-coordinates. A basic explanation is presented below. xDir deviations are characterized by the vertical contact that the male snap makes with the female snap (z -axis in local coordinates). *My* signatures strongly correlate Fz in the first half of the Rotation state. Thus, one can identify xDir deviations by averaging the MC's magnitude values that occur in the first half of the Rotation State of the *My* axis.

yDir deviations have noisy Fy signals are that are hard to interpret, however, a moment about z can easily be discerned as some rotation about x_W occurs when the male and female parts are in contact. One can identify yDir deviations by taking the MC's maximum amplitude signal that occur in the Rotation State of the Fz axis.

YawDir deviations generate force signals in the Fx axis during the Approach stage as opposed to having a near-zero reading when there is no rotation. In particular, the change in force value is apparent towards then end of the Approach stage. One can identify YawDir deviation by averaging the MC's magnitude values that occur in the last third of the Approach state in the Fx axis as shown in Fig. 4.

2) *Training*: After exemplars are identified the training phase computes a couple of values for both successful and unsuccessful tasks. (i) The Mean Averaged Value: it is the mean value of all training example MC magnitudes or maximum MC amplitudes; and (ii) The Upper and Lower Boundaries: created by taking the Mean Averaged value plus or minus 1 or 2 standard deviations respectively.. In effect one has these 3 statistical measures for our six exemplars: MyR, MzR1, MzR23, FzA1, FzA2, and FzA3. By collecting statistical measures for these exemplars, we will be more effective in classifying common values for failure assemblies that have deviations in the xDir, and/or, the yDir, and or the

YawDir. In essence, a type of map is being formulated in which regions with similar values can be classified.

During the training phase, 114 trials were run. Training samples show with red dots in Fig. 5. In the x - and y -directions, deviations range from $\pm 0.0075\text{m}$ to $\pm 0.0105\text{m}$. In the Yaw direction, rotations range from $\pm 0.08725\text{rad}$ to $\pm 0.5235\text{rad}$. Seven trials were run in $(+x\text{Dir}, +y\text{Dir}, -y\text{Dir})$ respectively. Six were run in $(+Yaw, -Yaw)$ respectively. Nine were run in $(+x\text{Dir}, +y\text{Dir})$ and $(+x\text{Dir}, -y\text{Dir})$ respectively. Eighteen were run in $(+x\text{Dir}, \pm Yaw)$ and $(\pm y\text{Dir}, +Yaw)$ directions with equally spaced intervals. Twenty-seven were run in the $(+x\text{Dir}, +y\text{Dir}, +Yaw)$ directions with equally spaced intervals. For success cases, 6 trials were used. During training, three statistical parameters were computed for exemplar's f_j across trials: (i) the Mean Average Value \bar{x} , (ii) the Upper Boundary Value (\bar{x}_{max}), and (iii) the Lower Boundary Value (\bar{x}_{min}). Note that the exemplar for an individual trial already represents an averaged quantity, either of magnitudes or amplitudes, but the Mean Average Value is of exemplars over trials. The Mean Average value computation and the standard deviation for a set of N values are shown in Eqtn. 1.

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad \sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (1)$$

Additionally, if Π is defined as the set of all exemplars in the training phase, then the upper boundary x_{max} for all exemplar averages \bar{x} and the lower boundary x_{min} is defined according to $\bar{x}_{max} = \bar{x} + k\sigma$ and $\bar{x}_{min} = \bar{x} - k\sigma$, where k has a constant value of 1 or 2.

3) *Testing*: During testing, 36 trials were run and shown in blue dots in Fig. 5. Four trials were run in $(+x\text{Dir}, \pm y\text{Dir}, \pm Yaw\text{Dir})$ respectively. Six were run in $(\pm x\text{Dir}, \pm y\text{Dir})$ and

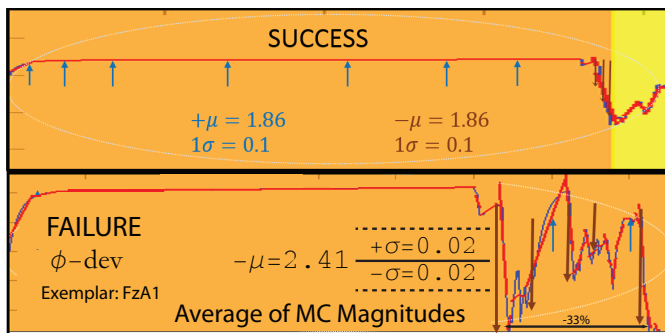


Fig. 4. Failed assembly introduced a Yaw deviation early in the trajectory. The figures shows a section of the Approach stage in the Fx axis. The red line segments represent primitives. Paired primitives form MCs represented here by blue arrows. Brown arrows in the failure case represent MC amplitudes. To compute the exemplar for a Yaw Deviation as a single deviation we look at the last 33% of the Approach stage where there is a large distortion in signals between the success and failure case. The mean and standard deviations of MC amplitudes in this region are computed. Notice the large difference between success and failure cases: 1.86 ± 0.1 for success vs. 2.41 ± 0.02 for failure cases. During training, this exemplar value is averaged with other iterations (we do not count outliers) of the same failed assembly task, resulting in an averaged set of statistical values used during the testing phase of the experiment.

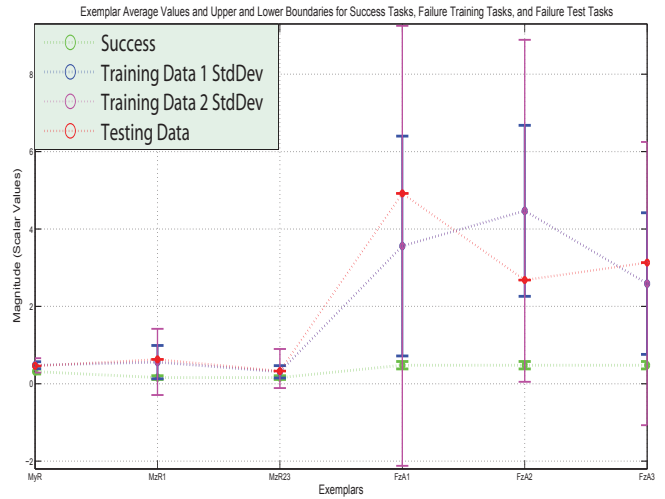


Fig. 6. Exemplar Mean Average Values and upper and lower boundaries for success assemblies in green, training with standard deviation $\sigma=1$ in blue, training with $\sigma = 2$ in magenta, and testing in red.

$(+x\text{Dir}, +Yaw\text{Dir})$, $(\pm y\text{Dir}, +Yaw\text{Dir})$, and $(+x\text{Dir}, +y\text{Dir}, \pm Yaw\text{Dir})$ directions.

Once the statistical measures were generated, the testing of new trials with deviations (in 1, 2, and 3 directions) were executed. Testing was designed to evaluate the system's performance in (i) failure detection for specific deviations, and (ii) failure mode identification. As mentioned in IV-C.2 failure-sub mode identification allows us to classify the failure with more accuracy as exemplar values change when the deviation magnitude increases or when multiple deviations are enacted simultaneously. To illustrate, consider a task in which there is only deviation in the YawDir. In this case the system will compute the mean exemplar values and (i) it will detect if the values fall within the boundaries of the successful task. Bounds for $x\text{Dir}$ and $y\text{Dir}$ should work but not so for YawDir. The system would then (ii) flag the YawDir with failure. Further, (iii) the system would then re-assess if the exemplar value computed falls within any of the six exemplars: MyR, MzR1, MzR23, FzA1, FzA2, and FzA3. If successfully classified, a flag should appear in FzA1.

V. RESULTS AND ANALYSIS

This section presents statistical measure computations during the training and testing phases as well as system's performance in detecting failure and failure modes.

In Fig. 6, note that average values between failure case exemplars and successful ones are markedly different and ratify their use for classification. Also note how the testing Average Mean Values (red color) fall within the boundaries of failure boundaries with $\sigma = 1$ (blue color) and $\sigma = 2$ (magenta color). Also note that the extrema between MzR1 and MzR23 is quite significant in net terms and the fact that the boundaries for FzA's in general are quite wide. Our intuition tells us that the Average Mean Values which are directly caused by contacts generated between the male and female parts vary widely even as we deviate along any

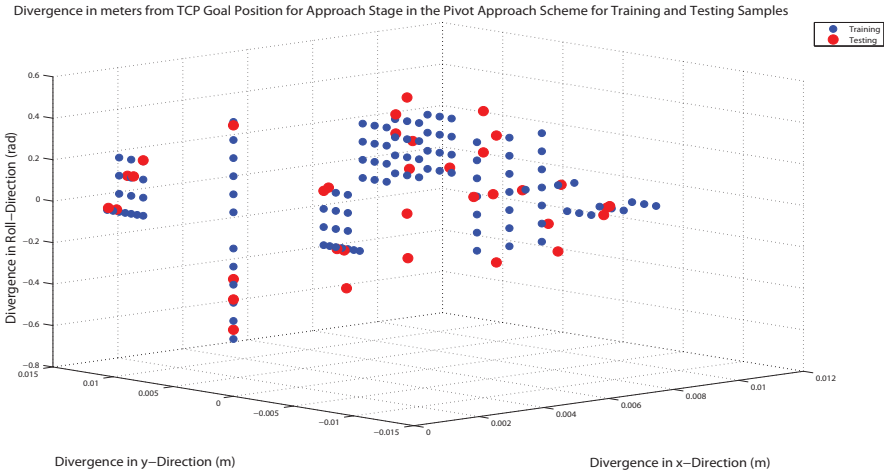


Fig. 5. Training and Testing trials. The scatter plot depicts selected deviation points in 1, 2, or 3 directions. Values for x- and y- directions range from $\pm 0.0075\text{m}$ to $\pm 0.0105\text{m}$. Values for yaw deviations range from 0.08725 to 0.5235rad . Training samples are represented by the blue dots and testing samples are represented by the red dots.

Failure Identification Accuracy

xDir	yDir	YawDir	FRComb
100.00%			100.0%
xDir	yDir	YawDir	FRComb
	100%		100.0%
xDir	yDir	YawDir	FRComb
		100%	100.0%
xDir	yDir	YawDir	FRComb
83%	50%		66.7%
xDir	yDir	YawDir	FRComb
83%		100%	91.7%
xDir	yDir	YawDir	FRComb
	83%	83%	83.3%
xDir	yDir	YawDir	FRComb
67%	50%	100%	72.2%
83%	71%	96%	88%

Fig. 7. Failure Detection Accuracy: accuracy percentages for failure detection by deviation subgroups are shown. Total percentages for failure accuracy are shown in the last row.

coordinate axes given the complex geometrical structure of the parts. We will discuss this more in Sec. VI. Another challenge is that different exemplars Average Mean Values and Boundaries overlap each other. Hence, when trying to identify failure modes, exemplars that should not be flagged will be raising a type of false-positive alarm.

We now present the system's effectiveness in detecting deviation-specific failures, correlation modes, and false positives. In Fig. 7, failure detection accuracy is organized by deviation directions and by exemplar classification. The table records: (i) Correct diagnostic's percentage for individual exemplars and (ii) an evaluation measures for the overall combined effect of multiple exemplars. In Fig. 8, the table records the percentage of failed exemplars that were able to be correlated. To clarify, consider an

assembly where there were deviations in the x- and y- directions, and in which My.Rot.AvgMag, Fz.Rot.AvgMag, and Mz.Rot.AvgAmp were flagged. We then consider if these exemplar's average values fell within the bounds derived during training. If so, we state that these exemplars are correlated and have a more reliable measure that the exemplars are not false-positive's.

A. Analysis

The analysis section studies failure detection and failure modes on three levels: (i) Individual Deviation Analysis: which exemplars did better when a given deviation was enacted?; (ii) Individual Exemplar Analysis: how did a given exemplar do across different deviations subgroups; and (iii) Output Computation Analysis: a study of the overall combined effect of exemplars for a given deviation.

1) *Failure Detection*: The first analysis, Individual Deviation Analysis, consists in taking: x , y , ϕ , xy , $x\phi$, $x\phi$, and x, y, ϕ deviations as separate subgroups. In doing a row-by-row analysis it is clear that failure detection in the yDir struggles more than any other direction. The reason is that as soon as the male part is moved in the +xDir or \pm YawDir, horizontal contact either disappears, or it is considerably different than what it was when there was only deviations in yDir. The second analysis, Individual Exemplar Analysis, consists in looking at an individual exemplar across all deviation subgroups. It can be considered a column-by-column analysis and shows that they Yaw deviation detection was the most accurate (96%), then xDir (83%), then yDir (71%). The third analysis, Output Computation Analysis, looks at average value of all exemplars. Generally, the more deviations one inserts, the poorer the accuracy. For a: 1-direction deviation (1D), exemplars assessed failure 100% accurately. For 2-direction deviations (2D): (x,Yaw) and (y,Yaw) both did well at 92% and 83% respectively. The (x,y) subgroup struggled more at 67%. For 3-direction-deviations (3D) exemplars assessed failure with 72% of

Failure Accuracy for Failure Mode Identification

1 Standard Deviation						
MyR	MzR1	MzR23	FzA1	FzA2	FzA3	CorrComb
75.0%						75.0%
MyR	MzR1	MzR23	FzA1	FzA2	FzA3	CorrComb
	75.0%					75.0%
MyR	MzR1	MzR23	FzA1	FzA2	FzA3	CorrComb
		75.0%				75.0%
MyR	MzR1	MzR23	FzA1	FzA2	FzA3	CorrComb
80.0%		33.3%				56.7%
MyR	MzR1	MzR23	FzA1	FzA2	FzA3	CorrComb
80.0%				50.0%		65.0%
MyR	MzR1	MzR23	FzA1	FzA2	FzA3	CorrComb
		40.0%		80.0%		60.0%
MyR	MzR1	MzR23	FzA1	FzA2	FzA3	CorrComb
75.0%		66.7%			66.7%	69.4%
77.5%	75.0%	46.7%	75.0%	65.0%	66.7%	68.0%
2 Standard Deviations						
MyR	MzR1	MzR23	FzA1	FzA2	FzA3	CorrComb
100.0%						100.0%
MyR	MzR1	MzR23	FzA1	FzA2	FzA3	CorrComb
	100.0%					100.0%
MyR	MzR1	MzR23	FzA1	FzA2	FzA3	CorrComb
			75.0%			75.0%
MyR	MzR1	MzR23	FzA1	FzA2	FzA3	CorrComb
80.0%		100.0%				90.0%
MyR	MzR1	MzR23	FzA1	FzA2	FzA3	CorrComb
100.0%				83.3%		91.7%
MyR	MzR1	MzR23	FzA1	FzA2	FzA3	CorrComb
		100.0%		100.0%		100.0%
MyR	MzR1	MzR23	FzA1	FzA2	FzA3	CorrComb
100.0%		100.0%			100.0%	100.0%
95.0%	100.0%	100.0%	75.0%	91.7%	100.0%	93.8%

Fig. 8. Failure Mode Identification: failure mode accuracy percentages by subgroups are presented in the table. The top table shows the results when upper and lower bounds are adjusted by 1 standard deviation. The lower table shows the results when 2 standard deviations are used. Total percentages are shown in the last row of each table.

accuracy. Surprisingly better than the (x,y) subgroup. It tells us that, in our current method, YawDir estimates are better and yDir estimates, a result that is different from [13], in which accuracy degraded as more deviations were inserted.

2) *Failure Mode Identification*: This section analyzes failure modes identification. It evaluates how accurately the correct deviation sub-group was detected for a given trial. The evaluation is done by a user offline. As with our previous research, here too failure sub-group identification shows to be more difficult than failure detection. We perform the same three types of analysis as in failure detection for both standard deviation values: $\sigma = 1$ and $\sigma = 2$.

1D has the highest accuracy for both $\sigma = 1$ and $\sigma = 2$. A different result compared to our previous research is that we do not get degrading performance as we add more deviations. By selecting the exemplars we have used in this work, 3D deviations are more accurate than 2D deviations. When using $\sigma = 2$ however, all accuracy measurement increase from 75% to 100% in 1D; and from 61% to 94% in 2D, and 69% to 100% in 3D.

While in failure detection yDir did the worst, in terms of failure mode detection it did the best. The reason is that the

range of boundary values for MzR1 and MzR23 are quite distinct to each other. Without the overlap it is easier to correctly identify the failure mode. For the FzA's however, the values have lower accuracy except for FzA3 whose value range is smaller than FzA1 or FzA2, helping it make more accurate assessments.

The total combined accuracy marker using the Correlation Combination measure is 68% for 1 standard deviation compared to 68% in [13]. However, when 2 standard deviations are used, the accuracy jumps to 93.8%. Hence, increasing the number of standard deviations increases the exemplar value range used to correctly classify a failure-sub mode. However, as the number of standard deviations increases, other exemplars begin to register false-positives.

While failure mode identification increased in accuracy, there are some challenges to this approach. Even in 1D deviation trials, extraneous exemplars may be flagged. During our exemplar identification stage, salient features may have been recognized in one axis, but deviation in one direction inevitably affected other force axis and states although not in salient ways. This factor can be understood further by realizing that due to the complex hardware configuration of the parts, even deviation motions in one direction may experience contact from other directions leading to unforeseen contact forces in the constrained deviation task. This will be further discussed in Sec. VI.

VI. DISCUSSION

Our work demonstrates that the contextualized early FC scheme performed equally well in detecting failure as in [13]. However, it improved failure-mode identification. Previously in [13], the Combined Output Measurement yielded 68% correct classifications; and with the Max Output Measurement it yielded 81% accuracy. Our work however was 68% to 94% accurate using the Combined Output Measurement. The characterization is more accurate since we consider the dual combination of simultaneous deviation direction analysis and bounded statistical measures instead of maximal and minimal bounds.

Our work provides further insights into how we can possibly begin to understand failure characterization. In essence, by starting with constrained deviations in motion trajectory, we can identify salient features that will aid in the classification of that particular type of error in the task. The challenge then becomes how to analyze and classify the state-space of failure modes. We believe that constraining the problem to “small deviations” is important. In [13], even when using one exemplar for assemblies with 1-3 deviation directions did not work because the exemplar value range was too wide and the system was unable to accurately detect failures and failure modes. As we further contextualize exemplars by separating them by deviation sub-groups, more narrow statistic parameters are obtained. Selecting adequate statistical boundary ranges is important but not trivial and if poorly selected can lead to broad value ranges. In doing so, false-positives occur not only in failure detection, but

also in failure mode identification. For example, in tasks with 3D deviations, such as: FzA1, FzA2, and FzA3, may exemplars could be flagged simultaneously. If there is no user to identify that the task had 3 deviations, the system will not know which one to choose. This work cannot classify the nature of trajectories, whether linear, simple, or compound; it can only infer where and how the contact might have taken place. We have also not studied boundary ranges and how this could be surmised. In future work, we will expand the method both by including more information, such as visual cues and joint information, as well as using more advanced probabilistic models.

The great variation seen in the range of values for exemplar sub-groups: yDir, xyDir, and yYawDir; along with the presence of false-positives lead us to believe that the exemplar space for failure tasks is more complex than originally conceived. We think that it may now be necessary to consider failure characterization by subdividing the task space in three dimensions, through voxel grids or a similar technique and assign each point in 3D space a set of exemplar values (that capture deviation magnitudes) and then use unsupervised classification methods to group the 3D space into similarly (FT) valued spaces. This would form a 3D contact map or failure-force map that can be used to enact failure correction mechanisms. The key is to have specific and accurate information about the magnitude and direction of the failure condition. As our work stands, the authors believe it would not be accurate enough to perform failure correction.

VII. CONCLUSION

In conclusion, a contextualized early failure detection scheme was implemented for cantilever snap assemblies. The method is able to provide early identification of situations in which an assembly trajectory motion deviates from the normal approach trajectory. Further, a contact-state map was generated through sampling the contact space during training. In so doing, we identified not only failure but also what type of failure occurred. The approach was effective in identifying failure modes by further contextualizing statistical measures about contact forces in the task.

REFERENCES

[1] K Althoefer, B Lara, YH Zweiri, and LD Seneviratne. Automated failure classification for assembly with self-tapping threaded fastenings using artificial neural networks. *Proc. of the Inst. of Mech. Engineers, Part C: Journal of Mechanical Engineering Science*, 222(6): 1081–1095, 2008.

[2] Rodney A Brooks. Symbolic error analysis and robot planning. *The Int. J. of Robotics Research*, 1(4):29–78, 1982.

[3] Sohyung Cho, Shihab Asfour, Arzu Onar, and Nandita Kaundinya. Tool breakage detection using support vector machine learning in a milling process. *Int. J. of Machine Tools and Manufacture*, 45(3):241–249, 2005.

[4] Enrico Di Lello, Markus Klotzbucher, Tinne De Laet, and Herman Bruyninckx. Bayesian time-series models for continuous fault detection and recognition in industrial robotic tasks. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ Int. Conf. on*, pages 5827–5833. IEEE, 2013.

[5] David M Fullmer. Parts assembly using signature analysis. US Patent 4,855,923, Aug. 1989.

[6] Yao-Wen Hsueh and Chan-Yun Yang. Prediction of tool breakage in face milling using support vector machine. *The Int. J. of Advanced Manufacturing Technology*, 37(9-10):872–880, 2008.

[7] F. Kanehiro, H. Hirukawana, and S. Kajita. Openhrp: Open architecture humanoid robotics platform. *Intl. J. of Robotics Res.*, 23(2):155–165, 2004.

[8] Ernesto Lopez-Mellado and Rachid Alami. A failure recovery scheme for assembly workcells. In *Robotics and Automation, 1990. Proceedings., 1990 IEEE Int. Conf. on*, pages 702–707. IEEE, 1990.

[9] W. Meeussen, J. Rutgeerts, K. Gadeyne, H. Bruyninckx, and Joris De Schutter. Contact-state segmentation using particle filters for programming by human demonstration in compliant-motion tasks. *IEEE Trans. on Robotics*, 23(2):218–231, 2007.

[10] Alberto Rodriguez, David Bourne, Mathew Mason, Gregory F. Rossano, and JianJun Wang. Failure detection in assembly: Force signature analysis. In *IEEE Conf. on Automation Science and Engineering*, 2010.

[11] Alberto Rodriguez, Matthew T Mason, Siddhartha S Srinivasa, Matthew Bernstein, and Alex Zirbel. Abort and retry in grasping. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ Int. Conf. on*, pages 1804–1810. IEEE, 2011.

[12] Juan Rojas, Kensuke Harada, Hiromu Onda, Natsuki Yamanobe, Eiichi Yoshida, Kazuyuki Nagata, and Yoshihiro Kawai. Towards snap sensing. *Int. J. of Mechatronics and Automation*, 3(2):69–93, 2013.

[13] Juan Rojas, Kensuke Harada, Hiromu Onda, Natsuki Yamanobe, Eiichi Yoshida, and Kazuyuki Nagata. Early failure characterization of cantilever snap assemblies using the pa-rcbht. In *IEEE Int. Conf. on Robotics and Automation (ICRA)*, pages 3370–3377, 2014. doi: 10.1109/ICRA.2014.6907344.

[14] A. Willsky. A survey of design methods for failure detection in dynamic systems. *Automatica*, 12(6): 601:611, 1976.