Engineering evaluation of an embedded computer-vision system for static dimensional screening of cartridge-case overall length

by

Timmy D. Bell

B.S., Bellevue University, 2024

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Integrated Systems Design and Dynamics
College of Technology and Aviation

KANSAS STATE UNIVERSITY
Salina, Kansas

2026

Approved by:

Co-Major Professor
Dr. Michael J. Pritchard

Approved by:

Co-Major Professor
Dr. Balaji Balasubramaniam

Copyright

© Timmy D. Bell 2026.

Abstract

This thesis evaluated CaseMetrology, an embedded computer-vision system developed for static dimensional measurement of cartridge-case overall length (OAL). The formal study was limited to overall length and used a frozen cohort of 60 specimens, an independent manual reference process, and three independently randomized computer-vision placements per specimen for each prospective system configuration.

The original V1 configuration used a principal component analysis (PCA) based estimator and established the prospective baseline. It achieved 100 percent measurement availability but met only one of six quantitative performance criteria. Its ninety-fifth percentile pairwise repositioning difference was 0.004831 inch, indicating substantial presentation-dependent variability.

Subsequent engineering work investigated acquisition stability, fixture behavior, estimator design, and physical presentation. The final V3 configuration combined controlled specimen presentation with a frozen Endface-Hybrid estimator and was prospectively evaluated after configuration freeze. V3 again achieved 100 percent availability. At the specimen level, mean bias was negative 0.000435 inch, mean absolute error was 0.001517 inch, root mean squared error was 0.001794 inch, the nearest-rank ninety-fifth percentile absolute error was 0.003170 inch, and the maximum absolute error was 0.003913 inch. The ninety-fifth percentile pairwise repositioning difference was 0.000633 inch. This approximately 86.9 percent V1-to-V3 reduction reflects changes in both the estimator and physical presentation, so it cannot be attributed to presentation alone. V3 met five of the six retained quantitative criteria. The only unmet criterion was the ninety-fifth percentile absolute error target of 0.0030 inch. Under the unguarded specification rule, V3 produced one false pass and three false fails.

The results demonstrate that repeated presentation stability and absolute agreement with the manual reference were distinct properties of the measurement system. The large improvement in repositioning did not eliminate residual specimen-level disagreement near the specification boundaries. A post-validation guard-band analysis showed that guarded decision logic could reduce observed false pass behavior within the study cohort, but those findings were retrospective and were not validated as a production decision rule.

The completed research supports further development of CaseMetrology as a controlled static dimensional screening architecture. It does not establish the prototype as a replacement for precision reference metrology, a production-ready inline inspection system, or an unconditional final-acceptance device across the full specification interval.

Table of Contents

List of Figures....ix
List of Tables....x
Chapter 1 - Introduction....1
1.1 Background and Motivation....1
1.2 Problem Statement....4
1.3 Purpose of the Study....5
1.4 Research Questions....6
Original Prospective Validation Question....6
Engineering Synthesis Question....7
Exploratory Industrial-Feasibility Question....7
1.5 Research Objectives....7
Original Validation Objectives....7
Subsequent Engineering Objectives....8
1.6 Scope and Delimitations....9
1.7 Significance of the Study....11
1.8 Organization of the Thesis....12
Chapter 2 - Literature Review....14
2.1 Introduction....14
2.2 Industrial Machine Vision and Dimensional Inspection....15
2.3 Backlit Dimensional Vision and Boundary Localization....16
2.3.1 Backlit silhouette measurement....16
2.3.2 Thresholding, contours, and subpixel localization....18
2.4 Optical Geometry and Lens Architecture....19
2.4.1 Calibration and measurement-plane geometry....19
2.4.2 Conventional and telecentric optics....20
2.5 Physical Presentation and Fixturing....21
2.6 Calibration, Traceability, and Measurement Uncertainty....23
2.6.1 Calibration and traceability....23
2.6.2 Reference measurements and uncertainty....24

2.7 Repeatability, Accuracy, and Measurement-System Agreement....25
2.7.1 Repeatability is not accuracy....25
2.7.2 Method agreement and validation design....27
2.8 Conformity Decisions, Guard Bands, and Screening....28
2.8.1 Decisions under measurement uncertainty....28
2.8.2 Guarded screening architectures....29
2.9 Optical and Computational Analysis of Cartridge Cases....30
2.9.1 Historical manufacturing inspection....30
2.9.2 Modern forensic imaging context....31
2.10 Literature Synthesis and Research Position....32
Chapter 3 - Methodology....35
3.1 Research Design and Experimental Chronology....35
3.2 CaseMetrology System Architecture....37
3.2.1 Embedded computing and imaging hardware....37
3.2.2 Image acquisition and segmentation....38
3.2.3 PCA OAL estimator....39
3.2.4 Endface-Hybrid OAL estimator....39
3.2.5 Configuration control and data preservation....40
3.3 Calibration and Manual Reference Measurement....42
3.3.1 Spatial and directional calibration....42
3.3.2 Manual reference instrument and session assurance....43
3.3.3 Formal manual measurement procedure....44
3.3.4 Reference uncertainty and near-limit classification....45
3.4 Specimen Population, Screening, and Frozen Cohort....46
3.5 V1 Prospective Formal Validation....47
3.5.1 Frozen V1 configuration and acquisition....47
3.5.2 Specimen-level analysis....48
3.5.3 Repositioning, availability, and classification....49
3.5.4 Prospectively frozen performance criteria....50
3.6 Post-V1 Diagnostic Investigation....51
3.6.1 Ambient-light diagnostic....51

3.6.2 Controlled-repositioning diagnostic....52
3.6.3 Fixture and presentation diagnostics....52
3.7 V2 Fixture Engineering Iteration and Formal Evaluation....53
3.8 Endface-Hybrid Development and Retrospective V2 Evaluation....54
3.9 V3 Controlled-Presentation Development, Freeze, and Prospective Validation....55
3.9.1 Development boundary and controlled-presentation rule....55
3.9.2 Final V3 freeze....56
3.9.3 Capture-only formal acquisition....56
3.9.4 Post-acquisition estimator processing and primary analysis....57
3.9.5 Comparative analysis boundaries....58
3.10 Exploratory Guard-Band Analysis....59
Chapter 4 - Results....61
4.1 Prospective V1 PCA Results....61
4.1.1 Availability, agreement, and repositioning....61
4.1.2 Classification and criteria....62
4.2 Post-V1 Diagnostic Results....63
4.3 Prospective V2 PCA Results....64
4.3.1 Availability, agreement, and repositioning....64
4.3.2 Classification and criteria....64
4.4 Retrospective V2 Endface-Hybrid Results....65
4.5 Controlled-Presentation Development Results....66
4.6 Prospective V3 Endface-Hybrid Results....66
4.6.1 Dataset completion and availability....66
4.6.2 Specimen-level agreement and repositioning....67
4.6.3 Classification and prospective criteria....67
4.7 Comparative V1–V3 Results....68
4.8 Exploratory Post-Hoc Guard-Band Results....69
4.9 Chapter Summary....70
Chapter 5 - Discussion and Conclusions....72
5.1 Answer to the Original Validation Question....72
5.2 Engineering Progression....74

5.3 Repeatability Versus Absolute Accuracy....76
5.4 Industrial Screening Feasibility....79
5.5 Guarded Decision Logic....80
5.6 Limitations....82
5.7 Future Work....85
5.8 Final Conclusion....87
References....90

List of Figures

Figure 2.1. Conceptual contact pathways between the case and fixture.....23
Figure 2.2. Conceptual distinction between repeatability and manual-reference agreement.....26
Figure 3.1. Schematic definition of cartridge-case OAL used in the formal analysis.....35
Figure 3.2. CaseMetrology physical system and specimen presentation.....38
Figure 3.3. Conceptual CaseMetrology measurement architecture.....41
Figure 4.1. Prospective V1–V3 comparison of P95 repositioning and P95 absolute specimen error.....69

List of Tables

Table 3.1. Experimental phases and inferential status....36
Table 3.2. Frozen formal-cohort allocation....46
Table 3.3. Prospectively frozen OAL performance criteria....50
Table 4.1. Prospective V1 PCA performance....62
Table 4.2. Prospective V1 criterion results....62
Table 4.3. Post-V1 fixed-specimen ambient-light diagnostic....63
Table 4.4. Prospective V2 PCA performance....64
Table 4.5. Prospective V2 PCA criterion results....64
Table 4.6. Retrospective V2 Endface-Hybrid performance....65
Table 4.7. Development-only Endface-Hybrid fixed-roll results....66
Table 4.8. Prospective V3 Endface-Hybrid performance....67
Table 4.9. Prospective V3 criterion results....68
Table 4.10. Comparison of OAL performance across system iterations....68
Table 4.11. Exploratory symmetric guard-band results for the V3 cohort....70

Chapter 1 - Introduction

1.1 Background and Motivation

Machine vision is widely used in manufacturing to inspect parts, identify features, monitor processes, and extract dimensional or quality-related information from images. Industrial vision systems can reduce dependence on manual visual inspection and provide repeatable computational methods for evaluating manufactured components when the imaging conditions, measurement objective, and decision process are appropriately defined (Malamas et al., 2003; Hashmi et al., 2022). For dimensional inspection, however, detecting an object or producing a numerical value from an image does not by itself establish that the result is suitable for measurement or manufacturing decisions.

Dimensional machine vision is a measurement-system problem as well as a computer-vision problem. Image formation, optical geometry, calibration, physical presentation of the workpiece, feature localization, reference standards, repeated-measurement behavior, and measurement uncertainty can all influence the reported dimension. Consequently, a computer-vision measurement should be evaluated not only by whether an algorithm successfully detects the intended boundaries, but also by how the resulting physical measurement behaves relative to an appropriate reference process and under repeated measurement conditions (Joint Committee for Guides in Metrology, 2008; Joint Committee for Guides in Metrology, 2012b).

Backlit silhouette imaging provides a useful architecture for this type of dimensional inspection because it can create high contrast between a manufactured component and its background. Previous dimensional machine-vision studies have used backlighting, edge or contour localization, calibrated image scale, geometric modeling, and controlled specimen presentation to measure manufactured parts (Shin et al., 2002; Xiang et al., 2018; Che & Ratnam,

2018). These studies illustrate that measurement performance depends on the complete physical and computational system rather than on image segmentation alone.

At the specimen–fixture interface, the location and stability of contact can also influence physical presentation. Contact mechanics and tribology provide a useful vocabulary for describing possible seating, friction, and wear pathways; the present study did not measure those pathways separately (Greenwood & Williamson, 1966; Archard, 1953).

Cartridge cases provide a practical test object for investigating these relationships. Optical and computerized analysis of cartridge cases is well established, particularly in forensic applications involving surface-feature acquisition and comparison. Historical work also demonstrates that electro-optical dimensional inspection of cartridge cases, including case-length measurement, predates the present research (Coleman et al., 1974; Coleman & Swinth, 1977; Swinth et al., 1979). The contribution of the present study therefore does not depend on claiming novelty for optical cartridge-case measurement itself. Instead, the research examines the engineering behavior of a specific embedded computer-vision architecture used to measure cartridge-case overall length under controlled static conditions.

Overall length (OAL) was selected as the formal dimensional characteristic because it provides a clearly defined end-to-end measurement while allowing the study to concentrate on one measurement problem. Restricting the formal work to OAL made it possible to evaluate calibration, reference measurement, repeated measurement, physical presentation, estimator behavior, and measurement agreement without simultaneously introducing additional dimensional features and measurement definitions.

An important distinction in this research is the difference between **precision dimensional metrology** and **dimensional screening**. A measurement system intended to replace a precision

reference instrument must demonstrate performance appropriate to that metrological role across the intended measurement range. A screening system can serve a different manufacturing function. Rather than requiring every image-derived measurement to function as a final reference-quality dimensional value, a screening system may be used to identify measurements that are comfortably within an acceptance region, distinguish clearly nonconforming measurements, and refer uncertain or near-limit measurements for secondary inspection. Such an architecture still requires quantitative characterization because measurement uncertainty and classification risk must be understood rather than ignored.

CaseMetrology was developed as a static, embedded, backlit computer-vision system for investigating this problem. The research progressed through successive formal validation and engineering iterations. The initial configuration established a prospective measurement baseline. Subsequent diagnostic and fixture-development work investigated sources of repeated-measurement variation, and later configurations introduced improved physical presentation control and a revised OAL estimator. The final prospective configuration provided an opportunity to determine whether improved presentation stability also improved absolute dimensional performance and whether the resulting system behavior supported a defensible industrial-screening interpretation.

Accordingly, the motivation for this thesis extends beyond determining whether a camera and image-processing algorithm can estimate cartridge-case length. The broader engineering question is whether a complete embedded vision system can produce sufficiently stable and characterized dimensional information to support practical inspection decisions, what limitations remain after physical presentation is controlled, and how those limitations affect the role that such a system could reasonably perform in an industrial setting.

1.2 Problem Statement

A computer-vision system can produce dimensional values that appear plausible while still being unsuitable for final inspection decisions. For a vision-derived dimension to support manufacturing use, the measurement process must demonstrate acceptable behavior relative to an independent reference, adequate repeatability under repeated presentation, reliable measurement availability, and decision performance appropriate to the intended application.

The CaseMetrology research addressed this problem through successive engineering and validation stages rather than through evaluation of a single image-processing algorithm. The original formal validation established a baseline for measurement agreement and repeated-measurement behavior. Those results showed that the initial system configuration did not satisfy all of the predeclared performance criteria and that placement-dependent variation was a substantial limitation.

Subsequent engineering work therefore focused on understanding and reducing sources of measurement variation. Fixture changes, controlled repositioning studies, estimator development, and controlled-presentation testing were used to determine whether the dominant repeated-measurement limitation could be reduced without changing the fundamental embedded, backlit imaging architecture.

The resulting V3 configuration substantially improved presentation repeatability, but absolute dimensional error remained materially larger than the repositioning error. This created a second engineering problem: improved repeatability alone did not establish that the prototype could be treated as interchangeable with a precision reference instrument across the entire specification interval.

The completed study therefore addresses two related but distinct questions. First, it evaluates whether successive engineering changes improved the measurement behavior of the CaseMetrology system under controlled static conditions. Second, it examines whether the resulting level of performance is more appropriately interpreted as precision metrology or as dimensional screening in which near-limit measurements require secondary inspection.

This distinction is important because the practical value of a dimensional vision system depends on how its demonstrated uncertainty and error characteristics relate to the decision being made. A system that is not suitable as a direct replacement for precision contact measurement may still provide useful screening capability if its measurement behavior is sufficiently stable, its limitations are characterized, and decision rules are designed to account for uncertainty near specification boundaries.

Accordingly, the central problem addressed by this thesis is not simply whether cartridge-case overall length can be estimated from an image. The problem is whether an embedded computer-vision measurement architecture can be engineered and characterized well enough to support defensible dimensional inspection decisions, and what limits remain after major sources of presentation-dependent variation are reduced.

1.3 Purpose of the Study

The purpose of this study is to evaluate the engineering performance and practical dimensional-inspection potential of the CaseMetrology embedded computer-vision system for static measurement of cartridge-case overall length.

The study first evaluates the system against prospectively defined measurement-performance criteria using an independent manual reference process. It then examines how successive engineering changes to fixture design, part presentation, and measurement estimation

affected repeated-measurement behavior and absolute dimensional performance across V1, V2, and V3.

A further purpose is to determine whether the final demonstrated system behavior is more appropriately interpreted as precision metrology or as dimensional screening. This distinction is evaluated using formal V3 performance results together with an exploratory post-validation guard-band analysis intended to characterize the tradeoff between automatic acceptance and secondary review near specification limits.

The study does not attempt to validate the prototype as a production-ready inline inspection system or as a replacement for precision contact measurement. Instead, it seeks to establish what the demonstrated system can and cannot support under controlled static conditions and to identify the engineering evidence required for any future production-oriented validation.

1.4 Research Questions

The completed research is organized around three related questions that reflect the actual chronology of the study. The first question governed the original prospective validation. The second and third questions synthesize the engineering evidence generated after the original validation and are not presented as prospectively defined V1 hypotheses.

Original Prospective Validation Question

Can the frozen CaseMetrology computer-vision system measure cartridge-case overall length with acceptable agreement and repeated-measurement performance relative to an independent manual reference measurement process under documented static inspection conditions?

This question is evaluated using the prospectively established performance criteria and the corresponding formal validation results. The thesis reports explicitly which criteria were satisfied and which were not.

Engineering Synthesis Question

What do successive CaseMetrology design iterations reveal about the roles of physical presentation, estimator design, and residual measurement error in the feasibility of embedded computer-vision OAL screening?

This question synthesizes evidence from V1, the post-V1 diagnostic program, V2, Endface-Hybrid estimator development, controlled-presentation development, and prospective V3 validation. Its purpose is to determine what the completed experimental sequence demonstrates about the principal engineering limitations of the system.

Exploratory Industrial-Feasibility Question

Can guarded decision logic provide a practical pathway for using the demonstrated measurement system as a dimensional screening tool while reducing observed false-pass risk near specification limits?

This question is addressed through post-validation exploratory analysis. It is not used to redefine the original acceptance criteria and does not constitute validation of a production guard-band rule.

1.5 Research Objectives

The objectives of the study reflect both the original prospective validation and the subsequent engineering work that followed from the observed results.

Original Validation Objectives

The original validation objectives were to:

establish and freeze the static backlit imaging configuration, acquisition settings, OAL measurement procedure, calibration model, and software environment before formal validation;
establish an independent manual reference-measurement process using documented instruments, check-standard controls, repeated measurements, and session-acceptance criteria;
select and preserve a formal validation cohort using a documented screening and stratified random-selection procedure;
acquire repeated manual-reference and computer-vision measurements under controlled conditions;
quantify availability, systematic bias, absolute error, repeated-measurement behavior, and classification performance using prospectively defined statistical criteria; and
determine whether the original CaseMetrology configuration satisfied those predefined validation criteria.

Subsequent Engineering Objectives

Following the original validation, the engineering objectives were expanded to:
investigate the principal sources of repeated-measurement variability observed during V1;
evaluate whether fixture and presentation changes reduced placement-dependent measurement variation;
evaluate the Endface-Hybrid estimator using development data and retrospective V2 analysis while preserving the distinction between retrospective and prospective evidence;
prospectively evaluate the frozen Endface-Hybrid estimator under the V3 controlled-presentation configuration;
determine whether substantial improvement in presentation repeatability produced a corresponding improvement in absolute dimensional accuracy; and

explore whether guarded decision logic could support a practical dimensional-screening architecture while reducing observed false-pass risk near specification boundaries.

These objectives preserve the chronology of the study rather than representing later engineering questions as though they had been part of the original V1 validation plan.

1.6 Scope and Delimitations

The formal dimensional characteristic evaluated in this thesis is cartridge-case overall length. Overall length was selected as the most suitable initial dimension for evaluating the measurement system and its fixturing. The system could be adapted to other dimensions with different fixturing, but each new measurand would require its own calibration and validation. Other cartridge-case characteristics, including rim geometry, primer-pocket dimensions, concentricity, wall thickness, and additional dimensional features, are outside the formal validation scope.

The system is evaluated under static inspection conditions. The study does not validate conveyor-based inspection, moving workpieces, robotic handling, automated part loading, or full-rate production operation. Conclusions regarding system performance therefore apply to the documented static measurement configuration rather than to an automated manufacturing line.

The experimental system uses an embedded NVIDIA Jetson Orin Nano Super Developer Kit, an Arducam USB 3.0 IMX296 global-shutter camera, a 16 mm CS-mount lens, and LED backlighting. The 16 mm lens was selected because it was readily available off the shelf at a practical cost for this prototype. The physical measurement geometry, imaging controls, calibration approach, fixturing, software environment, and frozen estimator configurations are part of the evaluated system. Results should not be generalized automatically to other cameras, lenses, working distances, illumination systems, or processing pipelines.

The source population consisted of 235 uniquely identified cartridge cases. Five specimens used during development were excluded from formal specimen selection. A 60-specimen stratified cohort was then selected for formal validation. The cohort intentionally included specimens distributed around the lower and upper specification limits and therefore was designed to challenge measurement and classification behavior rather than represent natural production yield.

The manual reference process used a single operator for all formal acquisitions. The study therefore characterizes within-operator repeated-measurement behavior under the documented conditions but does not estimate between-operator reproducibility.

The manual reference measurement is treated as the comparison basis for evaluating the computer-vision system, but it is not assumed to be error-free. Reference-measurement uncertainty, instrument resolution, operator influence, and physical case-end geometry remain relevant when interpreting small differences between the two methods.

V2 Endface-Hybrid results are retrospective because the estimator was applied to previously acquired V2 images. Those results are used for comparative and engineering interpretation but are not represented as prospective validation.

The V3 Endface-Hybrid evaluation is prospective with respect to the frozen estimator and controlled-presentation configuration. The estimator and V3 configuration were frozen before formal acquisition, and formal estimator-result interpretation was withheld until completion of the 180-image acquisition.

The guard-band analysis was conducted after the formal V3 results were available. It is therefore exploratory and retrospective. It demonstrates decision-behavior tradeoffs within the formal validation cohort but does not validate a production guard-band rule.

The study does not claim that the demonstrated prototype is production-ready, that it replaces precision contact metrology, or that it has been validated for universal cartridge-case populations or operating environments.

1.7 Significance of the Study

The significance of this research lies in the evaluation of dimensional machine vision as a complete engineering measurement system rather than as an image-processing task alone.

The study demonstrates how physical presentation, optical measurement, calibration, reference comparison, estimator behavior, and decision criteria interact in a practical embedded vision system. The successive V1, V2, and V3 configurations provide evidence that a system can become substantially more repeatable without achieving a corresponding proportional improvement in absolute accuracy.

This distinction is important for industrial dimensional inspection. A system that produces highly repeatable results may still contain systematic or residual measurement error that limits its suitability for direct replacement of a precision reference instrument. Conversely, a system that does not satisfy all precision-metrology requirements may still provide useful inspection value if its limitations are understood and incorporated into the decision process.

The V3 results are particularly important because P95 repositioning decreased from 0.004831 in for the V1 configuration to 0.000633 in for the final V3 configuration, an overall system-level improvement of approximately 86.9%. Because estimator and presentation conditions both changed across the full V1-to-V3 sequence, that reduction is not attributed to controlled presentation alone. At the same time, the specimen-level nearest-rank V3 P95 absolute error remained 0.003170 in. The completed experimental sequence therefore supports the

conclusion that repeated-presentation performance and residual absolute measurement error were distinct limitations of the prototype.

The contact-based interpretation is deliberately limited. More consistent seating and roll control are compatible with reduced presentation variation, but the V1-to-V3 difference includes an estimator change and does not demonstrate a measured reduction in friction, wear, or deformation. The broader lesson is that a dimensional result emerges from interactions among the specimen, fixture, optical sensor, reference process, and interpretation algorithm.

This finding shifts the practical interpretation of the system. The completed prototype was not demonstrated as interchangeable with precision contact measurement across the complete specification interval. However, the measured stability, availability, low systematic bias, and exploratory guard-band behavior support further evaluation of the architecture as a dimensional-screening system.

The research also contributes a documented engineering progression in which unsuccessful or partially successful validation results were used to guide subsequent design changes. V1 was not discarded when its criteria were not met. Instead, its results provided evidence that motivated diagnostic testing, fixture improvements, estimator development, and controlled-presentation validation. The resulting sequence illustrates the value of preserving negative and intermediate results as part of an engineering research record.

Finally, the study establishes a traceable evidence base for future work. The formal specimen cohort, acquisition records, estimator freezes, analysis definitions, performance results, and limitations have been preserved so that future studies can evaluate alternative optics, operators, presentation systems, decision rules, additional dimensions, or production-oriented automation without obscuring the basis of the present conclusions.

1.8 Organization of the Thesis

Chapter 1 introduces the dimensional-inspection problem, the motivation for the study, the research questions, objectives, scope, delimitations, and significance of the completed V1-through-V3 engineering program.

Chapter 2 reviews literature relevant to industrial machine vision, backlit dimensional inspection, vision-based dimensional metrology, optical geometry, calibration, measurement uncertainty, repeatability, physical presentation, dimensional decision rules, and optical analysis of cartridge cases.

Chapter 3 describes the CaseMetrology hardware and software architecture, calibration and manual reference processes, specimen population and formal cohort, V1 validation methodology, post-V1 diagnostic investigations, V2 engineering iteration, Endface-Hybrid estimator development, V3 controlled-presentation development and freeze, prospective V3 validation, and exploratory guard-band methodology.

Chapter 4 presents the experimental results in chronological order. It reports the V1 formal results, post-V1 diagnostic findings, V2 formal PCA results, retrospective V2 Endface-Hybrid results, controlled-presentation development evidence, prospective V3 formal results, comparative V1–V3 performance, and exploratory guard-band results.

Chapter 5 interprets the findings and answers the research questions. Particular emphasis is placed on the distinction between repeatability and absolute accuracy, the engineering progression from V1 through V3, the practical implications for dimensional screening, the limitations of the demonstrated system, and the requirements for future production-oriented validation.

Chapter 2 - Literature Review

2.1 Introduction

The use of machine vision for dimensional inspection lies at the intersection of industrial automation, computer vision, and measurement science. A camera-based system can detect a workpiece and calculate a dimension, but those functions alone do not establish that the reported value is suitable for a manufacturing decision. Dimensional validity also depends on how the image is formed, how physical boundaries are localized, how image coordinates are related to physical units, how the workpiece is presented, how repeated observations behave, and how measurement uncertainty is incorporated into the eventual decision.

These distinctions are especially important for the present study because its final engineering interpretation concerns dimensional screening rather than unrestricted replacement of precision contact metrology. A screening system and a reference-quality measuring system may use similar image-processing operations, yet their required evidence and permissible decisions are different. The literature therefore must address not only how dimensions can be extracted from images, but also how such measurements are characterized and how uncertainty affects acceptance near specification limits.

This chapter reviews nine connected topics: industrial machine vision and dimensional inspection; backlit silhouette measurement; boundary localization and subpixel methods; optical geometry and lens architecture; physical presentation and fixturing; calibration, traceability, and uncertainty; repeatability, accuracy, and method agreement; conformity decisions and guard bands; and prior optical analysis of cartridge cases. The chapter concludes by defining the research position of CaseMetrology without claiming novelty for optical cartridge-case measurement itself. Findings specific to CaseMetrology are reserved for Chapters 4 and 5.

2.2 Industrial Machine Vision and Dimensional Inspection

Industrial machine vision combines image acquisition, computation, and application-specific decision logic to inspect manufactured products and processes. Malamas et al. (2003) described industrial vision systems as supporting tasks such as identification, gauging, assembly verification, surface inspection, and process control. Hashmi et al. (2022) similarly reviewed measurement-oriented applications of machine vision in manufacturing. Across these applications, the usefulness of a vision system depends on a clearly defined inspection objective, appropriate image acquisition, and performance demonstrated under the conditions in which the system will be used.

Dimensional inspection differs from detection or classification because it requires a defensible relationship between image information and a physical measurand. The International Vocabulary of Metrology defines a measurand as the quantity intended to be measured and treats a measurement result as more than an instrument indication alone (Joint Committee for Guides in Metrology, 2012b). In image-based dimensional work, the measurand must therefore be defined independently of the pixels or contours used to estimate it. A segmented silhouette is evidence about a physical boundary; it is not itself the physical dimension.

Coveney's (2014) National Physical Laboratory guide describes dimensional vision systems as integrated arrangements for image formation and sensing, image processing, and communication of results. The guide emphasizes that illumination, lens selection, camera behavior, calibration, edge detection, distortion, and uncertainty all influence dimensional measurement. This complete-system view is consistent with Hemming's (2007) metrological treatment of machine vision and with the uncertainty framework developed for optical vision measuring machines by Lazzari and Iuculano (2004).

Industrial use also requires a defined relationship between the measurement and the action taken from it. Historical manufacturing systems have used optical measurements for automatic classification, rejection, and process feedback, including high-throughput cartridge-case inspection (Coleman et al., 1974; Coleman & Swinth, 1977; Swinth et al., 1979). Such examples demonstrate that image-derived dimensions can support production decisions. They do not establish that every optical measurement architecture has equivalent accuracy, uncertainty, or decision risk. Fitness for purpose remains an application-specific property that must be demonstrated for the implemented system.

This distinction supports two possible roles for dimensional machine vision. In a metrology role, the reported value is expected to perform as a defensible dimensional result across the intended range and operating conditions. In a screening role, the system may instead separate clear cases from measurements that require confirmation. Screening does not eliminate the need for validation. It changes the decision architecture: system limitations are managed through defined acceptance, review, and rejection logic rather than ignored.

2.3 Backlit Dimensional Vision and Boundary Localization

2.3.1 Backlit silhouette measurement

Backlighting is a widely used arrangement for dimensional vision because transmitted light can create strong contrast between an opaque workpiece and the illuminated background. Coveney (2014) identifies contrast and field uniformity as primary considerations when selecting illumination for dimensional measurement. A high-contrast silhouette can simplify the localization of external boundaries, but the measured dimension still depends on how those boundaries are imaged and interpreted.

Shin et al. (2002) developed a machine-vision system for dimensional inspection of ball studs using backlit silhouetted images. Their system extracted outlines at subpixel resolution, modeled boundary data with geometric primitives, and used jig fixtures and controlled motion to present different regions of the part. The study is relevant because it combines illumination, edge localization, geometric interpretation, and physical presentation within one dimensional measurement architecture.

Xiang et al. (2018) used a two-camera backlit arrangement to measure the spacing of automobile brake-pad bayonets. Their method incorporated a purpose-designed dimensional standard, relative edge-position measurement, edge detection, and line fitting. The work is also important because it evaluated both dimensional bias and repeatability against application-specific requirements. Its numerical results are not directly transferable to a single-camera cartridge-case system, but the experimental structure illustrates that a dimensional vision method should be evaluated as more than an edge-detection algorithm.

Che and Ratnam (2018) applied LED backlighting and subpixel edge localization to monitor workpiece diameter during turning. Vision-derived values were compared quantitatively with digital-caliper measurements. The workpiece remained constrained by the machine tool, so the presentation conditions differ from independent static placement. Nevertheless, the study provides precedent for converting silhouette-edge information into a physical dimension and evaluating it against a conventional contact method.

Collectively, these studies establish that backlit dimensional vision is an integrated measurement approach. Illumination creates the observable boundary, the optical system maps that boundary to the sensor, software estimates its location, calibration converts image position to physical units, and specimen presentation determines the geometry under which those

operations occur. Failure or variability in any one of those relationships can affect the final dimension.

2.3.2 Thresholding, contours, and subpixel localization

Silhouette processing often begins with separation of the object from its background. Otsu's (1979) method selects a global threshold from the gray-level histogram by optimizing between-class separation. It is a general segmentation method rather than a dimensional estimator. A binary mask produced by thresholding identifies foreground and background regions, but it does not by itself define which points correspond to the intended physical endpoints of a manufactured part.

After segmentation, dimensional systems may use contour extraction, fitted lines or circles, geometric primitives, principal directions, or local edge models. Principal component analysis provides a general method for determining dominant directions in multivariate data (Jolliffe & Cadima, 2016), but the conversion of a component direction into a physical end-to-end measurement remains application specific. Similarly, a fitted geometric model can stabilize a boundary estimate only when the chosen model corresponds appropriately to the measured feature.

Subpixel localization is frequently used when the desired dimensional discrimination is finer than one sensor pixel. Shin et al. (2002) and Che and Ratnam (2018) incorporated subpixel edge methods into backlit dimensional systems. Lai et al. (2026) combined a telecentric calibration framework with distortion correction, contour information, illumination compensation, and subpixel edge detection. Their work demonstrates that image-space localization finer than one pixel is possible under a carefully modeled system.

Subpixel output should not be equated automatically with subpixel physical accuracy. Pixel interpolation or model fitting can produce fractional-pixel coordinates, but the physical accuracy of those coordinates still depends on focus, optical distortion, illumination, threshold or edge model, calibration, workpiece geometry, and presentation. Coveney (2014) likewise treats threshold setting, magnification, distortion, pixel calibration, and measurement uncertainty as connected aspects of dimensional vision. The literature therefore supports subpixel localization as a measurement tool, not as an independent guarantee of metrological performance.

2.4 Optical Geometry and Lens Architecture

2.4.1 Calibration and measurement-plane geometry

Monocular dimensional measurement depends on the relationship among the camera, lens, calibrated coordinate system, measurement plane, and workpiece. For a conventional perspective lens, changes in object distance or orientation can alter projected size. Lateral field position can also matter when distortion or spatially varying magnification is not adequately controlled or modeled.

Wang et al. (2024) analyzed dimensional error arising when the physical measurement plane differs from the plane represented during camera calibration. Their theoretical and experimental work considered translation and rotation between calibration and measurement geometries and included both conventional industrial and telecentric optics. Under the assumptions evaluated, displacement along the optical axis could produce systematic scale error, while more general nonperpendicular geometry produced more complex relationships. The broader implication is that a calibration is valid within a physical geometry, not merely within a software file.

Coveney (2014) similarly recommends evaluating distortion and magnification across the usable image area and repeating pixel calibration within the area selected for measurement. The guide notes that distortion-related errors are often most apparent near the edge of the field. These principles are relevant to conventional area-camera systems because a single nominal pixels-per-unit value does not automatically describe all image locations and orientations with equal fidelity.

The literature establishes plausible mechanisms by which geometric change can influence a dimensional result. It does not justify assigning a particular error mechanism to another system without an experiment that isolates the relevant displacement, rotation, or field-position effect. In the present thesis, general optical-geometry literature supplies the rationale for controlling and documenting presentation geometry; CaseMetrology-specific causal conclusions must come from the project's own experiments.

2.4.2 Conventional and telecentric optics

Telecentric lenses are often used in high-precision dimensional vision because they reduce perspective effects and maintain approximately constant magnification over a defined depth range. Lai et al. (2026) developed a telecentric dimensional system with a calibration model matched to telecentric image formation. Wang et al. (2024) also included telecentric optics when experimentally evaluating external-geometry effects. These sources demonstrate why telecentric systems can be advantageous where object depth and perspective sensitivity are critical.

Telecentric optics do not eliminate the need for calibration, distortion characterization, boundary localization, illumination control, or uncertainty analysis. Lai et al. (2026) treated those elements jointly despite using a telecentric lens. Telecentric and conventional systems also

involve different imaging models, fields of view, hardware costs, and physical constraints. Performance reported for a telecentric system cannot therefore be transferred directly to a conventional-lens system.

For conventional optics, dimensional performance can still be evaluated empirically and may be adequate for a defined task. The appropriate conclusion from the literature is not that one optical architecture is universally required, but that the selected architecture, calibration model, working distance, measurement plane, and allowed presentation variation must be mutually consistent with the required measurement and decision performance.

2.5 Physical Presentation and Fixturing

Physical presentation determines how the workpiece occupies the calibrated imaging geometry. Relevant degrees of freedom can include translation within the field, rotation within the measurement plane, displacement along the optical axis, tilt relative to the measurement plane, and part-specific roll or seating behavior. Some of these changes may affect scale or distortion; others may alter which physical surface or edge becomes the apparent silhouette boundary.

Prior dimensional systems often controlled presentation as part of the measurement design. Shin et al. (2002) used jig fixtures and servo-controlled rotation and translation in the ball-stud system. Che and Ratnam’s (2018) workpiece remained mechanically constrained by the lathe. The historical cartridge-case profile monitors used dedicated mechanical handling, gauging fixtures, and precision reference surfaces rather than treating part placement as incidental (Coleman et al., 1974; Coleman & Swinth, 1977; Swinth et al., 1979).

The value of these examples is architectural. They show that a dimensional vision system can include mechanical registration and controlled motion as integral measurement components.

The examples do not establish a universal fixture design, nor do they prove that fixturing alone determines measurement accuracy. Hemming (2007) found that dominant uncertainty contributions differed by application and could arise from optical, mechanical, geometric, algorithmic, environmental, operator, or object-related effects.

A fixture may improve repeated presentation by constraining relevant degrees of freedom, but it can also introduce contact, obstruction, deformation, or a new datum relationship. Fixture evaluation should therefore consider whether the part is seated consistently, whether the measured edges remain visible, whether contact forces change the measurand, and whether the fixture maintains the workpiece within the calibrated geometry. A fixture's purpose is not simply to hold the object still; it is to establish repeatable measurement conditions appropriate to the measurand.

Contact mechanics offers a more precise way to frame that relationship. The real contact between nominally flat surfaces depends on local topography and deformation; contact area and wear can change with loading and rubbing (Greenwood & Williamson, 1966; Archard, 1953). In a fixture, the base support, mouth stop, and lateral locator may therefore establish slightly different seating states when a case is removed and replaced. This is a plausible pathway, not a measured explanation of the CaseMetrology result.

Studies of forming contacts show that surface finish, treatment, and coatings can influence galling and material transfer; controlled sliding experiments also show sensitivity to scratches and counterface material (Podgornik et al., 2006; Heinrichs et al., 2012). Multipass scratch testing demonstrates why coating durability depends on repeated contact conditions (Fallqvist et al., 2012). These studies concern different materials, loads, and sliding histories.

They motivate future fixture-material and surface-condition checks but provide no evidence that CaseMetrology developed galling, transfer films, or measurable wear.

This literature provides the conceptual foundation for investigating presentation-dependent variation. It also establishes an important boundary: the general importance of presentation and fixturing is established in prior work, whereas the magnitude of their effects in CaseMetrology is an experimental finding of this thesis and is not assumed from the literature.

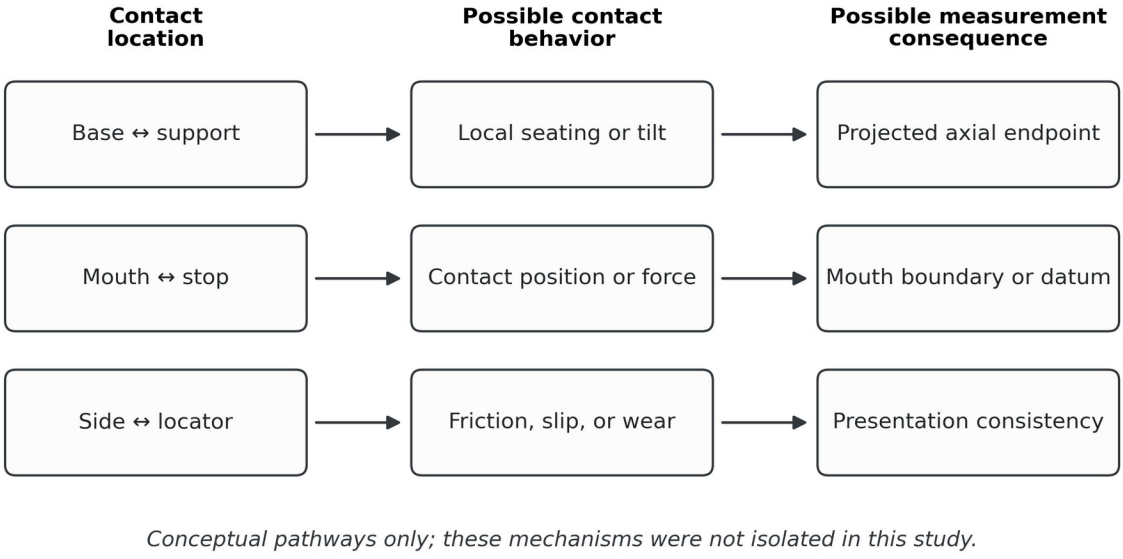


Figure 2.1. Conceptual contact pathways between the case and fixture. Potential influences on the silhouette and seating are shown as hypotheses, not isolated effects observed in this study.

2.6 Calibration, Traceability, and Measurement Uncertainty

2.6.1 Calibration and traceability

Calibration establishes a relationship between indications and known reference values under specified conditions. Metrological traceability requires a documented, unbroken chain of calibrations linking a measurement result to an appropriate reference, with each link contributing to measurement uncertainty (Joint Committee for Guides in Metrology, 2012b). For dimensional

vision, this commonly requires a physical length artifact or calibrated scale together with a defined model relating image coordinates to physical units.

Calibration is necessary but not sufficient. A valid scale factor does not characterize boundary localization, distortion, workpiece geometry, presentation variation, environmental effects, or reference-method behavior. Hemming (2007) demonstrated that machine-vision uncertainty budgets are application specific and that dominant contributions can differ among systems. In some applications, camera, lens, and edge-finding behavior were important; in others, mechanical or geometric effects dominated.

Lazzari and Iuculano (2004) treated a camera-equipped optical measuring machine as a metrological system and developed a GUM-oriented uncertainty model. Their work distinguished verification of instrument performance from evaluation of uncertainty associated with a measurement result. Related work from the same research program used certified dimensional standards and repeated measurements while separating machine-related and vision-related behavior (Lazzari et al., 2001). These studies provide precedent for examining subsystems individually while maintaining a complete-system measurement model.

The practical guidance from Coveney (2014) is consistent with those studies. Pixel calibration, threshold selection, distortion, magnification, repeated observations, calibration-standard uncertainty, and environmental conditions can all contribute to the final result. A manufacturer’s accuracy specification or a successful check measurement is not automatically equivalent to the uncertainty of a measurement made on a particular workpiece.

2.6.2 Reference measurements and uncertainty

The Guide to the Expression of Uncertainty in Measurement provides a general framework for identifying input quantities, evaluating uncertainty contributions, and combining

them into a measurement-uncertainty statement (Joint Committee for Guides in Metrology, 2008). The VIM further distinguishes error, bias, precision, repeatability, reproducibility, and uncertainty (Joint Committee for Guides in Metrology, 2012b). These terms describe different properties and should not be substituted for one another.

An independent comparator used in a validation study is therefore better described as a reference measurement process than as error-free truth. Its calibration, resolution, repeatability, operator procedure, contact force, alignment, environmental conditions, and workpiece geometry can affect the assigned reference value. Che and Ratnam (2018) provide precedent for comparing vision-derived dimensions with digital-caliper measurements, but accessible evidence from that study does not establish a complete uncertainty treatment for the comparator. The literature supports comparison to an external method while also requiring caution about the reference method's limitations.

CaseMetrology's manual reference process and its uncertainty are described in Chapter 3. The literature does not supply the project-specific uncertainty value, nor does the uncertainty of the manual reference process constitute a complete uncertainty budget for the computer-vision system. The latter was characterized empirically in this study through agreement, error, repeated-placement, availability, and classification metrics.

2.7 Repeatability, Accuracy, and Measurement-System Agreement

2.7.1 Repeatability is not accuracy

The VIM defines measurement precision in terms of agreement among replicate measurements under specified conditions and defines repeatability as precision under repeatability conditions (Joint Committee for Guides in Metrology, 2012b). Measurement accuracy concerns closeness between a measured value and the true value of the measurand; it is

related to both trueness and precision and is not itself assigned a numerical quantity value in the VIM framework.

These definitions establish why a repeatable system is not necessarily accurate. A system may return tightly grouped repeated values that share a systematic offset or another stable error. Conversely, a system may show little average bias across a cohort while producing substantial specimen-level scatter. Bias, absolute error, and repeated-measurement dispersion therefore answer different engineering questions.

Coveney (2014) explicitly warns that strong day-to-day repeatability does not preclude measurement error arising from calibration, environment, equipment, procedure, or operator effects. Xiang et al. (2018) likewise reported bias and repeatability as separate performance characteristics. Hemming (2007) and Lazzari and Iuculano (2004) provide the broader metrological basis for evaluating multiple contributions rather than treating one precision statistic as proof of complete measurement quality. Figure 2.2 illustrates this distinction without plotting study observations.

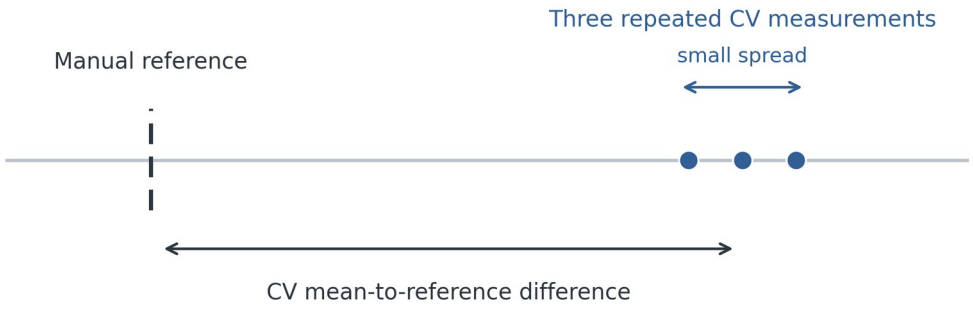


Figure 2.2. Conceptual distinction between repeatability and manual-reference agreement. Closely grouped CV results can differ from the reference result. Schematic; no CaseMetrology data are plotted. The reference result also has uncertainty.

This distinction is central to the final CaseMetrology architecture, but the literature review does not state the project's experimental outcome. It establishes the framework used later to interpret whether changes in repeated-presentation behavior were accompanied by corresponding changes in absolute agreement with the reference process.

2.7.2 Method agreement and validation design

Correlation is not a substitute for agreement between measurement methods. Bland and Altman (1986) showed that two methods can be strongly correlated while still differing systematically or by an amount that is unacceptable for the intended use. Difference-based analyses are therefore appropriate when evaluating whether an image-derived dimension agrees sufficiently with an independent reference.

A dimensional validation should also distinguish among measurement availability, mean bias, error magnitude, repeated-measurement variation, range-dependent behavior, and classification outcomes. No single metric answers all of these questions. For example, a low mean bias can coexist with large absolute errors of opposing sign, and perfect measurement availability does not imply acceptable dimensional accuracy.

Replication and randomization help characterize repeated observations and reduce confounding between acquisition order and uncontrolled temporal changes (NIST/SEMATECH, n.d.). The definition of replication is important: repeated processing of the same image evaluates algorithmic determinism, repeated images without moving the part evaluate short-term acquisition variation, and complete removal and replacement evaluate a broader presentation process. Repeatability values from different experimental conditions should not be compared as though they describe the same source of variation.

Sampling design also determines what can be inferred. Stratified sampling can intentionally increase representation in scientifically important regions, such as near specification limits, but results must be interpreted in relation to the design (Lohr, 2022). A cohort deliberately concentrated near decision boundaries is useful for challenging classification behavior; its raw acceptance fraction is not automatically an estimate of natural production yield.

2.8 Conformity Decisions, Guard Bands, and Screening

2.8.1 Decisions under measurement uncertainty

Dimensional specifications define permissible values of a product characteristic, whereas measurements provide uncertain information about those values. JCGM 106:2012 addresses conformity decisions made from measurement results and distinguishes tolerance limits from acceptance limits (Joint Committee for Guides in Metrology, 2012a). When an acceptance limit is displaced from a tolerance limit, the interval between them is a guard band. The purpose is to manage the probability and consequences of incorrect decisions, not to imply that measurement uncertainty has disappeared.

ISO 14253-1:2017 provides decision rules for verifying conformity or nonconformity of geometrical product specifications while taking measurement uncertainty into account, including cases in which measured values fall near specification limits (International Organization for Standardization, 2017). Together, JCGM 106 and ISO 14253-1 establish that direct comparison of a point estimate with a specification limit is only one possible decision rule.

Two incorrect outcomes are especially relevant. False acceptance occurs when a nonconforming item is accepted, while false rejection occurs when a conforming item is rejected. JCGM 106 relates these outcomes to consumer and producer risks and shows that changing the acceptance limits redistributes those risks. Narrowing the automatic-acceptance interval can

reduce false-acceptance risk, but it also increases the number of conforming items that are rejected or withheld from automatic acceptance.

A defensible guard band is therefore not simply the smallest margin that removes errors from one observed dataset. Its selection ordinarily depends on the measurement model, uncertainty, risk tolerance, consequences of each decision error, and the intended operating population. A margin selected retrospectively from a limited validation cohort can demonstrate observed decision behavior, but it cannot by itself establish a validated production decision rule or a general false-acceptance probability.

2.8.2 Guarded screening architectures

Guarded decisions can be implemented as either binary or multistage processes. In a binary process, a result within the acceptance interval is accepted and a result outside it is rejected. In a screening process, a third disposition is possible: measurements comfortably within the guarded interval may be accepted automatically, while near-limit measurements are referred to a more capable or independent measurement process. Measurements clearly outside specification may likewise be rejected, confirmed, or investigated according to the production procedure.

This multistage architecture separates screening from final reference measurement. Its value depends on the screened population, the performance of both measurement stages, the disposition of review-zone items, and the cost or consequence of incorrect decisions. The screening system still requires characterized measurement behavior because its acceptance and referral boundaries must be connected to empirical performance and risk.

Historical cartridge-case systems used optical measurements for automatic accept/reject decisions and incorporated calibration or known standards for system assurance (Coleman et al.,

1974; Coleman & Swinth, 1977; Swinth et al., 1979). Those systems demonstrate industrial precedent for integrating optical measurement with manufacturing decisions. JCGM 106:2012 and ISO 14253-1:2017 provide the modern metrological framework for understanding why decision limits and measurement uncertainty must be considered explicitly.

For the present thesis, this literature supports evaluation of guarded screening as an engineering pathway. It does not prevalidate any CaseMetrology-specific margin. The status, derivation, cohort dependence, and limitations of the project's exploratory guard-band analysis are addressed separately in Chapters 3 through 5.

2.9 Optical and Computational Analysis of Cartridge Cases

2.9.1 Historical manufacturing inspection

Electro-optical dimensional inspection of cartridge cases predates modern area-camera computer vision. Work at Pacific Northwest Laboratory and Battelle developed automatic high-speed cartridge-case inspection systems that combined mechanical handling, electro-optical profile measurement, computer processing, process monitoring, and rejection (Coleman et al., 1974; Coleman & Swinth, 1977; Swinth et al., 1979).

The systems measured several cartridge-case characteristics, including total length, head thickness, head diameter, extractor-groove diameter, and gas-seal length. Transmitted illumination created projected shadows that were analyzed with linear diode or CCD arrays. Known or calibrated standards could be inserted for system checks, and automatically measured cases could be compared with conventionally gauged values. The reported systems operated at manufacturing throughput and linked dimensional measurement to automated disposition.

These sources are the most direct prior art for the CaseMetrology application. They establish that optical measurement of cartridge-case total length, computer-assisted inspection,

calibration checks, comparison with conventional gauging, and automated classification were demonstrated decades before the present study.

The historical architecture differed materially from CaseMetrology. It used dedicated mechanical handling, precision gauging references, optical projection, fiber-optic image transfer, and linear sensor arrays. CaseMetrology uses a static two-dimensional area camera, embedded processing, software-defined silhouette measurement, and preserved raw images. Those differences define a distinct implementation and experimental study; they do not justify a first-ever claim for optical cartridge-case measurement.

2.9.2 Modern forensic imaging context

Modern cartridge-case imaging literature is dominated by forensic surface and topography analysis rather than manufacturing OAL measurement. NIST Standard Reference Material 2461 provides a standard cartridge case for quality assurance and performance evaluation of optical topography systems used in firearm identification (National Institute of Standards and Technology, n.d.). Other research has addressed automated segmentation of cartridge-case base regions and image matching of forensic marks (Sakarya et al., 2012; Yammen & Muneesawang, 2013).

This literature confirms that controlled optical acquisition and computational analysis of cartridge cases are established fields. The measurands and decisions, however, differ from dimensional manufacturing inspection. Surface-topography comparison, region-of-interest extraction, and forensic image matching do not establish OAL accuracy or production conformance performance. These sources therefore provide cartridge-specific optical context but are not substitutes for dimensional-metrology literature.

2.10 Literature Synthesis and Research Position

The literature supports six conclusions that frame the CaseMetrology study.

First, dimensional machine vision is a complete measurement-system problem. Image formation, illumination, optical geometry, boundary localization, calibration, mechanical relationships, specimen presentation, reference comparison, and decision logic can all affect the reported result (Hemming, 2007; Lazzari & Iuculano, 2004; Coveney, 2014).

Second, backlit silhouette measurement is an established method for manufacturing dimensions. Prior systems have combined transmitted illumination with edge or contour localization, geometric fitting, calibration, and external comparison (Shin et al., 2002; Xiang et al., 2018; Che & Ratnam, 2018). The research question is therefore not whether a silhouette can yield a dimension, but whether a specific system produces measurement behavior adequate for its intended role.

Third, physical presentation is part of the measurement architecture. Prior systems used fixtures, machine-tool constraints, servo motion, or dedicated handling, while optical-geometry research demonstrates why the relationship between the calibrated plane and the measured object matters (Shin et al., 2002; Wang et al., 2024). The literature justifies controlling and studying presentation, but the quantitative effect of that control in CaseMetrology remains a project-specific experimental result.

Tribology refines this premise: nominal contact geometry, friction, and changing surface condition are possible mechanisms linking a fixture to presentation stability (Greenwood & Williamson, 1966; Archard, 1953). Figure 2.1 locates these hypotheses within the measurement chain without assigning them a measured contribution to V3 performance.

Fourth, repeatability and absolute accuracy are distinct. Precision under repeated conditions does not prove closeness to the reference value, and low average bias does not prove small specimen-level error (Joint Committee for Guides in Metrology, 2012b; Coveney, 2014). A defensible evaluation therefore requires separate characterization of bias, absolute error, repeated-measurement behavior, availability, and decision outcomes.

Fifth, manufacturing decisions must account for uncertainty and risk near specification limits. JCGM 106:2012 and ISO 14253-1:2017 distinguish specification limits from acceptance limits and provide the basis for guarded decision rules. A guarded screening architecture can refer near-limit measurements for secondary inspection, but an empirically selected margin is not a validated production guard band without prospective evaluation and an appropriate risk basis.

Sixth, optical cartridge-case dimensional inspection is not novel. Historical systems measured total case length and other dimensions, used calibration and conventional-gauge comparisons, and performed automated rejection (Coleman et al., 1974; Coleman & Swinth, 1977; Swinth et al., 1979). The contribution of CaseMetrology must therefore be stated more narrowly.

CaseMetrology occupies a research space defined by the engineering evaluation of a modern embedded, static, backlit area-camera architecture for cartridge-case OAL. Its contribution lies in the documented progression from prospective baseline validation through diagnostic investigation, fixture and presentation control, estimator development, and prospective evaluation of the final frozen configuration. The research also evaluates the difference between repeated-presentation stability and absolute agreement and explores how measured limitations could be incorporated into a guarded screening process.

This position does not claim that CaseMetrology replaces precision contact metrology, that it is production-ready, or that its exploratory decision margins are validated for production. Instead, the literature supports the narrower question answered by the thesis: what a specific embedded dimensional-vision architecture can defensibly support under documented static conditions, what limitations remain, and how those limitations affect its potential role as an industrial screening tool.

Chapter 3 - Methodology

3.1 Research Design and Experimental Chronology

This study used a sequential engineering-validation design to evaluate a static computer-vision system for cartridge-case overall-length measurement. The work began with a prospectively specified formal validation of the original frozen system, followed by diagnostic investigation and two engineering iterations. The completed sequence was analyzed chronologically so that later findings were not represented as though they had been known or planned before the original validation.

The formal measurand was cartridge-case overall length (OAL), defined as the axial distance from the case base/head end to the case mouth end, as illustrated in Figure 3.1. The nominal specification interval used for research classification was 0.9900–1.0100 in, inclusive. All screening, manual reference, and formal computer-vision measurements were performed by one operator, Timmy D. Bell. The study therefore characterized single-operator performance under the documented conditions; it was not a multi-operator Gage R&R study.



Figure 3.1. Schematic definition of cartridge-case OAL used in the formal analysis. OAL was the axial distance from the case base/head end to the case mouth end; schematic not to scale.

Development observations and formal validation observations were kept analytically separate. Five specimens used during estimator and presentation development were excluded from formal sample eligibility. Formal datasets were not used to retune the configuration being evaluated, and an unfavorable result did not authorize a repeat measurement, specimen replacement, calibration change, estimator change, or relaxation of an acceptance criterion.

Table 3.1 summarizes the status of each phase. The distinction between prospective and retrospective work is central to the methodology.

Table 3.1. Experimental phases and inferential status

Phase	Configuration or activity	Primary purpose	Status
V1	PCA estimator; independent free placement	Establish the original formal performance baseline	Prospective formal validation
Post-V1 diagnostics	Ambient-light, controlled-repositioning, and fixture investigations	Investigate V1 variability and generate engineering hypotheses	Exploratory diagnostic work
V2	Revised mechanical fixture; PCA estimator retained	Evaluate the fixture change on the retained formal cohort	Prospective formal engineering iteration
Endface-Hybrid development	Geometry-derived axis with subpixel end-face localization	Develop an alternative endpoint estimator using excluded development data	Development work
V2 Endface-Hybrid analysis	Frozen Endface-Hybrid applied to previously acquired V2 images	Compare estimator behavior on a common formal image set	Retrospective exploratory analysis
Controlled-presentation development	Rocking control, fixed roll orientation, and fixture-retention verification	Define the V3 presentation procedure	Development and preacquisition verification
V3	Frozen Endface-Hybrid; controlled presentation	Evaluate the final frozen configuration	Prospective formal validation
Guard-band analysis	Symmetric guarded acceptance intervals applied to V3 specimen means	Explore dimensional-screening decision behavior	Retrospective exploratory analysis

The same frozen 60-specimen cohort and the same accepted manual-reference means were retained for V1, V2, and V3. This supported paired engineering comparisons while avoiding a change in specimen composition or reference assignment between iterations. Results from each phase are reported in Chapter 4; the present chapter describes how those results were generated and classified.

3.2 CaseMetrology System Architecture

3.2.1 Embedded computing and imaging hardware

CaseMetrology was implemented as a static backlit vision system. Image acquisition, segmentation, calibration lookup, and dimensional computation were executed on an NVIDIA Jetson Orin Nano Super Developer Kit. The documented study environment used NVIDIA L4T R39.2 with kernel 6.8.12-1021-tegra, Python 3.12.3, OpenCV 4.6.0, and NumPy 1.26.4. The Jetson operated in the documented 15 W power mode.

The imaging device was an Arducam B0499 USB 3.0 camera based on the Sony IMX296 global-shutter sensor. The project camera was identified by serial number Arducam_20231211_0001 and accessed through the persistent device path /dev/case-metrology-camera. Images were acquired at 1456 × 1088 pixels in YUYV format at 44 frames/s after a 20-frame warm-up. The installed optic was an Arducam 16 mm CS-mount lens, model CS2316ZM02. Repository history and a nondisturbing physical inspection supported continuity of this 16 mm lens through the validation sequence; no evidence supported the earlier recollection that a 25 mm lens had been used.

The camera-to-measurement-plane distance was 8.625 in, measured from the front edge of the lens barrel to the surface on which the reference artifact or specimen rested. A clear Lexan sheet with nominal thickness 0.250 in served as the working surface and measurement plane. Transmitted illumination was provided by a VALIKIY LED backlight sheet rated at 24 V and 36 W within the measurement enclosure.

The locked camera-control values were brightness 0, contrast 10, saturation 0, automatic white balance disabled, white-balance temperature 2498, gain 100, manual exposure mode, and exposure-time value 5. The acquisition software checked these controls against the expected

values. The camera, lens, focus, backlight, and calibrated working geometry were not intentionally changed across V1, V2, and V3. Figure 3.2 documents the physical system and specimen presentation.

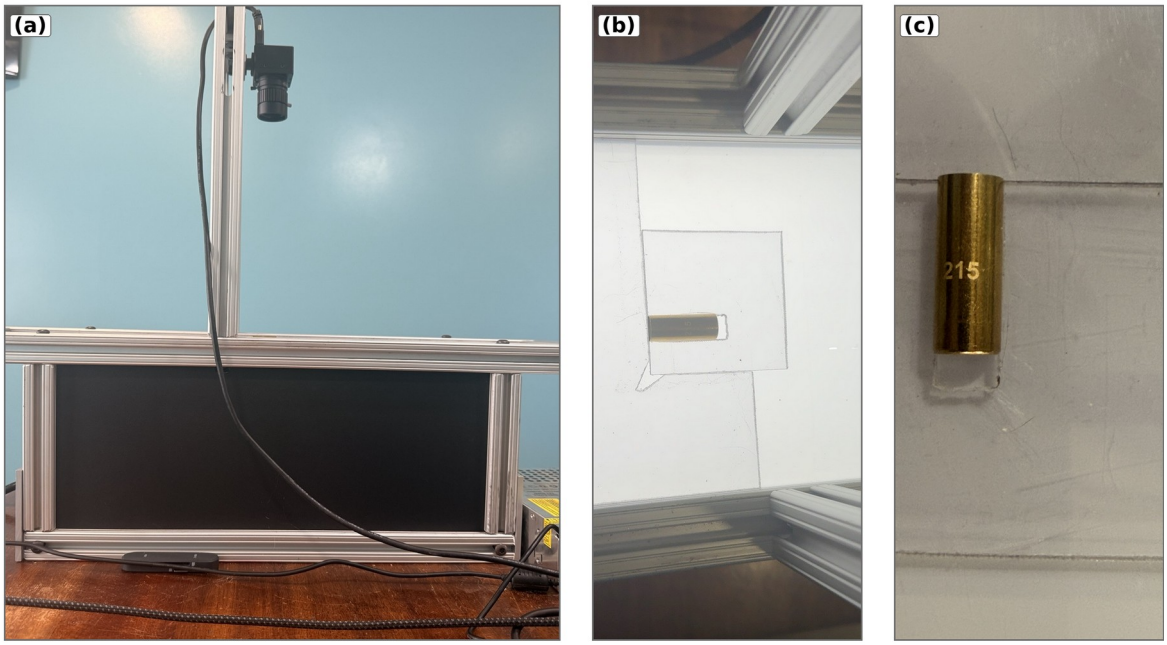


Figure 3.2. CaseMetrology physical system and specimen presentation: (a) camera, support frame, and backlight enclosure; (b) illuminated working surface; and (c) specimen seated between the clear Lexan fixtures. Photographs by the author.

3.2.2 Image acquisition and segmentation

Each formal computer-vision session began with a preserved blank-reference image acquired without a specimen. A specimen image was converted to grayscale and compared with the session blank using blank-minus-specimen subtraction so that regions darkened by the case silhouette were emphasized. The difference image was smoothed with a 5 × 5 Gaussian kernel, thresholded using Otsu's method, and processed with 3 × 3 morphological opening and closing (Otsu, 1979; Gonzalez & Woods, 2018). Connected components were then evaluated, and the largest foreground component was retained as the candidate case silhouette.

Segmentation generated the binary mask, component centroid, and quality information required by the measurement estimators. Thresholds, morphology settings, component selection, and quality-gate logic were not adjusted for individual formal specimens. The same segmentation and calibration lineage was used when multiple estimators were applied to a common image.

3.2.3 PCA OAL estimator

The primary V1 and V2 estimator was the frozen principal component analysis (PCA) and extreme-mask method implemented in src/measurement/topdown_oal.py. PCA provided an initial longitudinal direction for the foreground point set (Jolliffe & Cadima, 2016). The implementation refined the case axis from exposed silhouette points and projected accepted mask points onto that axis. The pixel-space OAL was the separation between the minimum and maximum accepted projections. Base and mouth identity followed specimen geometry rather than being inferred only from the sign of the projection.

The estimator returned the pixel length, centroid, measurement-axis angle, endpoint coordinates, component information, and pass/fail status. Its source identity and the segmentation implementation were frozen before V1 formal acquisition. The PCA estimator remained unchanged for the prospective V2 fixture evaluation and was retained as a historical comparator for V3.

3.2.4 Endface-Hybrid OAL estimator

The Endface-Hybrid estimator combined a geometry-derived longitudinal axis with grayscale subpixel localization at the two OAL end faces. It did not apply subpixel refinement to the case sidewalls. The binary geometry estimator supplied the axis, end seeds, centroid, case

width, and measurement angle. At each end, 41 transverse seed locations were distributed across a centered span equal to 60% of the estimated case width.

For each seed, grayscale intensity was sampled along the longitudinal direction from -6 to +6 pixels in 0.25-pixel increments. Interior and exterior plateau intensities were estimated from the prescribed sampling intervals, and a profile was accepted only when it met the frozen local-contrast, anchor-distance, and unique-crossing rules. The edge location was the 0.50 normalized-intensity crossing. At least 12 accepted profiles were required at each end. Accepted crossing points were fit using iteratively reweighted linear least squares, and pixel OAL was the axial separation between the fitted base and mouth intercepts.

The frozen numerical parameters included a minimum local contrast of 30 grayscale units and maximum anchor and crossing distances of 3 pixels. The estimator converted its pixel result to inches using the same spatial and directional calibration used by the PCA pipeline. The frozen implementation was scripts/evaluate_oal_endface_hybrid_dev.py, with SHA-256 3f332ef097f5f39e6e1447a61756c7ecc4c8e7c8170fe8933ab362daffb2a2af.

3.2.5 Configuration control and data preservation

Measurement-affecting source code, protocols, run orders, calibration records, and formal analysis scripts were maintained under Git version control. Raw images, blank references, acquisition metadata, observation tables, and derived outputs were preserved separately. SHA-256 manifests were used for the V3 measurement-system freeze and selected research artifacts. Version control and hashing established configuration identity and research chronology; they were not treated as substitutes for empirical measurement validation.

Formal records distinguished VALID, SYSTEM_FAILURE, and PROTOCOL_INVALID events. A system failure was an otherwise valid attempt in which the

frozen pipeline failed to return an accepted measurement; it remained in the availability denominator and could not be repeated merely to obtain a favorable result. A protocol-invalid event was an external procedural failure, such as an incorrect specimen or acquisition condition. Its original record was retained, and a linked replacement was permitted only under the frozen deviation procedure. No missing result was imputed.

Figure 3.3 summarizes the two measurement paths. A case seated on the Lexan plane and illuminated from behind produced the image used by the frozen CV pipeline; gauge-block measurements supplied its spatial and directional scale. The independently acquired caliper result supplied the specimen-level comparison reference. Fixture contact is part of the presentation conditions, but this study did not record normal force, friction, contact patch geometry, wear, or surface transfer.

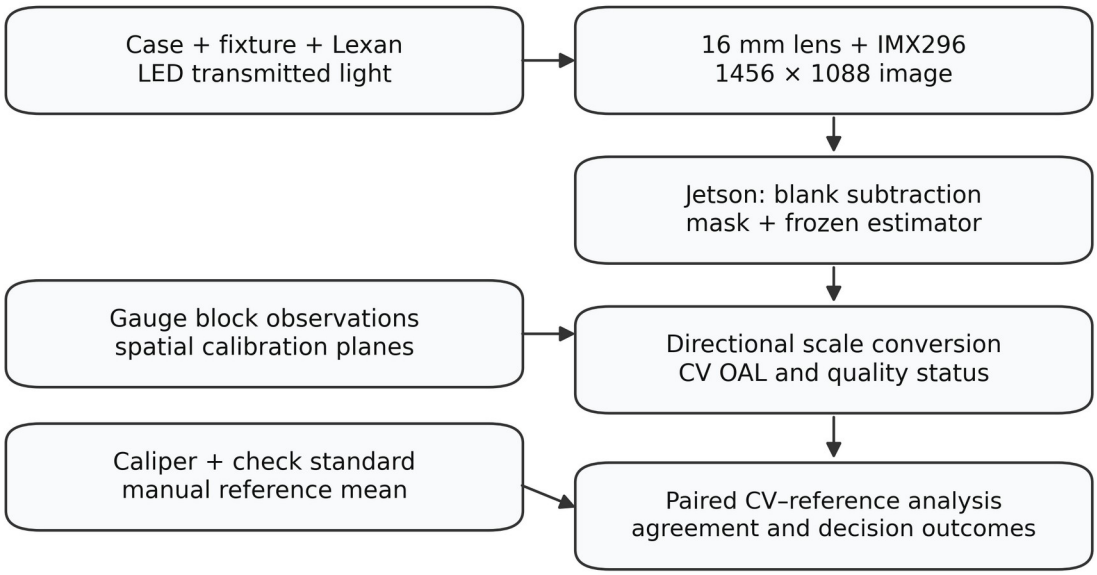


Figure 3.3. Conceptual CaseMetrology measurement architecture. The upper path produces CV OAL; gauge-block calibration and the separate caliper reference inform conversion and comparison, respectively. Arrows indicate information flow, not a quantified uncertainty budget.

3.3 Calibration and Manual Reference Measurement

3.3.1 Spatial and directional calibration

Image-space measurements were converted to physical units using a Mitutoyo Grade 0 steel gauge block, identification number 253568. Certificate T26A06624 assigned a central length of 1.0000024 in with expanded uncertainty of 0.0000024 in at coverage factor k = 2. Calibration characterization was performed on the installed measurement plane with the permanent Lexan assembly in place.

A coordinate-controlled 2 × 2 grid at positions A through D was measured in two principal gauge-block orientations, with three images acquired at each position-orientation combination. A fifth location near the center of the grid was withheld from fitting and used for principal-direction and oblique holdout checks. The calibration observations showed that effective scale varied with both image position and measurement direction, so one global pixels-per-inch value was not used.

Separate first-order spatial planes were fit to the accepted A–D observations:

$$PPI_0(x, y) = 516.907045 + 0.00843326411\,x + 0.000817091863\,y,$$

and

$$PPI_{90}(x, y) = 518.408825 + 0.00322556849\,x + 0.00442714650\,y.$$

The labels PPI_0 and PPI_{90} identify the historical gauge-block acquisition orientations. A prevalidation audit found that the gauge-block analyzer reported the block's long-axis orientation while measuring the certified short dimension perpendicular to that axis. The cartridge-case estimator reported the OAL measurement direction itself. The direction mapping was therefore corrected before formal validation without refitting the accepted planes.

For a segmented specimen centroid (x_c, y_c) and OAL direction θ, the frozen effective scale was

$$PPI_{eff} = \sqrt{\left[PPI_{90}(x_c, y_c)\cos\theta\right]^2 + \left[PPI_0(x_c, y_c)\sin\theta\right]^2}.$$

Physical OAL was then

$$L_{OAL} = \frac{L_{px}}{PPI_{eff}},$$

where L_{px} was the estimator's pixel-space length. Under this convention, a horizontal OAL direction resolved to the historical PPI_{90} surface and a vertical direction resolved to PPI_0. The corrected oblique holdout had an equivalent residual of approximately -0.000413 in over the approximately 1-in reference. The calibration therefore supported use within the tested region while retaining the limitation that the conventional optical system was not demonstrated to have universal sub-thousandth-inch accuracy at arbitrary positions and orientations.

3.3.2 Manual reference instrument and session assurance

The formal reference instrument was a Mitutoyo digital caliper, model CD-6”ASX, code 500-196-30, serial number B24254849, with 0.0005-in display resolution. Calibration certificate DDJ25P23P reported an as-left indication error of 0.0000 in at 1 in and expanded uncertainty of 0.000126 in at approximately 95% confidence. No numerical correction was applied to formal case measurements because the reported as-left error at 1 in was zero.

The Grade 0 gauge block served as a session-assurance standard rather than as a direct calibration correction for each specimen. Before and after every formal manual session, the operator obtained five independent gauge-block readings, completely removing and replacing the block between readings. The prospectively frozen inclusive limits were absolute mean indication error no greater than 0.0005 in relative to the expected 1.0000-in display, five-reading range no

greater than 0.0005 in, and pre-to-post mean drift no greater than 0.0005 in. A failed precheck prevented specimen measurement; a failed postcheck placed the session on hold for review.

The measurement equipment was allowed to equilibrate in a consistent indoor workspace. Room and gauge-block surface temperatures were recorded at the beginning and end of formal manual sessions using a Cole-Parmer 20250-91 thermometer, serial number 2605139793, under calibration report 1817067, with Type-T surface probes. The calibration report applied to the meter using simulated thermocouple inputs; the individual surface probes were not separately calibrated. The temperature records therefore documented stability and abnormalities, but no numerical thermal correction or environmental uncertainty component was introduced.

3.3.3 Formal manual measurement procedure

The 60 selected specimens were measured in three independently randomized rounds, one reading per specimen per round, using seed 2026081901. The three sessions were identified as MAN-S001, MAN-S002, and MAN-S003. For each observation, the operator verified the laser-etched specimen ID, placed the case lengthwise between the external caliper faces, established axial contact between the base/head and mouth using only sufficient force for gentle contact, recorded the stabilized display exactly, and completely removed the case before proceeding.

Screening values and previous formal readings were not consulted during acquisition. A reading was not repeated because it differed from screening, another replicate, the nominal specification, or an expectation. Two out-of-sequence readings in MAN-S002 were preserved as protocol-invalid records and replaced under the documented deviation procedure. The accepted dataset contained 180 displayed readings, and the invalid records remained in the audit trail.

For specimen i, the formal reference mean was

$$\overline{M}_i = \frac{M_{i1} + M_{i2} + M_{i3}}{3}.$$

These accepted three-round means were frozen after V1 and reused for the paired V2 and V3 evaluations. Reusing the same reference values preserved the intended V1-to-V3 comparison. A planned V3 manual stability-check round was removed by Protocol Amendment 001 before V3 acquisition; no new V3 manual readings replaced the accepted reference means.

3.3.4 Reference uncertainty and near-limit classification

The reference process was treated as a calibrated comparator with finite uncertainty, not as error-free truth. The standard calibration contribution was

$$u_{cal} = 0.000126 / 2 = 0.000063 \text{ in},$$

and the digital-resolution contribution was

$$u_{res} = 0.0005 / \sqrt{12}.$$

The pooled within-specimen manual standard deviation was calculated from the 60 triplicate sets:

$$s_p = \sqrt{\frac{\sum_i \sum_j \left(M_{ij} - \overline{M}_i\right)^2}{60\left(3 - 1\right)}}.$$

The repeated-measurement contribution to a three-reading specimen mean was

$$u_{rep} = \frac{s_p}{\sqrt{3}}.$$

Combined standard and expanded reference uncertainties were

$$u_{ref} = \sqrt{u_{cal}^2 + u_{res}^2 + u_{rep}^2}$$

and

$$U_{ref} = 2\,u_{ref}.$$

The primary reference classification used the point estimate $\overline{M}_i$ and the inclusive 0.9900–1.0100 in specification limits. A secondary descriptive flag identified a specimen when its expanded reference-uncertainty interval reached or crossed either limit. This flag did not reclassify the specimen, create an indeterminate primary category, authorize replacement, or impose a guard band.

3.4 Specimen Population, Screening, and Frozen Cohort

The source population comprised 235 cartridge cases laser etched with unique identifiers 1 through 235. The etched integer was used as part_id throughout the study. The population included cases distributed across the available OAL range, including cases shorter and longer than the nominal interval.

Specimens 121, 146, 178, 187, and 197 had been used during estimator or presentation development and were excluded from formal eligibility. The remaining 230 cases were screened using the designated manual process to characterize the OAL distribution for stratification. Screening observations were not included in the formal three-reading reference means.

The formal cohort was selected with a prospectively frozen, intentionally non-proportional stratified design. Within each stratum, eligible specimens were ranked using SHA-256 applied to the fixed string containing the project name, seed 2026090301, stratum, and specimen ID. The highest-ranked eligible cases were selected according to the frozen allocation shown in Table 3.2.

Table 3.2. Frozen formal-cohort allocation

Stratum	Screening definition	Eligible	Selected
Below LSL	$L < 0.9900$ in	10	8
Near LSL, in specification	$0.9900 \le L \le 0.9950$ in	34	14
Central nominal	$0.9950 < L < 1.0050$ in	137	16

Stratum	Screening definition	Eligible	Selected
Near USL, in specification	$1.0050 \le L \le 1.0100$ in	24	14
Above USL	$L > 1.0100$ in	25	8
Total		**230**	**60**

The deliberately enriched near-limit and out-of-specification strata increased the study's ability to challenge dimensional agreement and classification behavior. Because the selection fractions differed by stratum, unweighted acceptance and referral fractions from this cohort were not interpreted as production-yield estimates (Lohr, 2022).

The 60-specimen cohort was frozen before formal manual or CV acquisition and retained without performance-based replacement through V1, V2, and V3. Each formal CV phase used three independently randomized rounds containing all 60 specimens exactly once. Replicates were used to characterize repeated presentation and to form specimen means; they were not treated as 180 independent specimens.

3.5 V1 Prospective Formal Validation

3.5.1 Frozen V1 configuration and acquisition

V1 evaluated the original frozen OAL measurement system with the PCA/extreme-mask estimator. Formal CV acquisition followed completion of the manual reference phase. The operator did not consult formal manual readings, specimen means, screening values, sampling strata, or CV-versus-reference errors while acquiring CV observations.

The CV run order was generated independently of the manual order using seed 2026081902. Three sessions, CV-S001 through CV-S003, each contained all 60 specimens once. A separate blank-reference image was preserved for each session. For every scheduled observation, the operator verified the specimen ID, completely removed the case from the

measurement surface, placed it for a single acquisition, preserved the resulting image and metadata, and removed it before the next observation.

The physical placement condition used in V1 was documented after acquisition to resolve ambiguity in the earlier fixture description. The clear Lexan working surface was present, but no rails, stops, pockets, pins, or other locating features constrained case position. Cases were laid freely within the field of view with the long axis approximately vertical and with a consistent gross base-to-mouth direction. Exact image-plane location, small in-plane angular differences, roll orientation, and possible base-to-mouth rocking were not mechanically controlled. V1 repeatability therefore represented independent free placement rather than repeatability within a dedicated locating fixture.

Each scheduled position permitted one formal attempt. A surprising numerical result, pipeline rejection, or disagreement with another reading did not authorize a retake. Valid observations, failures, raw images, blank linkage, diagnostic metadata, effective scale, software revision, and protocol status were preserved.

3.5.2 Specimen-level analysis

The unique specimen was the primary unit for agreement analysis. For specimen i, the formal CV mean was

$$\overline{C}_i = \frac{C_{i1} + C_{i2} + C_{i3}}{3},$$

and signed specimen error was

$$e_i = \overline{C}_i - \overline{M}_i.$$

Positive error indicated that the CV specimen mean exceeded the manual-reference mean. Mean bias, mean absolute error, and root mean squared error were defined as

$$Bias = \bar{e} = \frac{1}{n}\sum_{i=1}^{n} e_i,$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \left|e_i\right),$$

and

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2}.$$

The empirical 95th percentile used the frozen nearest-rank definition. For $n = 60$, P95 absolute specimen error was the 57th ordered value of $\left|e_i\right)$. Maximum absolute error was the largest observed $\left|e_i\right)$. The two-sided 95% confidence interval for mean bias used the Student's *t* distribution:

$$\bar{e} \pm t_{0.975,n-1}\frac{s_e}{\sqrt{n}},$$

where s_e was the sample standard deviation of specimen errors. Difference-versus-magnitude and Bland–Altman analyses were also used to examine agreement. The descriptive 95% limits of agreement were $\bar{e} \pm 1.96\, s_e$. Correlation and regression, when reported, were secondary and did not replace direct error analysis (Bland & Altman, 1986).

3.5.3 Repositioning, availability, and classification

Repeated-presentation behavior was evaluated independently of specimen-level agreement. Three valid measurements generated three absolute pairwise differences per specimen:

$$d_{i,12} = \left|C_{i1} - C_{i2}\right), d_{i,13} = \left|C_{i1} - C_{i3}\right), d_{i,23} = \left|C_{i2} - C_{i3}\right).$$

For 60 complete specimens, the pooled repositioning distribution contained 180 pairwise differences. Its empirical P95 was the 171st ordered value. Specimen-level ranges and standard

deviations and the pooled median, mean, P95, and maximum pairwise differences were also calculated. The three pairwise differences from each specimen were treated as related descriptive values, not as independent inferential observations. This definition was retained for V2 and V3 so that repeated-presentation results were comparable.

Attempt-level measurement availability was

$$Availability = \frac{N_{VALID}}{N_{VALID} + N_{SYSTEM_FAILURE}}.$$

Protocol-invalid events were preserved but were not counted as validly executed measurement-system attempts. No failed measurement was imputed.

Both manual-reference and CV specimen means were classified with the inclusive 0.9900-in LSL and 1.0100-in USL. A false pass occurred when the CV mean was in specification and the manual-reference mean was not. A false fail occurred when the CV mean was outside specification and the manual-reference mean was in specification. The confusion matrix, sensitivity for detecting nonconforming specimens, specificity for conforming specimens, and identities of misclassified cases were retained.

3.5.4 Prospectively frozen performance criteria

The criteria in Table 3.3 were specified before V1 formal collection and were not relaxed after results were observed. The same quantitative targets were retained for the prospective V3 evaluation. Zero false passes was a preferred decision outcome rather than one of the six quantitative performance thresholds.

Table 3.3. Prospectively frozen OAL performance criteria

Endpoint	Target
Valid CV measurement availability	≥99%
Absolute mean bias	≤0.0010 in
RMSE	≤0.0020 in

Endpoint	Target
P95 absolute specimen error	≤0.0030 in
Maximum absolute specimen error	≤0.0040 in
P95 pairwise repositioning difference	≤0.0020 in
False passes	Zero preferred

Meeting availability or repeatability targets did not imply meeting absolute-agreement criteria. Conversely, failure of a criterion was retained as a study result and did not authorize validation-set tuning. Numerical results and criterion dispositions are reported in Chapter 4.

3.6 Post-V1 Diagnostic Investigation

V1 was closed and preserved before diagnostic work began. The post-V1 program was designed to investigate plausible contributors to the observed measurement variation and to guide engineering changes. These studies used excluded development specimens or non-formal images and did not replace, correct, or reclassify any V1 observation.

3.6.1 Ambient-light diagnostic

An ambient-light diagnostic used excluded specimen 121 while the V1 camera, lens, backlight, working surface, controls, calibration, segmentation, and estimator remained unchanged. Blank images were acquired under four ordered conditions: window blocked and ceiling light off; window exposed and ceiling light off; window blocked and ceiling light on; and a return to the initial blocked-window, light-off condition. Five blank images were acquired per condition.

The case was then placed once and left physically untouched while ten specimen images were collected under each of the same four lighting conditions. Blank-image intensity and fixed-specimen OAL were compared across conditions. Because the specimen remained unmoved, any repeatable change could be attributed to acquisition or illumination state rather than deliberate

repositioning. The diagnostic was not designed to prove or disprove the lighting conditions that existed during V1.

3.6.2 Controlled-repositioning diagnostic

A separate controlled-repositioning study used specimen 121 under blocked-window and ceiling-light-off conditions with one frozen blank reference. The case was removed and reseated ten times using the existing locator's natural seated position. Three consecutive images were acquired without touching the case after each seating. The design separated short-term frame-to-frame variation within a placement from variation among independent placements. No estimator, segmentation, calibration, exposure, focus, or working-geometry parameter was changed.

3.6.3 Fixture and presentation diagnostics

Subsequent fixture studies examined whether mechanical registration, base-to-mouth rocking, and roll orientation contributed to repeated-presentation variation. These investigations used development specimens 121, 146, 178, 187, and 197. The specimen was removed and replaced between placements, and fixed-position image sequences were used where needed to separate acquisition stability from physical presentation. Later controlled-presentation trials held roll orientation constant by orienting the engraved identifier toward the camera and constrained base-to-mouth rocking.

The diagnostic sequence was interpreted as engineering evidence rather than confirmatory validation. Its role was to identify a defensible change for the next frozen configuration. Numerical diagnostic results are reported separately in Chapter 4.

A contact-related explanation remained a hypothesis. The development trials changed how cases seated and rolled, but did not manipulate surface roughness, contact force, material

pairing, lubrication, or wear as independent factors. The observed improvement therefore cannot be attributed to a specific tribological mechanism.

3.7 V2 Fixture Engineering Iteration and Formal Evaluation

V2 prospectively evaluated whether improved mechanical registration reduced the placement-dependent variation observed in V1. The sole intentional system change was the specimen fixture. Fixture 1 provided a 90-degree mouth stop constrained by the frame rails. Fixture 2 was a rebuilt U-shaped lateral locator; its left side was shortened by approximately 0.505 in, and its groove was narrowed and hand-fitted so cases rested on the unchanged base Lexan without binding. For an accepted placement, the case rested on the base Lexan, contacted the Fixture 1 mouth stop, and occupied its natural constrained position within Fixture 2.

The Arducam IMX296 camera, 16 mm lens, mounting geometry, backlight, Lexan measurement plane, camera controls, spatial calibration, segmentation, PCA estimator, physical-unit conversion, and analysis definitions were inherited unchanged from V1. The same 60 formal specimens and accepted V1 manual-reference means were retained. Development observations were excluded from the formal V2 dataset.

V2 comprised three independently randomized rounds of 60 placements, for 180 planned observations. The V2 run order used seed 2026090901. Each observation required complete removal and independent reseating before one image acquisition, and each session used its own preserved blank reference. An unexpected OAL did not authorize reseating or reacquisition. The PCA pipeline was the prospective formal V2 estimator.

V2 used the same specimen-level agreement, pooled pairwise repositioning, availability, classification, and error definitions used for V1. Because the specimen cohort and manual reference were unchanged, V1 and V2 could be compared on a paired specimen basis. This

comparison evaluated the combined consequence of the revised physical registration under the retained PCA measurement chain; it did not treat the fixture change as proof that any particular physical error mechanism had been eliminated.

3.8 Endface-Hybrid Development and Retrospective V2 Evaluation

The Endface-Hybrid estimator was developed after V2 using only non-formal diagnostic images, including images of specimens excluded from the formal cohort. Its purpose was to test whether retaining a geometry-derived axis while applying subpixel localization only to the two OAL end faces could reduce endpoint-related variability without changing the optical hardware or calibration. The estimator architecture and numerical parameters described in Section 3.2.4 were frozen before any Endface-Hybrid result was generated from the V2 formal images.

The estimator freeze identified the source implementation and SHA-256 hash and prohibited any architecture or parameter change in response to the retrospective V2 results. The complete V2 raw-image set was first regenerated through the frozen PCA pipeline as an integrity check against the recorded V2 PCA values. The frozen Endface-Hybrid estimator was then applied to the same 180 previously acquired V2 images. Availability, specimen means, agreement errors, pairwise repositioning, and classification were recalculated using the original V2 run structure and the same manual reference.

This analysis was retrospective because V2 images already existed when the Endface-Hybrid estimator was finalized. It provided comparative evidence and informed the decision to proceed to a later prospective evaluation, but it was not an independent prospective validation of the estimator. Formal V2 remained the PCA evaluation defined by the V2 freeze. No V2 image, specimen, or measurement result was used to retune the frozen Hybrid estimator after the retrospective analysis.

Geometry and other experimental estimators were retained as secondary diagnostic comparators. Their outputs were not substituted for the prospective V2 PCA outcome.

3.9 V3 Controlled-Presentation Development, Freeze, and Prospective Validation

3.9.1 Development boundary and controlled-presentation rule

Post-V2 development examined presentation more directly after estimator-only investigation indicated that a substantial residual was shared across endpoint methods. Controlled-presentation trials were conducted with excluded specimens 121, 146, 178, 187, and 197. These trials constrained base-to-mouth rocking and registered roll orientation by placing the laser engraving toward the camera. The work was completed before V3 formal acquisition and was not included in the V3 performance dataset.

The final V3 presentation procedure required the operator to remove the case completely, replace it independently, seat it against the established fixture datum, mechanically control base-to-mouth rocking, orient the engraving toward the camera, confirm seating, acquire one image, and remove the case. The procedure was not changed in response to measurement outcomes.

Fixture 2 was retained against the established left datum using a friction-based holding method located outside the camera field of view. Protocol Amendment 002 prospectively documented this method. Before formal acquisition, a blank image verified that the retention mechanism did not intrude into the field of view, and excluded specimen 178 was used for five independent presentation checks with the planned V3 procedure. This activity was a mechanical verification only and was excluded from formal analysis.

3.9.2 Final V3 freeze

The end of controlled-presentation development was recorded at baseline commit cdd76a7. The prospective V3 configuration was then frozen at configuration commit e5e634d. The final evidence identified Endface-Hybrid as the primary estimator, PCA as the historical comparator, and Geometry as a secondary exploratory estimator. The measurement-system manifest preserved hashes for the cohort, run orders, plan, acquisition scripts, estimator source, segmentation, calibration, camera-control code, and comparator estimators.

The preacquisition audit verified a 60-specimen cohort identical to the established formal cohort, absence of the five development specimens, the expected stratum allocation, and three complete 60-specimen randomized run orders. The primary-estimator source hash was 3f332ef097f5f39e6e1447a61756c7ecc4c8e7c8170fe8933ab362daffb2a2af. No estimator, calibration, segmentation, camera, lens, lighting, or presentation change was permitted after final freeze without formal disposition.

The original V3 plan included a randomized manual stability-check round. Protocol Amendment 001 removed that round before any V3 formal measurement was collected. The accepted V1 three-round manual-reference means remained the frozen reference values. This amendment did not alter the cohort, CV run orders, presentation, estimator hierarchy, or performance targets.

3.9.3 Capture-only formal acquisition

V3 used three independently randomized sessions, V3-CV-S001 through V3-CV-S003, generated with seed 2026091202. Each session contained every formal specimen once, for 180 planned placements. A separate session blank was acquired under the locked camera-control procedure.

The V3 runner used a capture-only architecture. Before each image, it checked the expected round, sequence, specimen ID, existence of the session blank, completion of preceding positions, frozen manifest, and camera controls. It preserved one raw image and its metadata but did not execute or display an OAL estimator. The operator therefore had no Endface-Hybrid, PCA, or Geometry value available to influence seating, repetition, or reacquisition. All three rounds had to be completed before estimator processing was allowed.

During V3 Round 2, sequence 13 was initially captured under the scheduled label for specimen 35 while the wrong specimen had been placed. Acquisition stopped when the mismatch was recognized. The mislabeled capture and metadata were preserved in the protocol-invalid record and excluded from the formal observation table. The correct scheduled specimen 35 was then acquired under the documented replacement rule. No unfavorable numerical result prompted the replacement because no estimator output was displayed during acquisition.

After collection, the run-order and observation tables were audited for round coverage, identity, file linkage, and uniqueness. The formal dataset contained three scheduled images for each of the 60 specimens. The preserved protocol-invalid capture remained outside those 180 validly executed formal positions.

3.9.4 Post-acquisition estimator processing and primary analysis

Only after all 180 formal images had been acquired were the frozen estimators applied. Endface-Hybrid was the primary V3 estimator. PCA was processed on the same images to support historical and presentation comparisons, and Geometry was processed as a secondary estimator. Per-observation JSON diagnostics and the V3 observation table preserved estimator validity, pixel and inch results, effective scale, accepted end-profile counts, image linkage, and session information.

The primary experimental unit remained the specimen. Each specimen's three Endface-Hybrid results were averaged and compared with its frozen manual-reference mean. The agreement, P95 absolute-error, maximum-error, availability, pairwise-repositioning, and classification definitions in Section 3.5 were retained. V3 therefore evaluated the frozen Endface-Hybrid system prospectively under controlled presentation, even though the estimator calculations were intentionally deferred until acquisition was complete.

The prospective V3 targets were those in Table 3.3. The V3 plan also precluded introducing a post-hoc guard band into the primary formal classification. Failure of any target, false pass, false fail, estimator disagreement, or unusual specimen geometry was a reportable result rather than an acquisition stop condition. Hardware movement, configuration change, failed required verification, loss of traceability, or inability to reproduce the frozen presentation procedure required a stop or hold.

3.9.5 Comparative analysis boundaries

The completed chronology permitted several comparisons, but their evidentiary status differed. V1 PCA and V2 PCA were prospective evaluations of their respective physical configurations. V2 Endface-Hybrid was a retrospective estimator comparison using preserved V2 images. V3 Endface-Hybrid was the prospective primary evaluation of the frozen estimator and controlled presentation. V3 PCA and Geometry were same-image comparator analyses rather than separate acquisition experiments.

Paired comparisons used the same 60 specimens and the same reference means. Changes in repeated-presentation metrics were interpreted in relation to the physical presentation and estimator combination used in each phase. Absolute agreement and repeatability were analyzed separately; improvement in one was not treated as proof of improvement in the other.

3.10 Exploratory Guard-Band Analysis

The guard-band analysis was performed after V3 formal results were available. It was therefore retrospective and exploratory and did not alter the primary V3 confusion matrix, performance targets, estimator, reference classifications, or specimen measurements.

For a symmetric guard g, the automatic-acceptance interval was defined as

$$LSL + g \leq \overline{C}_i \leq USL - g,$$

where LSL = 0.9900 in and USL = 1.0100 in. A CV specimen mean inside this guarded interval was a candidate for automatic acceptance. A mean within the original specification interval but outside the guarded interval entered a near-limit review zone for secondary inspection. A mean outside the original specification interval was not automatically accepted and would require rejection, confirmation, or investigation under an eventual production procedure.

Candidate symmetric margins were applied to all 60 V3 specimen means. For each margin, the analysis recorded the number automatically accepted, the number routed to near-limit review, observed false passes among automatically accepted specimens, and conforming specimens not automatically accepted. The minimum observed margin that eliminated false passes in this cohort was also identified. A 0.0010-in margin was retained as a simple engineering example for thesis discussion.

The analysis did not estimate production yield or a general false-acceptance probability. The cohort was intentionally stratified around the specification limits, and the same cohort was used both to observe formal performance and to examine candidate margins. Consequently, any margin selected from these data remained a cohort-dependent feasibility result. A production decision rule would require a prospectively specified risk basis, an appropriate uncertainty

model, and independent validation on the intended operating population (Joint Committee for Guides in Metrology, 2012a; International Organization for Standardization, 2017).

The methodology supports a three-way screening interpretation—automatic acceptance, secondary review, and nonacceptance—but does not claim that the demonstrated prototype was production-ready, interchangeable with precision contact metrology, validated for conveyor operation, or validated with a production guard band. Those claim boundaries are retained in the interpretation of Chapters 4 and 5.

Chapter 4 - Results

This chapter reports the OAL results in the order in which the evidence was generated. V1 and V2 were prospective evaluations of the PCA estimator under their respective physical configurations. The Endface-Hybrid evaluation on V2 images was retrospective. V3 was the prospective evaluation of the frozen Endface-Hybrid estimator under the controlled-presentation configuration. Guard-band results were generated after V3 and are reported separately as exploratory results.

The same frozen cohort of 60 specimens and the same accepted manual-reference means were used throughout V1, V2, and V3. Each prospective CV evaluation contained three independently placed observations per specimen, for 180 scheduled observations. The primary agreement unit was the specimen mean. Availability was evaluated at the observation level, and repositioning was evaluated from the three absolute pairwise differences generated by each specimen's three placements.

The manual-reference dataset contained 180 valid observations. All three manual sessions met the frozen session-acceptance criteria. The pooled within-specimen manual standard deviation was 0.000553 in, and the expanded reference uncertainty was 0.000712 in at $k = 2$. Thirteen specimens had reference-uncertainty intervals that overlapped a nominal specification limit. The overlap flag did not change the primary point-estimate classification.

4.1 Prospective V1 PCA Results

4.1.1 Availability, agreement, and repositioning

All 180 scheduled V1 PCA observations were valid, giving 100.000% availability. The 60 specimen-level errors had a mean bias of +0.001613 in with a 95% confidence interval from

+0.001222 to +0.002005 in. Table 4.1 reports the remaining agreement and repeated-presentation results.

Table 4.1. Prospective V1 PCA performance

Endpoint	V1 result
Valid observations	180/180
Availability	100.000%
Bias	+0.001613 in
MAE	0.001815 in
RMSE	0.002205 in
Nearest-rank P95 absolute specimen error	0.003960 in
Maximum absolute specimen error	0.004799 in
Nearest-rank P95 pairwise repositioning difference	0.004831 in
Maximum pairwise repositioning difference	0.007081 in

The largest V1 pairwise difference occurred for specimen 70. Its three PCA results were 1.003218, 1.010299, and 1.008805 in, producing the 0.007081-in maximum pairwise difference.

4.1.2 Classification and criteria

Classification used the inclusive 0.9900–1.0100 in limits for both the manual-reference mean and the CV specimen mean. V1 produced 37 true-conforming classifications, 14 true-nonconforming classifications, eight false fails, and one false pass. The false-pass specimen was 143. The false-fail specimens were 15, 35, 51, 104, 118, 135, 144, and 151. Sensitivity for detecting nonconforming specimens was 93.33%, and specificity for conforming specimens was 82.22%.

Table 4.2 compares the V1 results with the six prospectively frozen quantitative criteria. The zero-false-pass condition is listed separately because it was a preferred classification outcome rather than one of the six quantitative thresholds.

Table 4.2. Prospective V1 criterion results

Criterion	Observed	Target	Outcome
Availability	100.000%	$\geq 99\,\%$	Met

Criterion	Observed	Target	Outcome
Absolute mean bias	0.001613 in	≤ 0.0010 in	Not met
RMSE	0.002205 in	≤ 0.0020 in	Not met
P95 absolute specimen error	0.003960 in	≤ 0.0030 in	Not met
Maximum absolute specimen error	0.004799 in	≤ 0.0040 in	Not met
P95 pairwise repositioning difference	0.004831 in	≤ 0.0020 in	Not met
False passes, preferred outcome	1	0 preferred	Not met

V1 met one of the six quantitative criteria.

4.2 Post-V1 Diagnostic Results

The post-V1 studies used excluded development specimens and did not modify any V1 result. The ambient-light study held specimen 121 fixed while four lighting states were applied. Table 4.3 gives the fixed-specimen OAL results.

Table 4.3. Post-V1 fixed-specimen ambient-light diagnostic

Lighting condition	Mean OAL	Within-condition P95 pairwise difference
Window blocked, ceiling light off (A)	0.999597 in	0.000107 in
Window exposed, ceiling light off (B)	0.999620 in	0.000146 in
Window blocked, ceiling light on (C)	0.999675 in	0.000171 in
Return to window blocked, ceiling light off (A2)	0.999681 in	0.000152 in

Relative to the initial A condition, the B-A difference was +0.000023 in, the C-A difference was +0.000078 in, and the A2-A return difference was +0.000083 in.

A separate controlled-repositioning study produced 30 valid PCA observations from ten placements of specimen 121, with three untouched frames per placement. The within-placement P95 pairwise difference was 0.000519 in. The placement-mean standard deviation was 0.000523 in, the placement-mean range was 0.001780 in, and the between-placement P95 pairwise difference was 0.001629 in. For reference, the formal V1 independent-placement P95 was 0.004831 in.

4.3 Prospective V2 PCA Results

4.3.1 Availability, agreement, and repositioning

All 180 scheduled V2 PCA observations were valid, again giving 100.000% availability. The 60 specimen-level errors had a mean bias of +0.000277 in with a 95% confidence interval from -0.000117 to +0.000672 in. Table 4.4 reports the full V2 result set.

Table 4.4. Prospective V2 PCA performance

Endpoint	V2 result
Valid observations	180/180
Availability	100.000%
Bias	+0.000277 in
MAE	0.001191 in
RMSE	0.001538 in
Nearest-rank P95 absolute specimen error	0.003065 in
Maximum absolute specimen error	0.004290 in
Nearest-rank P95 pairwise repositioning difference	0.003994 in
Maximum pairwise repositioning difference	0.005199 in

4.3.2 Classification and criteria

V2 PCA produced 40 true-conforming classifications, 14 true-nonconforming classifications, five false fails, and one false pass. The false-pass specimen was 143. The false-fail specimens were 9, 14, 104, 135, and 144. Sensitivity for detecting nonconforming specimens was 93.33%, and specificity for conforming specimens was 88.89%.

Table 4.5. Prospective V2 PCA criterion results

Criterion	Observed	Target	Outcome
Availability	100.000%	$\geq 99\%$	Met
Absolute mean bias	0.000277 in	≤ 0.0010 in	Met
RMSE	0.001538 in	≤ 0.0020 in	Met
P95 absolute specimen error	0.003065 in	≤ 0.0030 in	Not met
Maximum absolute specimen error	0.004290 in	≤ 0.0040 in	Not met
P95 pairwise repositioning difference	0.003994 in	≤ 0.0020 in	Not met

Criterion	Observed	Target	Outcome
False passes, preferred outcome	1	0 preferred	Not met

V2 PCA met three of the six quantitative criteria.

4.4 Retrospective V2 Endface-Hybrid Results

The frozen Endface-Hybrid estimator was applied retrospectively to all 180 preserved V2 images. Before comparison, regeneration of the V2 PCA results produced a maximum difference of 0.000000000000 in from the recorded PCA values. Endface-Hybrid returned valid results for all 180 observations and complete three-placement results for all 60 specimens.

Table 4.6. Retrospective V2 Endface-Hybrid performance

Endpoint	Retrospective V2 Endface-Hybrid result
Valid observations	180/180
Availability	100.000%
Bias	-0.000505 in
MAE	0.001257 in
RMSE	0.001601 in
Nearest-rank P95 absolute specimen error	0.003109 in
Maximum absolute specimen error	0.003842 in
Nearest-rank P95 pairwise repositioning difference	0.003021 in
Maximum pairwise repositioning difference	0.004557 in

The 95% confidence interval for mean bias extended from -0.000900 to -0.000109 in. Classification produced 40 true-conforming specimens, 15 true-nonconforming specimens, five false fails, and no false passes. The false-fail specimens were 9, 14, 58, 151, and 221.

These results describe retrospective application of the frozen estimator to previously acquired V2 images. They are not a prospective V2 validation result, and no criterion-pass count is assigned to them.

4.5 Controlled-Presentation Development Results

Controlled-presentation development used specimens excluded from the formal cohort. Removing base-to-mouth rocking for ten independent placements of specimen 146 reduced the recorded Endface-Hybrid pairwise P95 from 0.003271 in under the earlier Fixture 2 placement condition to 0.001486 in. The corresponding PCA result decreased from 0.003708 to 0.001641 in, and the Geometry result decreased from 0.003282 to 0.001844 in.

Fixed-roll trials then held the engraved roll orientation toward the camera. Table 4.7 reports the Endface-Hybrid pairwise P95 recorded for each ten-placement run.

Table 4.7. Development-only Endface-Hybrid fixed-roll results

Development run	Valid placements	Pairwise P95
Specimen 178, run 1	10/10	0.001240 in
Specimen 178, run 2	10/10	0.001616 in
Specimen 187	10/10	0.000170 in
Specimen 197	10/10	0.000979 in

Across these four fixed-roll runs, the mean of the run-specific P95 values was 0.001001 in, the median was 0.001109 in, and the largest was 0.001616 in. These development observations were excluded from the V3 formal dataset.

4.6 Prospective V3 Endface-Hybrid Results

4.6.1 Dataset completion and availability

The V3 acquisition audit confirmed 180 scheduled observations, 60 unique specimens, three formal placements per specimen, no missing formal observations, and no duplicate scheduled positions. The protocol-invalid Round 2 capture made with the wrong physical specimen under the scheduled specimen-35 label was preserved outside the formal observation table. The correctly identified scheduled observation was subsequently acquired under the protocol-defined replacement rule.

After completion of all three capture-only rounds, the frozen Endface-Hybrid estimator returned valid results for all 180 formal images. V3 Endface-Hybrid availability was therefore 100.000%.

4.6.2 Specimen-level agreement and repositioning

The primary V3 agreement calculation used the 60 specimen means specified in Chapter 3. Mean bias was -0.000435 in, and its 95% confidence interval extended from -0.000889 to +0.000018 in. Table 4.8 reports the remaining prospective results.

Table 4.8. Prospective V3 Endface-Hybrid performance

Endpoint	V3 result
Valid observations	180/180
Availability	100.000%
Bias	-0.000435 in
MAE	0.001517 in
RMSE	0.001794 in
Nearest-rank P95 absolute specimen error	0.003170 in
Maximum absolute specimen error	0.003913 in
Nearest-rank P95 pairwise repositioning difference	0.000633 in
Maximum pairwise repositioning difference	0.001164 in

The maximum absolute specimen error occurred for specimen 100. Its manual-reference mean was 1.000000 in and its V3 Endface-Hybrid mean was 1.003913 in. The P95 absolute-error value was the 57th ordered absolute error among the 60 specimen-level errors. The P95 repositioning value was the 171st ordered difference among 180 pooled pairwise differences.

4.6.3 Classification and prospective criteria

Under the unguarded 0.9900–1.0100 in specification rule, V3 produced 42 true-conforming classifications, 14 true-nonconforming classifications, three false fails, and one false pass. The false-pass specimen was 143. Its manual-reference mean was 0.989833 in and its V3 mean was 0.990504 in. The false-fail specimens were 9, 151, and 194.

Table 4.9. Prospective V3 criterion results

Criterion	Observed	Target	Outcome
Availability	100.000%	$\geq 99\,\%$	Met
Absolute mean bias	0.000435 in	≤ 0.0010 in	Met
RMSE	0.001794 in	≤ 0.0020 in	Met
P95 absolute specimen error	0.003170 in	≤ 0.0030 in	Not met
Maximum absolute specimen error	0.003913 in	≤ 0.0040 in	Met
P95 pairwise repositioning difference	0.000633 in	≤ 0.0020 in	Met
False passes, preferred outcome	1	0 preferred	Not met

V3 met five of the six quantitative criteria. The separate preferred outcome of zero false passes was not met under the unguarded decision rule.

4.7 Comparative V1–V3 Results

Table 4.10 presents the common endpoints for the three prospective configurations and the retrospective V2 Endface-Hybrid analysis. All agreement results in the table use 60 specimen means. Availability uses the 180 attempted observations in each configuration, and repositioning uses 180 pooled pairwise differences. The V2 Endface-Hybrid column is marked retrospective to distinguish its evidentiary status from the three prospective result sets.

Table 4.10. Comparison of OAL performance across system iterations

Endpoint	V1 PCA, prospective	V2 PCA, prospective	V2 Endface-Hybrid, retrospective	V3 Endface-Hybrid, prospective
Availability	100.000%	100.000%	100.000%	100.000%
Bias (in)	+0.001613	+0.000277	-0.000505	-0.000435
MAE (in)	0.001815	0.001191	0.001257	0.001517
RMSE (in)	0.002205	0.001538	0.001601	0.001794
P95 absolute specimen error (in)	0.003960	0.003065	0.003109	0.003170
Maximum absolute specimen error (in)	0.004799	0.004290	0.003842	0.003913
P95 pairwise repositioning difference (in)	0.004831	0.003994	0.003021	0.000633
True conforming	37	40	40	42
True nonconforming	14	14	15	14
False passes	1	1	0	1
False fails	8	5	5	3

Across the prospective iterations, P95 pairwise repositioning decreased from 0.004831 in for V1 to 0.003994 in for V2 and 0.000633 in for V3. Figure 4.1 summarizes this progression, including the 86.9% V1-to-V3 reduction. Over the same sequence, P95 absolute specimen error was 0.003960, 0.003065, and 0.003170 in, respectively. These are descriptive comparisons of the observed result sets; their engineering interpretation is developed in Chapter 5.

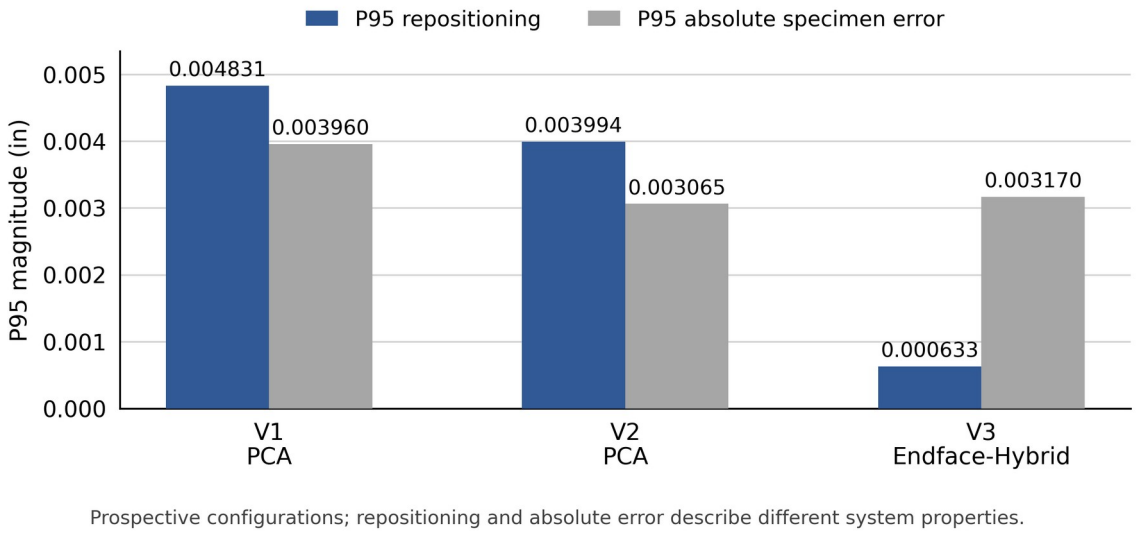


Figure 4.1. Prospective V1–V3 comparison of P95 repositioning and P95 absolute specimen error. The total V1-to-V3 change reflects combined configuration differences and does not isolate a single cause.

4.8 Exploratory Post-Hoc Guard-Band Results

The exploratory guard-band analysis was applied after completion of the prospective V3 analysis. It did not alter the formal V3 classification reported in Section 4.7. For a symmetric guard g, only V3 specimen means within $0.9900 + g$ to $1.0100 - g$ in were counted as automatically accepted. Specimen means inside the original specification interval but outside the guarded interval were assigned to the near-limit review zone.

Table 4.11. Exploratory symmetric guard-band results for the V3 cohort

Guard	Automatic-acceptance interval	Automatically accepted	Near-limit review	Observed false passes among automatically accepted	Conforming specimens not automatically accepted
0.0000 in	0.9900–1.0100 in	43/60	0	1	3
0.0005 in	0.9905–1.0095 in	39/60	4	1	7
0.0006 in	0.9906–1.0094 in	38/60	5	0	7
0.0010 in	0.9910–1.0090 in	34/60	9	0	11
0.0015 in	0.9915–1.0085 in	31/60	12	0	14
0.0020 in	0.9920–1.0080 in	30/60	13	0	15
0.0025 in	0.9925–1.0075 in	27/60	16	0	18
0.0030 in	0.9930–1.0070 in	24/60	19	0	21
0.0035 in	0.9935–1.0065 in	22/60	21	0	23
0.0040 in	0.9940–1.0060 in	21/60	22	0	24

The smallest symmetric margin in the evaluated cohort that produced zero observed false passes was 0.0006 in. Under that rule, 38 of 60 specimens were automatically accepted, five were in the near-limit review zone, and seven reference-conforming specimens were not automatically accepted. The five review-zone specimens were 35, 58, 104, 143, and 221. Specimen 143, the unguarded V3 false pass, was no longer automatically accepted.

The 0.0010-in engineering example automatically accepted 34 of 60 specimens, assigned nine specimens to near-limit review, produced no observed false passes among automatically accepted specimens, and left 11 reference-conforming specimens outside automatic acceptance.

These counts apply only to the intentionally stratified 60-specimen cohort. The 0.0006-in margin was selected after observing the same cohort and was not prospectively validated.

4.9 Chapter Summary

V1, V2, and V3 each produced valid measurements for all 180 scheduled CV observations. V1 PCA met one of six prospective quantitative criteria, V2 PCA met three of six, and V3 Endface-Hybrid met five of six. V3 met the maximum-absolute-error criterion with a specimen-level maximum of 0.003913 in but exceeded the 0.0030-in P95 absolute-error limit with a nearest-rank value of 0.003170 in.

P95 pairwise repositioning was 0.004831 in for V1 PCA, 0.003994 in for V2 PCA, 0.003021 in for the retrospective V2 Endface-Hybrid analysis, and 0.000633 in for prospective V3 Endface-Hybrid. Unguarded classification produced one false pass in each prospective configuration; false fails decreased from eight in V1 to five in V2 and three in V3. Retrospective V2 Endface-Hybrid produced no false passes and five false fails.

The post-hoc guard-band analysis found that a symmetric 0.0006-in margin was the smallest evaluated cohort margin that eliminated observed false passes, while a 0.0010-in example automatically accepted 34 of 60 specimens with zero observed false passes. These guard-band results remain exploratory and are interpreted in Chapter 5.

Chapter 5 - Discussion and Conclusions

5.1 Answer to the Original Validation Question

The original prospective validation question asked whether the frozen CaseMetrology computer-vision system could measure cartridge-case overall length (OAL) with acceptable agreement and repeated-measurement performance relative to an independent manual reference process under documented static inspection conditions. For the original V1 configuration, the answer is no.

V1 completed all 180 planned computer-vision acquisitions and therefore achieved 100% measurement availability. Availability, however, was the only one of the six quantitative performance criteria that V1 satisfied. Mean bias was +0.001613 in, RMSE was 0.002205 in, nearest-rank P95 absolute specimen error was 0.003960 in, maximum absolute specimen error was 0.004799 in, and P95 pairwise repositioning difference was 0.004831 in. The unguarded classification result also included one false pass and eight false fails, contrary to the prospectively preferred zero-false-pass outcome.

The V1 result should remain an immutable answer to the original prospective question. It established that the original free-placement PCA configuration did not satisfy the complete OAL validation requirements under the tested conditions. The later engineering program does not erase, correct, or reinterpret that outcome. Instead, V1 provided the preserved quantitative baseline that motivated the diagnostic and redesign sequence.

The final V3 configuration addressed a later and different engineering question: whether a modified, frozen system could improve the limitations revealed by V1. V3 combined controlled physical presentation with the frozen Endface-Hybrid estimator and was evaluated prospectively after the configuration had been frozen. All 180 planned V3 images produced valid

measurements. Using the 60 specimen means for agreement with the manual reference, V3 mean bias was -0.000435 in, MAE was 0.001517 in, RMSE was 0.001794 in, nearest-rank P95 absolute specimen error was 0.003170 in, and maximum absolute specimen error was 0.003913 in. P95 pairwise repositioning difference was 0.000633 in.

V3 therefore satisfied five of the six retained quantitative criteria. Availability met the >=99% target; absolute mean bias met the <=0.0010-in target; RMSE met the <=0.0020-in target; maximum absolute specimen error met the <=0.0040-in target; and P95 pairwise repositioning met the <=0.0020-in target. The one unmet quantitative criterion was P95 absolute specimen error, which was 0.003170 in compared with the <=0.0030-in target, an exceedance of 0.000170 in.

The preferred zero-false-pass decision outcome was separate from those six quantitative thresholds. Under the unguarded 0.9900–1.0100 in specification rule, V3 produced one false pass and three false fails. The final configuration therefore cannot be described as having satisfied every predefined performance expectation or as having been validated for unconditional final acceptance across the full specification interval.

The completed formal evidence consequently supports two distinct statements. First, the original V1 system did not satisfy the complete prospective validation requirements. Second, the later V3 system achieved substantial engineering improvement and met five of six quantitative criteria, but residual specimen-level absolute error and one observed false pass still limited the permissible claim. The final system is therefore better interpreted as evidence of a potentially useful static dimensional-screening architecture than as a demonstrated replacement for precision reference metrology.

5.2 Engineering Progression

The V1-through-V3 sequence is most informative when interpreted as an evidence-driven engineering progression rather than as a set of isolated experiments. Each stage was used to narrow the technical problem while preserving the evidentiary status of the work that preceded it.

V1 established the prospective baseline. The original PCA configuration returned valid measurements for all planned images, but its independent free-placement performance was not sufficiently stable. P95 pairwise repositioning was 0.004831 in compared with the 0.0020-in target. V1 also exceeded the limits for absolute mean bias, RMSE, P95 absolute specimen error, and maximum absolute specimen error. These results demonstrated that successful image acquisition and dimensional output were not sufficient to establish acceptable metrological behavior.

Post-V1 diagnostic studies investigated acquisition stability, repositioning sensitivity, fixture behavior, rocking, seating, and roll orientation. Those studies were exploratory and were not designed to retroactively explain the V1 result with a single causal mechanism. They did, however, show that repeated measurements could become more stable when physical presentation was constrained, thereby providing a defensible basis for fixture and presentation changes.

V2 introduced fixture modifications while retaining the PCA estimator. Relative to V1, mean bias decreased from +0.001613 in to +0.000277 in, RMSE decreased from 0.002205 in to 0.001538 in, and P95 absolute error decreased from 0.003960 in to 0.003065 in. P95 repositioning also decreased, from 0.004831 in to 0.003994 in, but remained above the 0.0020-in target. V2 therefore showed that fixture improvement alone did not fully resolve repeated-presentation variability.

The Endface-Hybrid estimator was subsequently developed using excluded development data and then applied retrospectively to the already acquired V2 images. Because the image set was unchanged, the V2 PCA-versus-Hybrid comparison is particularly useful for examining estimator effects without claiming prospective validation of the Hybrid result. P95 repositioning decreased from 0.003994 in with PCA to 0.003021 in with Endface-Hybrid. In contrast, RMSE changed from 0.001538 in to 0.001601 in and P95 absolute error changed from 0.003065 in to 0.003109 in. The retrospective comparison therefore showed that estimator design affected repeated-measurement behavior but did not eliminate the broader absolute-agreement limitation.

Controlled-presentation development then addressed physical presentation more directly. The final V3 configuration retained the Endface-Hybrid estimator while constraining seating, roll orientation, and fixture retention. The estimator and physical configuration were frozen before formal acquisition, and formal estimator-result interpretation was withheld until all 180 V3 images had been acquired.

The resulting system-level improvement in repeated-presentation behavior was substantial. P95 repositioning decreased from 0.004831 in for V1 to 0.000633 in for V3, a reduction of approximately 86.9%. That total V1-to-V3 change should not be attributed to a single component because both estimator and presentation conditions changed across the full engineering sequence.

Contact mechanics is consistent with treating the support, stop, and locator as parts of the measurement interface. It does not separate the contributions of individual contacts to the 0.000633-in V3 statistic; Figure 2.1 describes possibilities for subsequent study rather than a causal decomposition of these data.

The more specific V2 Endface-Hybrid-to-V3 comparison provides stronger evidence regarding presentation control. The same Endface-Hybrid estimator was used for both analyses, while the later V3 acquisition incorporated the controlled-presentation procedure. P95 repositioning decreased from 0.003021 in for the retrospective V2 Hybrid analysis to 0.000633 in for prospective V3, a reduction of approximately 79.0%. Because the datasets were acquired in different phases rather than in a randomized factorial experiment, this comparison still does not isolate a pure causal effect. It is nevertheless consistent with the diagnostic evidence that controlled presentation was a major contributor to the improved repeated-measurement behavior.

The engineering synthesis question can therefore be answered with three conclusions. First, physical presentation is part of the measurement architecture rather than an incidental setup detail. Second, estimator development and fixture changes can alter measurement behavior, but neither should be assumed to resolve all error modes. Third, once repeated-presentation variation was reduced in V3, the dominant remaining limitation was disagreement between specimen-level CV means and the manual reference. The completed sequence therefore separates a largely controlled repeatability problem from a residual absolute-agreement problem that requires a different technical explanation.

5.3 Repeatability Versus Absolute Accuracy

The distinction between repeated-presentation performance and absolute agreement is the central technical finding of the completed research.

Measurement repeatability concerns agreement among repeated measurements under specified conditions, whereas measurement agreement with a reference concerns the magnitude and structure of the difference between the two measurement processes. These properties are related but not interchangeable (Joint Committee for Guides in Metrology, 2012b). A system can

produce closely grouped repeated measurements while remaining offset from a reference, and a system can exhibit small mean bias while still producing comparatively large specimen-specific errors.

The V3 results demonstrate both points. P95 pairwise repositioning difference was 0.000633 in, well inside the 0.0020-in criterion. In contrast, the specimen-level agreement analysis produced an MAE of 0.001517 in, RMSE of 0.001794 in, nearest-rank P95 absolute error of 0.003170 in, and maximum absolute error of 0.003913 in. Mean bias was only -0.000435 in, but the much larger MAE and upper-tail errors show that a small average signed difference did not imply small error for every specimen. Positive and negative specimen-level differences can partially cancel in the mean.

The repositioning and absolute-error statistics should also not be treated as two estimates of the same quantity. Repositioning was calculated from pairwise differences among repeated CV presentations, whereas the primary agreement statistics compared each specimen's mean CV result with its accepted manual-reference mean. Their numerical separation is therefore interpreted structurally rather than as a direct subtraction of like quantities. The important observation is that V3 repeated-presentation behavior was well within its criterion while the upper tail of specimen-level reference disagreement still exceeded its criterion.

This distinction changes the engineering diagnosis. Under V1, repeated physical presentation was a major practical limitation and was therefore an appropriate target for fixture and procedure development. Under V3, repeated-presentation variability was substantially smaller, so further reductions in placement variation alone cannot be assumed to eliminate the remaining specimen-level disagreement. The residual error requires investigation of factors that affect the relationship between the image-derived dimension and the reference process.

Several mechanisms could plausibly contribute to that residual disagreement in a conventional-lens silhouette system, including spatial or directional calibration limitations, field-dependent magnification, optical distortion, measurement-plane geometry, endpoint localization, specimen end-face geometry, or interactions among those factors. The manual reference process can also contribute finite variability through instrument resolution, alignment, contact force, and the physical geometry of the case ends. The formal V3 results demonstrate the existence of residual CV-versus-reference disagreement but do not uniquely identify any one of these mechanisms as its cause.

The reference process must therefore be interpreted appropriately. The manual measurements were obtained using a calibrated instrument and documented session controls, and the reference uncertainty was explicitly characterized rather than assumed to be zero. Nevertheless, the manual reference is a measurement process, not an error-free realization of true OAL. The reported V3 errors quantify agreement with the accepted reference process under the study conditions.

This interpretation avoids two opposite errors. It would be an overstatement to characterize V3 as generally inaccurate merely because one upper-tail criterion was exceeded by 0.000170 in; five of six quantitative criteria were satisfied and repeated-presentation performance improved markedly. It would be equally incorrect to treat V3 as interchangeable with the reference process because mean bias and RMSE met their criteria. The P95 absolute-error result and the observed false pass show why upper-tail agreement and decision behavior remain important even when average performance is favorable.

5.4 Industrial Screening Feasibility

The final CaseMetrology configuration is more defensibly interpreted as a candidate dimensional-screening architecture than as a final reference-measurement replacement.

That interpretation follows from the combination of strengths and limitations demonstrated in V3. The system produced valid measurements for all 180 planned acquisition opportunities, and P95 pairwise repositioning was 0.000633 in. Mean bias, RMSE, and maximum absolute specimen error also met their retained quantitative targets. These results show that the final configuration can produce stable, consistently available OAL information under the documented static presentation procedure.

The same evidence also defines the limitation. Nearest-rank P95 absolute specimen error was 0.003170 in, slightly above the 0.0030-in criterion, and the unguarded specification rule produced one false pass. A binary decision process that automatically treats every CV mean within 0.9900–1.0100 in as a final conforming decision would therefore disregard an error mode that was actually observed in the validation cohort.

A screening architecture can handle that limitation differently. Measurements sufficiently far from a specification boundary can be considered candidates for automated disposition, while measurements near a boundary can be routed to an independent or higher-capability secondary measurement process. The value of such an architecture is not that it converts the vision system into reference metrology. Its value is that the decision process can be designed around the demonstrated measurement behavior rather than assuming the measurement error is negligible.

This interpretation is consistent with conformity-assessment principles in which specification limits and acceptance limits need not be identical when measurement uncertainty and decision risk are material (Joint Committee for Guides in Metrology, 2012a; International

Organization for Standardization, 2017). For CaseMetrology, however, those principles provide a framework rather than a validated production rule. The thesis demonstrates the feasibility of applying guarded screening logic to the observed V3 measurements; it does not establish the production risk associated with any specific acceptance boundary.

The intentionally stratified cohort is important to this conclusion. The 60 formal specimens were selected to provide representation near and beyond both specification limits, not to reproduce a natural manufacturing distribution. The observed proportions of true conforming, true nonconforming, false-pass, false-fail, automatic-acceptance, or review-zone outcomes therefore cannot be interpreted as production yield or expected production frequencies.

The study also did not evaluate production throughput, automated handling, conveyor motion, long-duration operation, economic benefit, or integration into a manufacturing quality system. The industrial conclusion is therefore deliberately narrow: the final static architecture produced measurement behavior that justifies further evaluation of a two-stage dimensional-screening process, but the present thesis does not demonstrate a production-ready inspection system.

5.5 Guarded Decision Logic

The post-validation guard-band analysis explored how the V3 measurement behavior could be incorporated into a three-way screening decision. Because the analysis was performed after the formal V3 results were known, it is retrospective and exploratory. It did not modify the V3 estimator, specimen measurements, primary confusion matrix, or prospective performance criteria.

For a symmetric guard (g), the candidate automatic-acceptance interval was

$$0.9900 + g \leq \acute{C}_i \leq 1.0100 - g .$$

A specimen mean inside that interval was considered a candidate for automatic acceptance. A mean remaining within the nominal specification interval but outside the guarded interval entered a near-limit review zone for secondary inspection. Measurements outside the nominal specification interval were not automatically accepted.

Within the 60-specimen validation cohort, the smallest observed symmetric margin that eliminated false passes was 0.0006 in. The corresponding candidate automatic-acceptance interval was 0.9906–1.0094 in. This is an empirical property of the observed cohort, not evidence that a 0.0006-in margin is generally safe.

A 0.0010-in symmetric margin provides a simpler engineering example. It narrows the candidate automatic-acceptance interval to 0.9910–1.0090 in. In the formal cohort, 34 of 60 specimens were automatically accepted under that exploratory rule, nine were placed in the near-limit review zone, and zero false passes were observed among the automatically accepted specimens.

The principal result of this analysis is the decision tradeoff rather than the numerical margin itself. Moving the acceptance boundary inward reduced the observed opportunity for false acceptance but also reduced the fraction of specimens eligible for automatic acceptance. A production implementation would therefore need to balance false-acceptance risk, false-rejection or referral burden, measurement uncertainty, the cost and capacity of secondary inspection, and the actual distribution of production OAL values (Joint Committee for Guides in Metrology, 2012a).

Because the same validation cohort was used both to observe V3 performance and to identify candidate margins, the zero-false-pass result is susceptible to cohort-specific selection. The analysis does not estimate a future false-pass probability and cannot demonstrate that the

same result would occur in an independent production population. A production guard band would require a prospectively specified risk basis and independent validation.

The exploratory industrial-feasibility question can therefore be answered cautiously. Guarded decision logic provides a technically plausible pathway for using the demonstrated V3 measurement behavior in a screening architecture, and the observed cohort shows the expected tradeoff between automatic acceptance and secondary review. The present data do not establish a validated production margin or a general false-acceptance rate.

5.6 Limitations

The conclusions of this study are bounded by the experimental design and by the specific CaseMetrology configuration that was evaluated.

First, one operator performed all formal manual-reference and computer-vision acquisitions. The reference data therefore characterize within-operator behavior, and the controlled V3 presentation procedure has not been shown to be reproducible across operators.

Second, the formal cohort contained 60 specimens selected through an intentionally non-proportional stratified design from the available 235-case population after exclusion of development specimens. This design strengthened evaluation near the specification boundaries but does not represent the natural distribution of a manufacturing population.

Third, the same frozen 60-specimen cohort and reference means were retained across V1, V2, and V3. This paired design is a strength for engineering comparison because specimen composition did not change between iterations, but it is not an independent replication of the final result on a new population.

Fourth, the study evaluated OAL only. No conclusion should be extended to rim diameter, primer-pocket geometry, concentricity, wall thickness, or other cartridge-case

characteristics without independent measurand definitions, calibration, reference methods, and validation.

Fifth, all formal computer-vision measurements were static. The study did not validate conveyor motion, robotic loading, automated orientation, vibration, line-speed imaging, or continuous production operation. The V3 controlled-roll and seating procedure should therefore be understood as an experimental presentation control, not as a demonstrated production-handling solution.

Sixth, the results are configuration specific. They apply to the documented Jetson platform, IMX296 camera, 16 mm lens, LED backlight, installed measurement plane, camera controls, calibration model, fixture, estimator, and software configuration. A materially different optical, mechanical, illumination, calibration, or software configuration should be treated as a new measurement configuration requiring appropriate verification or validation.

Seventh, the conventional-lens calibration model was characterized over the tested measurement region and included spatial and directional scale effects. It should not be interpreted as proof of universal sub-thousandth-inch dimensional accuracy at arbitrary image positions, orientations, or measurement-plane offsets.

Eighth, the manual comparator has finite uncertainty. Instrument resolution, calibration uncertainty, within-operator repeatability, alignment, contact behavior, and specimen end geometry can influence the assigned reference result. The study accounted for reference uncertainty but did not establish a complete uncertainty budget for every possible contributor to the final CV-versus-reference difference.

Ninth, the physical cause of the residual V3 absolute disagreement remains unresolved. The experiments show that repeated-presentation behavior improved substantially, but they do

not uniquely identify whether the remaining error arose primarily from optics, calibration, feature localization, specimen geometry, the reference process, or an interaction among factors.

The study did not independently measure fixture contact forces, surface topography, scratches, galling, transfer, coating condition, or wear progression. Contact-related changes at the case or fixture could affect presentation or the manual comparator, but their presence and contribution to residual absolute error remain untested.

Tenth, the V2 Endface-Hybrid results are retrospective. They provide valuable comparative evidence because the frozen estimator was applied to the already acquired V2 images, but they do not have the same prospective status as V1 PCA, V2 PCA, or V3 Endface-Hybrid formal acquisition.

Eleventh, the guard-band analysis is post hoc. The 0.0006-in zero-observed-false-pass margin and the 0.0010-in engineering example were derived and evaluated using the same 60-specimen cohort. Neither margin is a validated production decision rule.

Twelfth, the empirical P95 and maximum-error statistics are finite-sample results. The nearest-rank P95 is an empirical order statistic from the tested cohort, not a universal tolerance bound. The reported maximum is the largest error observed among the tested specimens, not a guarantee that future errors cannot be larger.

Finally, the study did not measure production throughput, economic benefit, comparative labor cost, long-term calibration stability, maintenance requirements, or production-environment robustness. Claims concerning speed, cost advantage, full-rate inspection, or production readiness therefore remain outside the evidence generated by this thesis.

5.7 Future Work

Future work should be directed toward the limitations that remained after V3 rather than repeating development that the present study already resolved.

The highest-priority technical question is the source of residual absolute disagreement. A designed diagnostic study could evaluate field position, working height, specimen tilt, roll, calibration location, scale-model behavior, and endpoint localization while maintaining the stable V3 presentation controls. A structured design would be preferable to unrestricted iterative tuning because it could separate competing explanations while preserving inferential clarity.

A second priority is stronger characterization of the reference relationship. Future work could use tighter control of manual contact force and end-face alignment, additional repeated reference measurements, or an independent higher-capability dimensional method. The objective would be to reduce ambiguity regarding how much of the observed CV-versus-manual disagreement arises from the vision system and how much may arise from the comparator or specimen geometry.

A separate fixture-interface study could vary contact geometry and material systematically, document insertion force and seating, inspect both contact surfaces for scratches or transfer, and repeat measurements after controlled contact cycles. Coatings or wear-resistant materials would be candidates only after the load, material pairing, and cleaning requirements are specified (Podgornik et al., 2006; Fallqvist et al., 2012). Such tests would evaluate a new configuration; they are not required to complete the present OAL validation.

If CaseMetrology is developed specifically as a screening system, the next decision study should define the guard-band rule prospectively before evaluating it. The margin should be tied to a stated risk objective and measurement model rather than selected because it removes errors

from the validation data. It should then be evaluated on an independent cohort representing the intended operating population, with false-acceptance, false-rejection, automatic-acceptance, and secondary-review rates reported separately.

Operator and handling effects should also be evaluated. A multi-operator study could determine whether the V3 presentation procedure is reproducible across users. Alternatively, a mechanically keyed or automated presentation mechanism could reduce operator dependence, but that mechanism would constitute a new system configuration and would require its own evaluation.

Production-oriented work would require a separate test program addressing computational cycle time, image-acquisition latency, automated part presentation, motion tolerance, throughput, recovery from failed measurements, environmental robustness, calibration maintenance, and integration with a secondary inspection path. These questions are distinct from the static metrology questions answered in this thesis.

Alternative optical configurations may be investigated if future requirements demand tighter absolute agreement. Higher magnification, improved distortion characterization, or telecentric optics could be evaluated prospectively against the present conventional 16 mm architecture. Such work would be a new research phase rather than a modification of the preserved V1–V3 evidence.

A further computational study could test alternatives at specific stages of the vision pipeline. A You Only Look Once (YOLO) detector could locate cartridge cases in a less constrained or moving scene, but its detection box alone would not provide a calibrated OAL measurement (Redmon et al., 2016). Density-Based Spatial Clustering of Applications with Noise (DBSCAN) could be evaluated for separating case edge points from isolated reflections,

debris, or fixture artifacts (Ester et al., 1996). A High-Resolution Network (HRNet) segmentation model could be tested for preserving fine base and mouth boundary detail under changes in lighting or presentation (Wang et al., 2021). YOLO and HRNet would require appropriately labeled development data, and all three methods would require an independent evaluation against the frozen V3 pipeline. Any comparison should report specimen-level agreement with the reference, repositioning, decision errors, measurement availability, and computational time; improved detection or segmentation alone would not establish improved dimensional accuracy.

Finally, the engineering framework could be extended to other cartridge-case dimensions. Each new characteristic should begin with its own measurand definition, reference method, calibration approach, acceptance criteria, and validation plan rather than assuming that the OAL results transfer automatically.

None of these extensions is required to preserve the conclusion of the present thesis. The completed V1–V3 program provides sufficient evidence to characterize the demonstrated static OAL system, define its current claim boundaries, and identify the specific questions that remain open.

5.8 Final Conclusion

This thesis evaluated CaseMetrology as a complete embedded dimensional-vision measurement system for cartridge-case OAL rather than as an image-processing algorithm in isolation. The work preserved the original prospective baseline, investigated the limitations revealed by that baseline, implemented evidence-driven engineering changes, and prospectively evaluated the final frozen configuration.

The answer to the original prospective validation question is that V1 did not satisfy the complete validation requirements. It achieved 100% measurement availability but failed the other five quantitative criteria and produced one false pass under the unguarded specification rule.

The subsequent engineering program materially improved the system. In V3, P95 pairwise repositioning was 0.000633 in, compared with 0.004831 in for V1, an overall system-level reduction of approximately 86.9%. Corrected specimen-level V3 agreement results were a mean bias of -0.000435 in, MAE of 0.001517 in, RMSE of 0.001794 in, nearest-rank P95 absolute error of 0.003170 in, and maximum absolute error of 0.003913 in. V3 met five of the six retained quantitative criteria. The remaining quantitative miss was the P95 absolute-error criterion, exceeded by 0.000170 in, and the unguarded decision rule still produced one false pass.

The completed engineering sequence demonstrates that repeated-presentation stability and absolute agreement were distinct system properties. The large improvement in V3 repeatability did not produce a proportional reduction in specimen-level disagreement with the manual reference. The V2 Hybrid-to-V3 comparison, together with the controlled-presentation development evidence, is consistent with physical presentation being a major contributor to the earlier repeatability problem, while the residual absolute-error component remains unresolved.

The broader theoretical implication is that a measurement is not solely a property of the object; it is the outcome of an interaction among the object, fixture, sensor, reference system, and interpretation algorithm. The CaseMetrology evidence illustrates this principle through a large reduction in repeated-presentation variation alongside persistent specimen-level disagreement. It does not determine which contact, optical, geometric, or reference mechanism produced the remaining disagreement.

The industrial-feasibility evidence is therefore bounded but useful. The final configuration demonstrated highly available and stable static OAL measurement behavior, and exploratory guard-band analysis showed a plausible way to separate automatic-acceptance candidates from near-limit measurements requiring secondary review. Because that decision analysis was retrospective and cohort dependent, no production guard band or future false-acceptance probability was validated.

Future studies could compare telecentric optics at a suitable magnification and test moving parts across realistic changes in position, in-plane angle, and roll. A prospectively validated pass, fail, or manual-review rule could then drive sorting, while part-level measurements feed statistical process control and an operator dashboard. These applications require independent evaluation of false passes, throughput, and reliability.

Accordingly, the thesis supports further development of CaseMetrology as a static dimensional-screening architecture under controlled presentation. It does not establish the prototype as interchangeable with precision reference metrology, as a production-ready inline inspection system, or as a validated unconditional final-acceptance device across the full specification interval. The principal contribution is a traceable engineering demonstration of what was improved, what remained unresolved, and how those measured limitations define the appropriate next stage of development.

References

Archard, J. F. (1953). Contact and rubbing of flat surfaces. Journal of Applied Physics, 24(8), 981–988. https://doi.org/10.1063/1.1721448

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet, 327*(8476), 307–310. https://doi.org/10.1016/S0140-6736(86)90837-8

Che, J. K., & Ratnam, M. M. (2018). Real-time monitoring of workpiece diameter during turning by vision method. *Measurement, 126*, 369–377. https://doi.org/10.1016/j.measurement.2018.05.089

Coleman, W. J., Reich, F. R., Erickson, M. D., & Kelly, W. S. (1974). *On-line automatic high speed inspection of cartridge cases* (BNWL-SA-5154; CONF-741020-2). Pacific Northwest Laboratory. https://doi.org/10.2172/4272487

Coleman, W. J., & Swinth, K. L. (1977). Automatic cartridge case inspection and process control monitor. *Proceedings of SPIE, 129*, 74–89. https://doi.org/10.1117/12.956017

Coveney, T. (2014). *Dimensional measurement using vision systems* (Measurement Good Practice Guide No. 39, 2nd ed.). National Physical Laboratory. https://eprintspublications.npl.co.uk/2042/

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)* (pp. 226–231). AAAI Press. https://www.aaai.org/Papers/KDD/1996/KDD96-037.pdf

Fallqvist, M., M'Saoubi, R., Andersson, J. M., & Olsson, M. (2012). Mechanical and tribological properties of PVD-coated cemented carbide as evaluated by a new multipass scratch-testing method. Advances in Tribology, 2012, Article 305209. https://doi.org/10.1155/2012/305209

Gonzalez, R. C., & Woods, R. E. (2018). *Digital image processing* (4th ed.). Pearson.

Greenwood, J. A., & Williamson, J. B. P. (1966). Contact of nominally flat surfaces. Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences, 295(1442), 300–319. https://doi.org/10.1098/rspa.1966.0242

Hashmi, A. W., Mali, H. S., Meena, A., Khilji, I. A., Hashmi, M. F., & Saffe, S. N. B. M. (2022). Machine vision for the measurement of machining parameters: A review. *Materials Today: Proceedings, 56*(4), 1939–1946. https://doi.org/10.1016/j.matpr.2021.11.271

Heinrichs, J., Olsson, M., & Jacobson, S. (2012). Mechanisms of material transfer studied in situ in the SEM: Explanations to the success of DLC coated tools in aluminium forming. Wear, 292–293, 49–60. https://doi.org/10.1016/j.wear.2012.05.033

Hemming, B. (2007). *Measurement traceability and uncertainty in machine vision applications* [Doctoral dissertation, Helsinki University of Technology].

International Organization for Standardization. (2017). *Geometrical product specifications (GPS)—Inspection by measurement of workpieces and measuring equipment—Part 1: Decision rules for verifying conformity or nonconformity with specifications* (ISO 14253-1:2017). https://www.iso.org/standard/70137.html

Joint Committee for Guides in Metrology. (2008). *Evaluation of measurement data—Guide to the expression of uncertainty in measurement* (JCGM 100:2008). https://doi.org/10.59161/JCGM100-2008E

Joint Committee for Guides in Metrology. (2012a). *Evaluation of measurement data—The role of measurement uncertainty in conformity assessment* (JCGM 106:2012). https://doi.org/10.59161/JCGM106-2012

Joint Committee for Guides in Metrology. (2012b). *International vocabulary of metrology—Basic and general concepts and associated terms (VIM)* (3rd ed., JCGM 200:2012). https://doi.org/10.59161/JCGM200-2012

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374*(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

Lai, Z., Li, H., Wang, N., & Zhang, X. (2026). A telecentric vision system tailored for high-precision dimensional measurement: Calibration and subpixel edge detection. *Measurement, 274*, 121239. https://doi.org/10.1016/j.measurement.2026.121239

Lazzari, A., & Iuculano, G. (2004). Evaluation of the uncertainty of an optical machine with a vision system for contact-less three-dimensional measurement. *Measurement, 36*(3–4), 215–231. https://doi.org/10.1016/j.measurement.2004.09.003

Lazzari, A., Iuculano, G., & Ponterio, A. (2001). Uncertainty analysis on an optical machine with vision system for three-dimensional measuring without contact. In *Proceedings of the 11th IMEKO TC4 Symposium on Trends in Electrical Measurements and Instrumentation and 6th IMEKO TC4 Workshop on ADC Modelling and Testing* (pp. 396–400). Lisbon, Portugal.

Lohr, S. L. (2022). *Sampling: Design and analysis* (3rd ed.). Chapman & Hall/CRC Press.

Malamas, E. N., Petrakis, E. G. M., Zervakis, M., Petit, L., & Legat, J.-D. (2003). A survey on industrial vision systems, applications and tools. *Image and Vision Computing, 21*(2), 171–188. https://doi.org/10.1016/S0262-8856(02)00152-X

National Institute of Standards and Technology. (n.d.). *Standard Reference Material 2461: Standard cartridge case*. https://www.nist.gov/pml/sensor-science/surface-and-interface-metrology/standard-bullets-and-cartridge-cases

NIST/SEMATECH. (n.d.). *e-Handbook of statistical methods*. National Institute of Standards and Technology. https://doi.org/10.18434/M32189

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics, 9*(1), 62–66. https://doi.org/10.1109/TSMC.1979.4310076

Podgornik, B., Vižintin, J., & Hogmark, S. (2006). Improvement in galling performance through surface engineering. Surface Engineering, 22(4), 235–238. https://doi.org/10.1179/174329406X122946

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 779–788). https://doi.org/10.1109/CVPR.2016.91

Sakarya, U., Topçu, O., Leloğlu, U. M., Soysal, M., & Tunalı, E. (2012). Automated region segmentation on cartridge case base. *Forensic Science International*, *222*(1–3), 277–287. https://doi.org/10.1016/j.forsciint.2012.07.004

Shin, D., Han, C., & Moon, Y. S. (2002). Ball stud inspection system using machine vision. *ISA Transactions, 41*(1), 3–11. https://doi.org/10.1016/S0019-0578(07)60197-9

Swinth, K. L., Walker, R. A., & Nedrow, N. G. (1979). High throughput dimensional measurements with electro-optics. *Proceedings of SPIE, 193*, 282–290. https://doi.org/10.1117/12.957902

Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., Liu, W., & Xiao, B. (2021). Deep high-resolution representation learning for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 43*(10), 3349–3364. https://doi.org/10.1109/TPAMI.2020.2983686

Wang, S., Li, X., Zhang, Y., & Xu, K. (2024). Effects of camera external parameters error on measurement accuracy in monocular vision. *Measurement, 229*, 114413. https://doi.org/10.1016/j.measurement.2024.114413

Xiang, R., He, W., Zhang, X., Wang, D., & Shan, Y. (2018). Size measurement based on a two-camera machine vision system for the bayonets of automobile brake pads. *Measurement, 122*, 106–116. https://doi.org/10.1016/j.measurement.2018.03.017

Yammen, S., & Muneesawang, P. (2013). Cartridge case image matching using effective correlation area based method. *Forensic Science International, 229*(1–3), 27–42. https://doi.org/10.1016/j.forsciint.2013.03.015