Proceedings of the 2005 Steel Founders' Society of America – Technical and Operating Conference

Measurement Error of Visual Casting Surface Inspection

Gokcer "Mike" Daricilar
Frank Peters

Industrial & Manufacturing Systems Engineering Department
Iowa State University
Ames, Iowa

Abstract

Visual inspection of steel casting surfaces is a critical step in the processing of castings. Several inspections are conducted by inspectors, operators, supervisors, management, and ultimately the customer. Over-inspection will cause unnecessary processing times and costs. Under-inspection will cause additional rework or customer returns. A method to assess the amount of measurement error that exists in a subjective inspection task, such as surface inspection, was developed. The method was used to determine the amount of measurement variation that exists at three steel foundries. The results showed that there is significant repeatability error (variation within an operator's performance) when the same inspector inspected the same castings. However, the reproducibility error (variation between operators) was worse. This indicates a need for better training and communication of customer requirements via work instructions or comparator plates.

INTRODUCTION

Visual evaluation of casting surfaces is conducted several times during the production of steel castings. The most obvious persons conducting these visual evaluations are the inspectors which mark up areas to be upgraded. However, evaluations are also done by operators while they are processing the castings, by management, by sales, and ultimately by the customer. Studies show that currently there is no satisfactory method, whereby the surface quality requirements can be communicated throughout the manufacturing and purchasing phases of casting production. The lack of a reasonable measurement system for quality causes several implications, including uncontrolled processing times. Undetected surface defects will result in additional rework cycles or worse, returns from the customer. Marking acceptable anomalies as defects will result in excessive processing.

The goal of this study is to assess the amount of variation introduced by these inspections. To accomplish this goal, a new methodology is developed in this paper that is needed to quantify the amount of measurement variation in terms of repeatability (variation within the same operator) and reproducibility (variation between different operators).

PREVIOUS WORK

A review of research in the areas of surface inspection and surface measurement were analyzed as a starting point for this study. Although no studies in the visual assessment of casting surface inspections were found, many studies in the optical casting surface defect detection, recognition, and classification were noted. A study with encouraging results looked at an automated visual inspection system for the detection of defects in a range from images of castings (Newman et al., 1995). The system used computer-aided design model information and inspection algorithms in several stages, including surface classification and inspection. The inspection system was able to correctly classify over 90% of the casting images. Some not so successful attempts were also made in the automated detection of surface defects in machined castings. A study by Woods and Allen (1989) utilized computer vision to automate the inspection of machined steel components for cracks using the fluorescent magnetic particle method. Unfortunately, the study's results indicated that although the candidate generation stage was promising, reliable detection of all defects was achieved only at the cost of an unacceptably high overall false positive rate.

Although some success was achieved in the automated surface inspection systems, they were limited to a specific type of surface defect, such as surface cracks. These automated systems are usually not practical, as the castings need to be inspected for a number of different types of surface defects. A study by Someji et al. (1997) explained that because of the various defect types and their complicated shapes, defect inspection of castings is dependent on the human eye. However, the study also claimed that this process could get unstable by the tiredness of a person or a change in the environment, and become highly subjective.

A gage R&R study by Lee et al. (1992) on subjective evaluation of image data from the medical industry was also examined. The study used nine medical x-ray CT head images from three patients as test cases. Six radiologists participated in reading the 99 images (some were duplicates) compressed at four different image compression ratios, original, 5:1, 10:1, and 15:1. The study found that the six readers agreed more than by chance alone and their agreement was statistically significant, but there were large variations among readers as well as within a reader. What this study analyzed is very similar to the visual casting surface inspections, as both are highly subjective because of their reliance on the human eye.

DEFINING MEASUREMENT ERROR

Defect detection is a very important task that is dependent on the human eye during the casting surface inspection process. In order to assess the amount of variation introduced by the visual casting surface inspections, the measurement error associated with this process has to be defined.

In practice, visual surface inspections of the castings are performed by operators who identify areas of the casting that need grinding or welding as defects, and mark those areas with a marker. This study utilized inspectors that perform these tasks on a daily basis. This required a total of four inspection trials for each casting, as every casting needed to go through two operators twice. After the castings were marked for defects, round stickers were used to cover the markings as shown in Figure 1. This made locating and quantifying the size of the marked areas possible. More detail about the sticker size and the application process can be found later. After the location of the stickers was recorded, the casting was cleaned and presented to the other operator. At least a day transpired before the same operator inspected the castings the second time.

Figure 1. Example of the casting with surface anomalies, the markings made by the inspector, and the stickers used to identify the defect size and location.

In this study, a cluster of stickers in the same region is called a master cluster and defined as a group of stickers that have contact with each other and located anywhere among the four combined inspection trials of the same casting. The master cluster concept is introduced to characterize a marked area by an operator as a supposed defect region. Master clusters represent the supposed defect regions identified and marked by the operators during the inspection trials. A search zone is created about each sticker, and all other stickers from any trial are checked to see if they are in this zone. If they are, the stickers are assigned to the same master cluster. The radius of the search zone is the radius of the sticker multiplied by a user defined search zone coefficient (with a default value of 2.5). It is possible for a sticker to have multiple touching stickers if the center positions of more than one sticker fall within the search zone, especially from different inspection trials.

A visual representation of the master cluster concept is displayed in Figure 2. In this case, operator 1's trial 1 contains 7 stickers for 3 supposed defect regions, and trial 2 contains 4 stickers for 3 supposed defect regions for the same casting. Operator 2's trial 1 contains 5 stickers for 3 supposed defect regions, and trial 2 contains 4 stickers for 3 supposed defect regions for the same casting. Therefore, the four combined inspection trials of the same casting by two different operators result in 4 master cluster regions. The combining of the inspection trials of a casting can also be described as superimposing its inspection images on top of each other. This combination operation is displayed with a union ($\cup$) symbol.

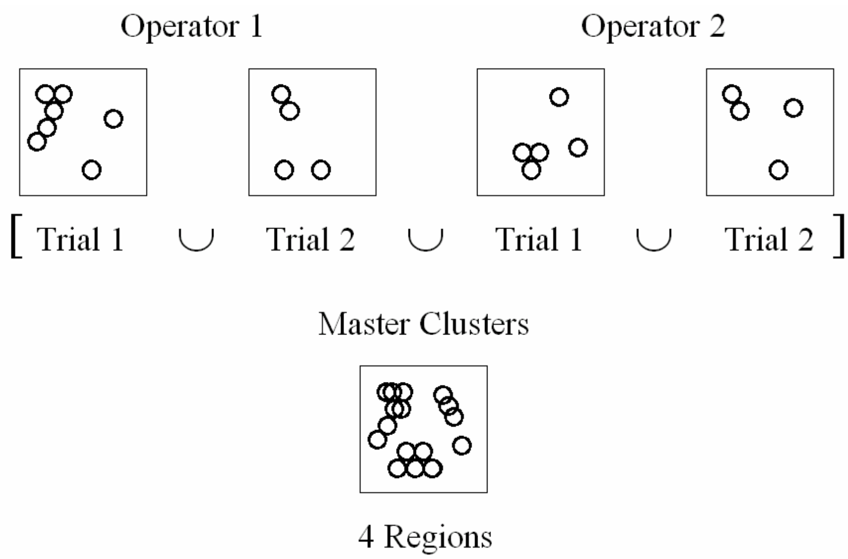

Figure 2. Defining master clusters with the four combined inspection trials of the same casting. The two stickers at the bottom of the casting from trial 2 of operator 1 both falls in the same master cluster region, since they are connected by stickers from other trials.

Defining Repeatability

Repeatability error of an operator is the amount of variation between the two inspections of that operator of the same casting. Repeatability is defined here by two aspects. The first represents how well an operator performs in identifying the same supposed defect regions between the two inspection trials of the same casting. The other aspect represents how well the operator performs in defining the size of the supposed defect regions between the two inspection trials of the casting.

The two inspection trials are compared by one-to-one sticker matching to determine the repeatability of that operator. This one-to-one sticker matching process is also done with a search zone. During the one-to-one sticker matching process, if the center position of another sticker from the other inspection trial of the same casting falls within this zone, then the two stickers are considered to be matched. It is also possible for a sticker to have multiple matching stickers if the center positions of more than one sticker fall within the search zone.

The repeatability of an operator is reported as two different percentage values. The first aspect of repeatability indicates how well the operator performed in identifying the same supposed defect regions between the two inspection trials of the same casting, reported as percent defect match. The second percentage value indicates how well the operator performed in defining the size of the supposed defect regions between the two inspection trials of the same casting, reported as percent sticker match. Higher values indicate better repeatability.

A visual representation for the repeatability of two different operators can be seen in Figure 3. Operator 1's trial 1 contains 7 stickers for 3 supposed defect regions, and trial 2 contains 4 stickers for 3 supposed defect regions for the same casting. The two inspection trials of the same casting are then compared. This comparing operation is displayed with an intersection (∩) symbol. Operator 1's two inspection trials for the same casting match 7 stickers out of 11, and 2 master clusters out of 3 in total. For this hypothetical example, operator 1's repeatability is 67% and 64% for master cluster and sticker matching, respectively. This was repeated for each operator in the study.

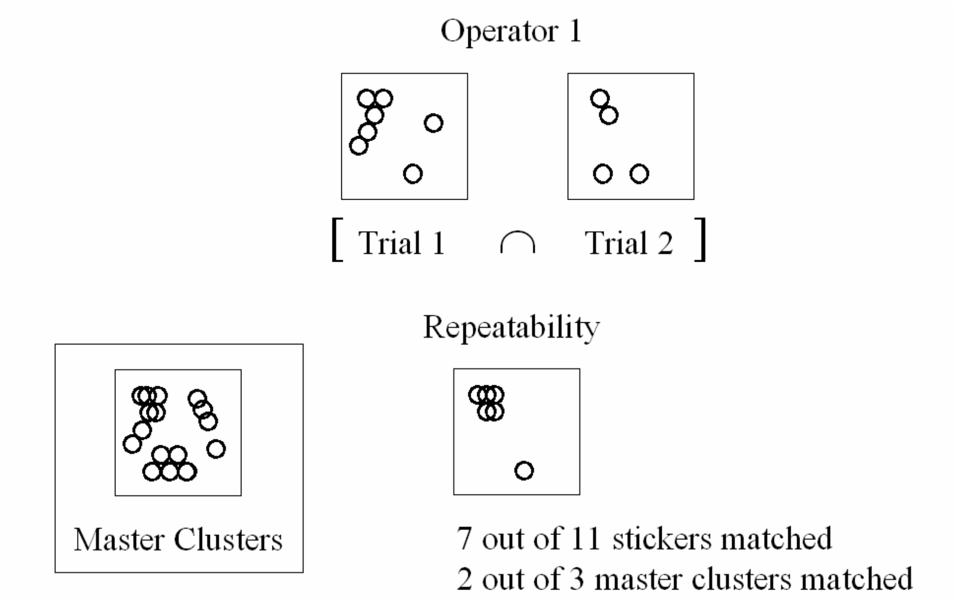

Figure 3. Defining repeatability of operator 1 with the same casting.

Defining Reproducibility

Reproducibility error is the amount of variation caused by differences between the two operators. The two inspection trials of the operators are combined to represent the best evaluation of each operator for that casting. Therefore, better repeatability from the operators will result in better reproducibility between the operators.

The combined inspections of the two operators are compared by one-to-one sticker matching to determine the reproducibility of the operators. This one-to-one sticker matching process is also performed with a search zone. During the one-to-one sticker matching process, if the center position of a sticker from the other operator's inspection trial of the same casting falls within this zone, then the two stickers are considered to be matched. It is also possible for a sticker to have multiple matching stickers if the center positions of more than one sticker fall within the search zone. Similar to repeatability, the reproducibility of two operators is also reported as two different percentage values. The first value indicates how well the operators performed in identifying the same supposed defect regions, reported as percent defect match. The second value indicates how well the operators performed in using the same size markings to define the same supposed defect regions, reported as percent sticker match. The higher values correspond to better reproducibility.

A visual representation for the reproducibility of two operators can be seen in Figure 4. In this case, operator 1's trial 1 contains 7 stickers for 3 supposed defect regions, and trial 2 contains 4 stickers for 3 supposed defect regions for the same casting. Operator 2's trial 1 contains 5 stickers for 3 supposed defect regions, and trial 2 contains 4 stickers for 3 supposed defect regions for the same casting. Then, the two inspection trials of the operators are combined to represent the best evaluation of each operator for that casting. Operator 1's combined trial contains 11 stickers for 4 supposed defect regions, and operator 2's combined trial contains 9 stickers for 4 supposed defect regions. The two combined inspection trials of the operators for the same casting are then compared. The operators' two inspection trials for the same casting match 12 stickers out of 20, and 3 master clusters out of 4 in total. For this hypothetical example, the two operators' reproducibility is 75% for master cluster matching and 60% for sticker matching.

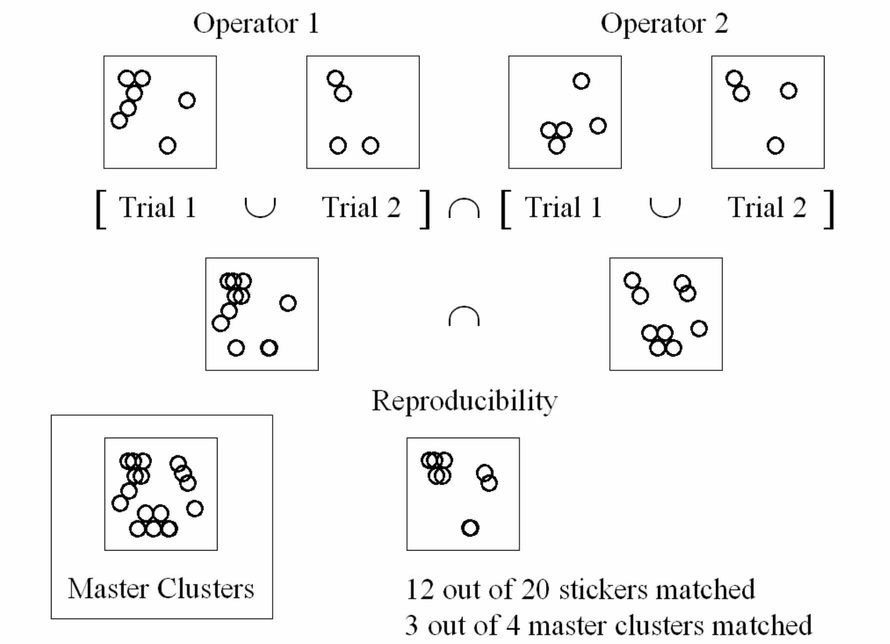

Figure 4. Defining reproducibility of two operators with the same casting.

METHODOLOGY

The first step was the collection of the visual quality inspection image data of the castings from three steel foundries. The next stage was the extraction of the coordinate data from the images. The final stage was the analysis of this coordinate data.

Image Data Collection

Image data was collected at three steel foundries, which collectively represent the North American steel casting industry quite well, outside of the railroad market. The companies ranged from 150 to 300 employees, and produced castings for a variety of construction equipment, pump and valve, and industrial equipment applications. One of these companies almost exclusively pours high alloy castings, another almost exclusively carbon and low alloy steel castings, and the third pours about 75% carbon and low alloy with the remainder being high alloy and wear resistant grades.

A similar procedure was used at each foundry for the image data collection. The setup included 2 visual inspectors and 6 castings in Foundry 1, and 10 castings in Foundries 2 and 3. The castings were marked for defects by the operators employing the same method that they typically use to inspect castings. The only difference was that only one side of the castings was inspected for this study. The order was chosen randomly by the trial moderator. Each of the two operators inspected each casting twice, on different days, to reduce bias. Between trials, the castings were shot blasted. Figure 5 shows an example casting as marked by an operator and after the stickers were applied

Specific instructions were provided to the trial moderator of each experiment as to how the stickers needed to be placed; some examples are shown in Figure 6. Careful attention was paid to carry out this process as consistently as possible. Then, digital pictures of the sticker-covered castings were taken.

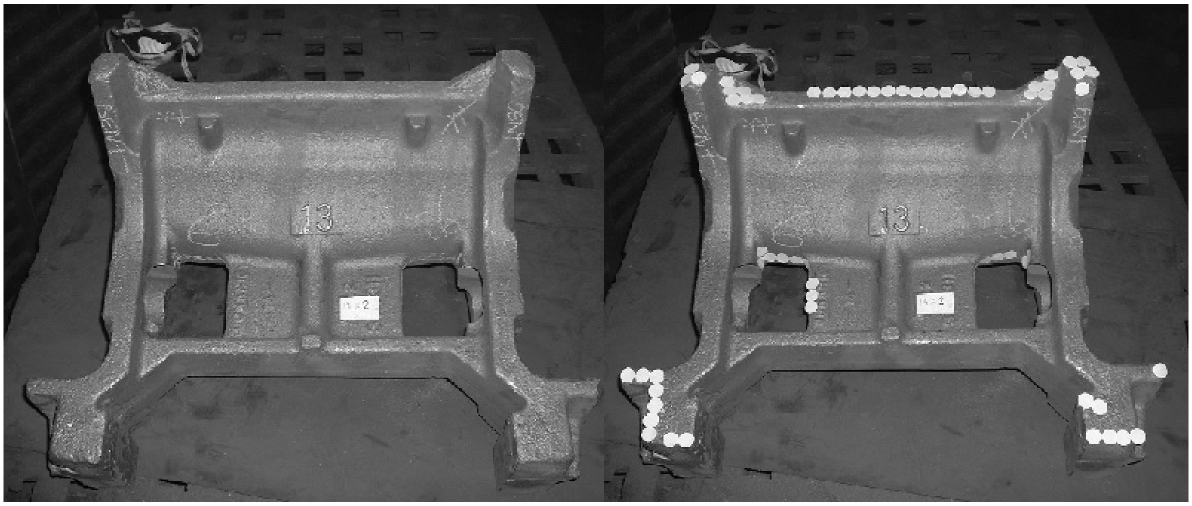

Figure 5. Casting marked by an operator (a) and stickers applied to the markings (b).

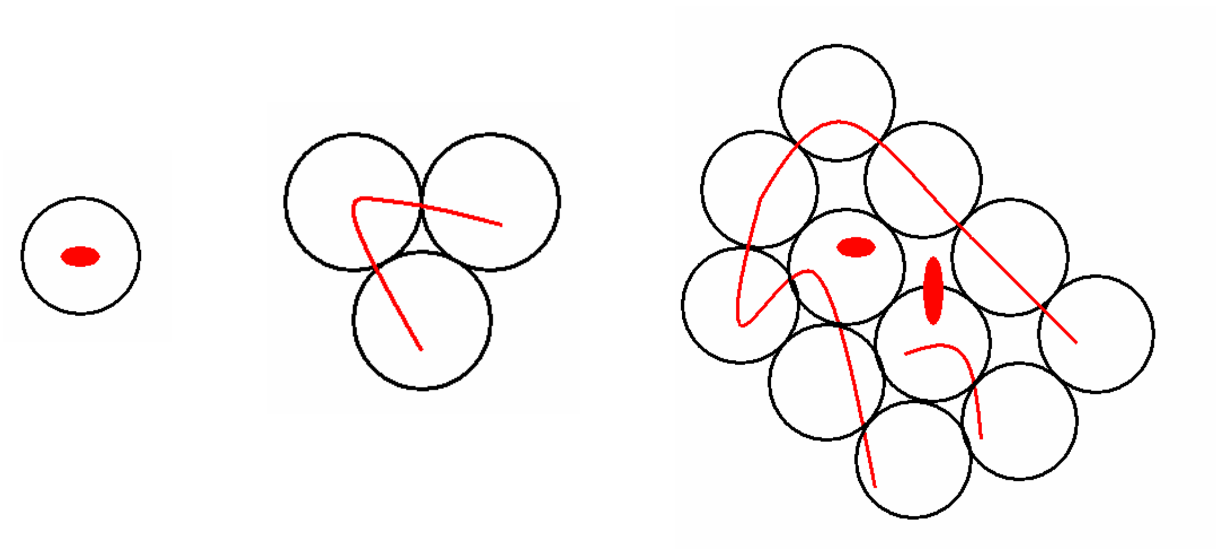

Figure 6. Example defect markings and how they were represented with stickers.

Sticker Coordinate Data Extraction

All images went through a series of processes to satisfy specific requirements of the image analysis software, and eliminate noise. The image cleaning process was done using Adobe Photoshop. Key steps in the image cleaning process included manually selecting the outlines of the castings and stickers. The image cleaning eliminated image noise, but did not change the location of the stickers.

After all the images are cleaned, they were loaded into DVT FrameWork image analysis software. DVT FrameWork is most commonly used in analyzing images for quality control purposes and was used to determine the coordinates of the stickers. The software compensated for the orientation and translation of the casting within the image. The sticker coordinate data was then outputted to MS Excel.

Analysis

After the sticker coordinate data was saved in a spreadsheet, another MS Excel file was used for the analysis, via algorithms implemented in macros. The master cluster operation was used to take the sorted sticker coordinate data and group the stickers into master clusters. The default search zone coefficient for the master cluster search was 2.5. If two stickers were touching, their centers would be 2 times the radius apart. The default was set at 2.5 to catch stickers which were close to touching. The default coefficient value was used for the Foundry 1 data; a value of 3.0 (a 20% increase of the default value) was used for the data from Foundries 2 and 3 because of the lower image quality and higher number of stickers that were at an angle to the image view. Further discussion on the determination of search zone coefficients and a sensitivity analysis can be found later.

The repeatability operation was used to take the sorted sticker data grouped into master clusters and conduct a one-to-one sticker matching between the different trials of the same castings inspected by the same operator. This process was done for each operator separately and repeated for each casting to determine the repeatability of each operator. The default search zone coefficient for the repeatability search was set at 1.5. The goal of the search zone in this case is to find stickers which are in the same location, but on different trial images. A value of 1.5 finds stickers whose centers are within 1.5 times the sticker radius. Although Foundry 1 used the default 1.5 for the search zone coefficient, Foundries 2 and 3 used 1.8 (a 20% increase from the default value) because of the lower image quality and higher number of stickers that were at an angle to the image view.

The reproducibility operation was used to take the sorted sticker data grouped into master clusters and conduct a one-to-one sticker matching between the same trials of the same castings inspected by different operators. This process was done with the combined images of the two trials of the same operator for the same casting and repeated for each casting to determine the reproducibility of the two operators. The default search zone coefficient for the reproducibility search was 1.5. Foundry 1 used the default coefficient value, and Foundries 2 and 3 used 1.8 (a 20% increase from the default).

RESULTS

The results of the analysis showed that there is a significant amount of repeatability and reproducibility error in the visual casting surface inspections. The repeatability of the two operators from the same foundry was found to be very similar. The repeatability for each operator is displayed in Figure 7. The reproducibility for each foundry (pair of operators) is displayed in Figure 8.

There was some concern that the blasting operations between trials might affect the casting surface quality, and bias the results. Pictures of the unmarked castings before and after the inspections do not show significant changes. There was also not a significant difference in the defect detection rates of the operators between their first and second inspection trials.

The search zone coefficient is a number, which, when multiplied by the radius of a sticker, creates a circular search zone around the sticker's center position. If the center position of another sticker from any of the four inspection trials of the same casting fell within this zone in the master cluster operation, then the two stickers were considered to be touching and assigned to the same master cluster. If the center position of another sticker fell within the respective zone in the repeatability or reproducibility operations, then the two stickers were considered to be matched. A sensitivity analysis of the search zone coefficient showed that the size of the search zone was appropriate, as reported by Daricilar (2005).

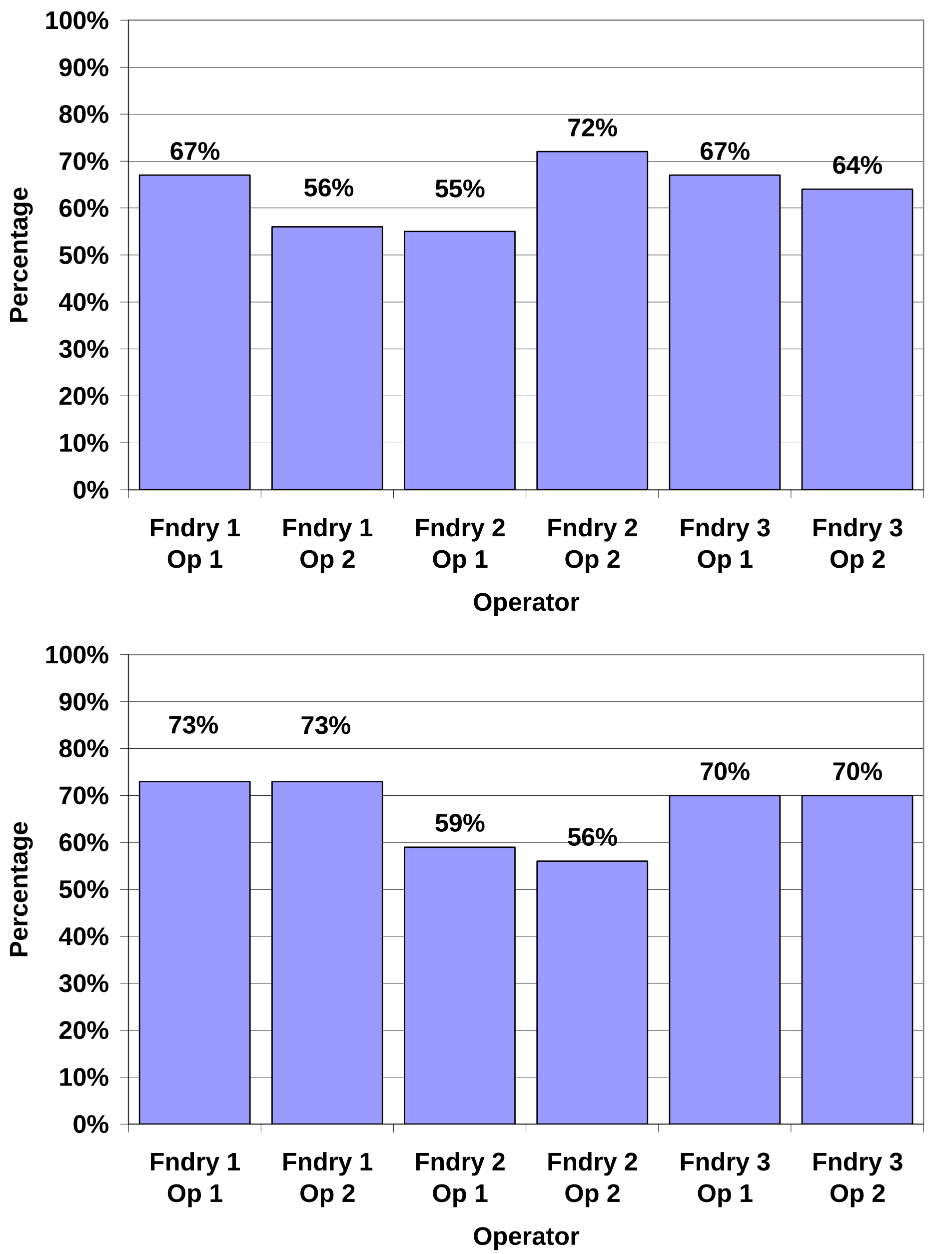

Figure 7: Repeatability values for each operator a) Percent defect match b) Percent sticker match

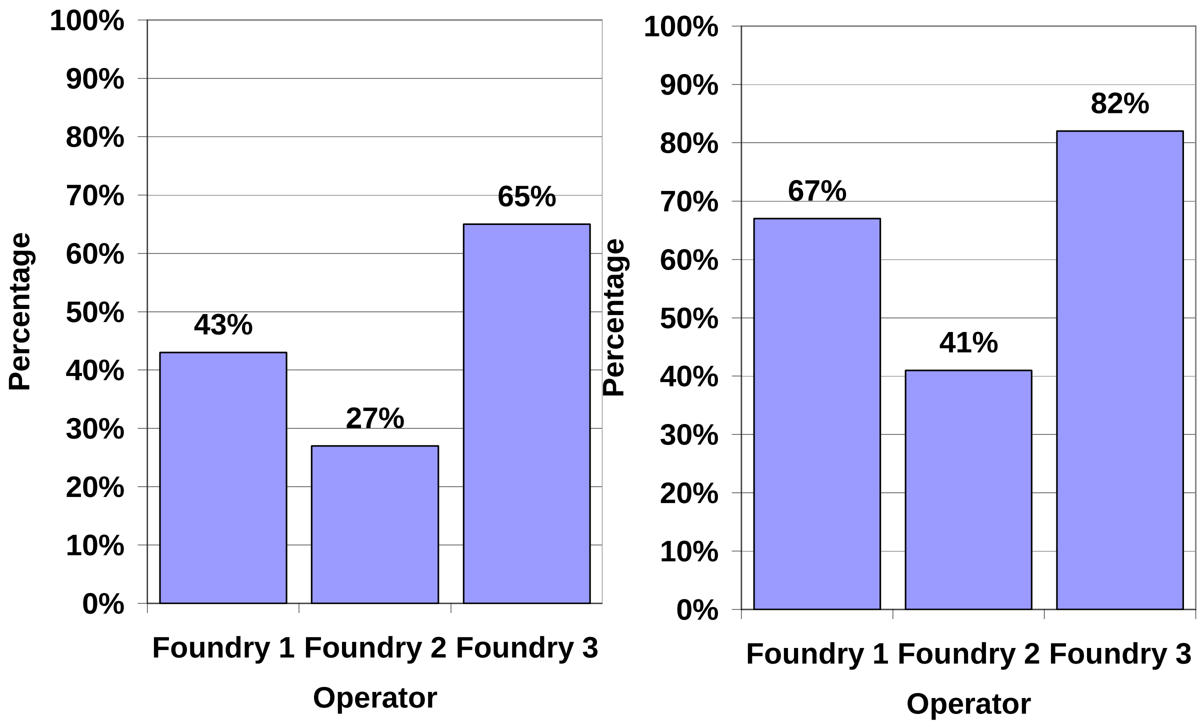

Figure 8: Reproducibility between the 2 operators at each foundry a) Percent defect match b) Percent sticker match

CONCLUSIONS

The results of the analysis showed that there was significant variation in both repeatability and reproducibility measurements from all three foundries. Although the repeatability measurements were somewhat consistent within the foundries, the reproducibility measurements displayed considerably more variation. This poses a particularly big problem in the industry, indicating a need for more training and the use of tools such as comparator plates and work instructions that detail customer requirements.

Another cause for concern was the higher variability detected in the defect match compared with the sticker match percentages in the repeatability and reproducibility measurements from all three foundries. The error associated with the defect match creates a bigger problem for the foundries, as incorrectly identifying or missing a whole defect region is much worse than identifying an already detected defect region as a bigger or smaller area. This variability in the defect detection process means that some defects go undetected from the inspection process and create rework or they reach the customer. Meanwhile, some allowable anomalies are incorrectly identified as defects. The uncertainty in the surface defect detection causes uncontrolled processing times leading to more inefficiency in cleaning room operations, such as excessive grinding and welding, and adversely affects other aspects such as production scheduling and material handling. All of these implications lead to increased cost of operations at steel foundries. Harwood (2004) reported that the cost of cleaning room operations are within the range of $0.76 to $5.34 per square inch (depending on the depth of the defect area) for welding operations, and $0.46 to $1.25 per square inch (depending on the surface quality after welding) for grinding operations. This means that there are high costs associated with every unnecessary defect marking placed by the casting surface inspection operators.

REFERENCES CITED

Daricilar, G. (2005), *Measurement Error of Visual Casting Surface Inspection*, M.S. Thesis, Iowa State University.

Harwood, B. (2004), *Unpublished Study of Cleaning Room Costs at a Steel Foundry*.

Lee, H., Rowberg, A. H., Frank, M. S., Choi, H. S. and Kim, Y. (1992), "Subjective evaluation of compressed image quality," *Proceedings of SPIE – The International Society for Optical Engineering*, vol. 1653, pp. 241-251.

Newman, T. S., Jain, A. K. (1995), "System for 3D CAD-based inspection using range images," *Pattern Recognition*, vol. 28, no. 10, pp. 1555-1574.

Someji, T., Yoshimura, T., Akiyama, N. (1997), "Development of automatic surface inspection system of casting," *Seimitsu Kogaku Kaishi/Journal of the Japan Society for Precision Engineering*, vol. 63, no. 10, pp. 1412-1416.

Woods, P. W., Allen, D. (1989), "Automated detection of surface defects in machined castings," *British Journal of Non-Destructive Testing*, vol. 31, no. 12, pp. 665-670.

ACKNOWLEDGEMENT

This material is based upon work supported by the U.S. Department of Energy under Award No. DE-FC07-02ID14228. Any opinions, findings, or conclusions and recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Department of Energy.