

ITSnoise: An Image Quality Dataset With Sensor Noise

**Joel Dumke
Margaret H. Pinson**



Technical Memorandum

ITSnoise: An Image Quality Dataset With Sensor Noise

**Joel Dumke
Margaret H. Pinson**



U.S. DEPARTMENT OF COMMERCE

Alan Davidson
Assistant Secretary of Commerce for Communications and Information
National Telecommunications and Information Administration

September 2022

DISCLAIMER

Certain commercial equipment and materials are identified in this report to specify adequately the technical aspects of the reported results. In no case does such identification imply recommendation or endorsement by the National Telecommunications and Information Administration, nor does it imply that the material or equipment identified is the best available for this purpose.

PREFACE

This memorandum is part of a series of NTIA Technical Memorandums. Each publication describes a subjective experiment that is named in series and distributed freely on the Consumer Digital Video Library (CDVL, www.cdvl.org) for research and development purposes. These experiments provide training data for no-reference (NR) metrics that focus on consumer camera applications. The reader is expected to have some knowledge of subjective experiments. A tutorial on this subject can be found in “Video Quality Assessment: Subjective testing of entertainment scenes,” by Margaret H. Pinson, Lucjan Janowski, and Zdzislaw Papir, published in *IEEE Signal Processing Magazine*, January 2015.

CONTENTS

Figures	v
Tables	vi
Glossary of Terms.....	vii
1. Introduction.....	1
2. Dataset Design	3
2.1 Scenes	3
2.2 Image Capture Methods.....	4
3. Subjective Test Design.....	6
4. Mean Opinion Score Analysis	8
5. References.....	10
Appendix A : MOS Values and Number of Ratings.....	11
Appendix B : Test Subject Instructions.....	13

FIGURES

Figure 1. MOS values.....	8
---------------------------	---

TABLES

Table 1. ITSnoise Scenes	3
Table 2. ITSnoise Devices and ICMs	4
Table 3. Skip Statistics	6
Table 4. Subject Demographics	7
Table A-1. MOS Values.....	11
Table A-2. Number of Ratings	11

GLOSSARY OF TERMS

ACR	absolute category rating
AI Focus	Artificial Intelligence Focus
Auto-servo focus	A Nikon auto-focus mode where the camera detects whether a subject is stationary or in motion and automatically elects to use either continuous or single servo (locked) autofocus
CCRIQ	Consumer Content Resolution and Image Quality Dataset
CDVL	Consumer Digital Video Library
HN	high noise
HRC	hypothetical reference circuit
ICM	image capture method
ISO	International Organization for Standardization refers to an exposure control value that impacts the brightness of a photograph
ITS	Institute for Telecommunication Sciences
LN	low noise
MOS	mean opinion score
NR	no reference
NTIA	National Telecommunications and Information Administration
P mode	Programmed Automatic (settings)
SLR	Single-Lens Reflex

ITSNOISE: AN IMAGE QUALITY DATASET WITH SENSOR NOISE

Joel Dumke and Margaret H. Pinson¹

This technical memorandum provides details for the image quality subjective experiment ITSnoise. This experiment includes 24 scenes, each digitally photographed using 12 different Image Capture Methods (ICMs) for a total of 288 images. The ICMs were designed to reflect the way images are captured in the most common public safety applications and also to generate a range of different sensor noise levels in the resulting images, from very low to very high. This dataset was intended to provide training and testing data for potential no-reference sensor noise metrics that could automatically predict the impact on perceived quality of sensor noise within a given image. This dataset is freely available for research and development purposes.

Keywords: Image quality, public safety, subjective testing

1. INTRODUCTION

A wide variety of factors can impact the way a human observer perceives the quality of an image or video sequence. Algorithms that can account for these various factors and assess image quality are highly desired because they can allow imaging systems to be engineered without the time and expense associated with gathering quality assessments from human viewers. Developing and validating these algorithms usually requires a significant amount of subjective quality data.

This paper presents a new subjective quality dataset known as ITSnoise. The quality factor of particular interest in the ITSnoise dataset is sensor noise. In bright lighting conditions, modern digital imaging sensors can precisely gauge the amount of light that is optically directed toward each picture element, or pixel. In dark lighting conditions, the sensors can still be reasonably accurate, but their precision is greatly reduced. This means that pixels that should be receiving about the same amount of light can instead show significant variations, creating the appearance of “noise” in an image. This noise is random and typically appears unrelated to the desired content of an image, but it may have an effect on how certain textures are perceived. This problem motivates photographers to carefully control their lighting conditions; however, such control is not always possible. For example, if law enforcement officers are conducting surveillance on a person, they will likely prefer not to be seen, which makes it impractical to use flash bulbs or other devices. These public safety applications are of particular interest to the Institute for Telecommunication Sciences (ITS).

¹ The authors are with the Institute for Telecommunication Sciences, National Telecommunications and Information Administration, U.S. Department of Commerce, Boulder, CO 80305.

Sensor noise is present in every image even if it is imperceptible, and there is no way to capture a pristine version of the image without noise. This means that any sensor noise detection algorithm would have to fall into the category of a no-reference (NR) metric because a reference is simply not available. ITSnoise was designed to facilitate the development of such a metric by collecting subjective quality data on images that contain a wide variety of noise levels from very high to very low. While noiseless reference images are not available, the data set comprises scenes captured in multiple ways designed to produce a range of noise levels. The amount and character of the sensor noise is certainly not the only difference between two images from a given scene. One can never truly take the same picture twice. Nevertheless, we believe there will be a benefit to having the image content be as similar as we can make it across various image capture methods (ICMs) that we have used for this dataset. This approach was inspired by the Consumer Content Resolution and Image Quality Dataset (CCRIQ) [2].

The ITSnoise dataset is focused on devices that would typically be seen in public safety applications. Forensics experts are increasingly using three-dimensional imaging technologies that are out of the current scope of our image quality experiments; however, two-dimensional images are still widely used. These images may be gathered by first responders or by civilians. This dataset includes images captured by three low-to-mid-level Single-Lens Reflex (SLR) cameras designed to reflect the equipment that a typical police department would use to gather evidence. The dataset also includes images captured by three consumer devices, two of which are phones, which reflects the kind of images gathered by the public that may also contain important evidence.

The following sections provide details on the ITSnoise dataset as well as subjective data gathered during the ITSnoise experiment.

2. DATASET DESIGN

2.1 Scenes

The ITSnoise dataset is made up 24 scenes, listed and briefly described in Table 1.

Table 1. ITSnoise Scenes

Name	Description
Aloe	Moderately low light with interesting colors
Bench	Simulated surveillance, outdoors, with very low light
Billboard	Moderately low light, outdoors
Boxes	One of the better-lit scenes with a repeating pattern and legible numbers
Cantina	Moderately low light with some visible text
Door	Moderately low light simulated surveillance
Firehouse	Moderately low light and a public safety theme
Garage	Moderately low light and a prominent reflected image
Heavy	Extremely low light which presents significant differences across ICMs, but also some interesting behaviors from some sensors
Kettle	Moderately low light with some particularly interesting textures
Keys	A relatively well-lit scene
Lamp	Moderately low light with an interesting drywall texture
Mail	Moderately low light, simulated surveillance
Mural	Some interesting colors and somewhat problematic lighting
Shed	Very low light with a few large textured regions
Street	Moderately low light, suggestive of surveillance
Strings	Relatively well-lit with some fine details
Tavern	Moderately low light with a near/far composition
Tree	Moderately low light with some interesting color
Truck	Moderately low light, and of interest to the public safety community
Venus	Moderately low light, some interesting and potentially useful textures
Vinyl	Moderately low light, high spatial frequency content in certain directions
Walk	Moderately low light, simulated surveillance from a typical camera angle
Yemiats	Moderately low light with some interesting textures

2.2 Image Capture Methods

Each scene was captured with 12 ICMs. The concept of an ICM is loosely analogous to that of a Hypothetical Reference Circuit (HRC) used in other subjective video quality experiments. The most important difference is that different HRCs are typically given identical input. When different ICMs are used, minor variations are unavoidable even if we do everything we reasonably can to capture the same content.

As previously mentioned, the ICMs used in ITSnoise were designed to reflect the way images are typically gathered in many important public safety or forensic applications. For this dataset we gathered six devices that we believed best represent the range of what would typically be used to capture two-dimensional images. These devices are designated A, B, C, D, E, and F. With each of these devices we were interested in capturing images with the highest level of sensor noise that we could produce under the conditions of each scene. These are the high-noise (HN) ICMs, known as A-HN, B-HN, C-HN, D-HN, E-HN, and F-HN.

Under low-light conditions it can be difficult to produce images without much noise, but we felt these would be useful in developing a noise metric. It is easy to imagine a noise detector which produces false positives in situations where a low-noise image has a texture that coincidentally matches whatever statistical properties the designer associated with noise. To produce a robust dataset that could help avoid such problems, ITSnoise also includes a low-noise (LN) ICM for each device, known as A-LN, B-LN, C-LN, D-LN, E-LN, and F-LN.

For every ICM, JPEG image compression was used. This was chosen largely because it is very rare for users to deviate from this compression standard, and because it was a format that was easy to generate from each device. All the devices used were relatively new (less than three years old) except for C, which has been used consistently, multiple times a week for over nine years. Table 2 includes important details for each ICM.

Table 2. ITSnoise Devices and ICMs

ICM	Device description	General Settings	Specific Settings
A-HN	Canon EOS 90D SLR with EFS 18-55 mm IS STM lens	P mode, AI Focus, image stabilizer on, auto white balance	ISO 25600
A-LN			ISO 100, except for the Bench scene where ISO 400 was used to get a proper exposure
B-HN	Nikon D5600 SLR with AF-P DX NIKKOR 18-55 mm VR lens	P mode, Auto-servo focus, auto white balance	ISO 25600
B-LN			ISO 100, except for the Bench scene where ISO 200 was used to get a proper exposure
C-HN	Canon EOS Rebel T3 with EFS 18-55 mm IS lens	P mode, auto focus, image stabilizer on, auto white balance	ISO 6400
C-LN			ISO 100, except for the Bench scene where ISO 400 was used to get a proper exposure
D-HN	Kodak Pixpro FZ43	Manual mode, auto focus, auto white balance	ISO 1600 except for the Walk scene where ISO 800 was used
D-LN			Lowest ISO setting in the range from 80 to 400 that would allow a proper exposure, depending on the scene
E-HN		Custom exposure	ISO 3200

ICM	Device description	General Settings	Specific Settings
E-LN	Apple iPhone 11 running Yamera		Lowest ISO setting in the range from 32 to 320 that would allow a proper exposure, depending on the scene
F-HN	LG Phoenix 4, LM-X210APM running Footej Camera 2	Manual exposure, auto white balance	ISO 800
F-LN			Lowest ISO setting in the range from 50 to 300 which would allow a proper exposure depending on the scene

3. SUBJECTIVE TEST DESIGN

Subjective data was collected for the ITSnoise dataset by running an absolute category rating (ACR) experiment [1]. To prepare the images for the test, each one was first cropped to a 16:9 aspect ratio in such a way that the center of the image was preserved. Then, each image was resampled to a resolution of 1920×1080 pixels. This was done because the test laptops being used had a display resolution of 1920×1080 pixels. Both the original and resampled images are part of the ITSnoise dataset. Finally, each image was converted to a one frame-per-second video lasting five seconds. In practice, this five-second video was typically only displayed for a little over four seconds, due to quirks of the subjective testing software.

All the images in the ITSnoise dataset were used in testing. The images were distributed among the four test laptops in such a way that all the ICMs were equally represented. A subset of the 24 available scenes was chosen for each laptop. The scenes called Aloe, Door, Heavy, Kettle, Lamp, Strings, Truck, and Vinyl were shown twice as often as the other scenes because ITS researchers judged them to be particularly interesting based on previous subjective testing experience. Therefore, twice as many subject ratings are available for these scenes. Table A-2 in Appendix A gives the number of ratings that were collected for each image. Under these guidelines, 6 ICMs from each of 16 scenes were chosen for each of the four laptops. This means every test subject was shown a total of 96 images.

The avrateNG test software (available at <https://github.com/Telecommunication-Telemedia-Assessment/avrateNG>) was used. This software uses a web browser interface to display a sequence of images and allows the viewer to rate each one by clicking radio buttons. We chose a question template that allowed subjects to skip rating an image if they were not paying attention. Statistics on the use of the skip option are summarized in Table 3. Otherwise, the standard ACR methodology was followed.

Table 3. Skip Statistics

Number of skips used	0	1	2	3	4	5	6	7
Number of subjects	43	10	4	3	0	1	1	1

Data was collected from 63 subjects at the 2022 PSCR Stakeholder Meeting, at the Westin® Hotel & Resorts, in San Diego, CA. We collected ratings from 15 or 16 subjects for each laptop. Meeting participants were informed of the experiment during the meeting sessions and asked to participate. Some subjects entered the room and offered to participate. The remaining subjects were recruited by the test proctor, who approached people on breaks, described the goals of the experiment, and asked for their assistance.

The test environment and methodology were similar to that presented in [2]. The experiment was conducted in a brightly lit, quiet room with a combination of natural and artificial light. This room was adjacent to the main conference room and registration desk. The instructions in Appendix B were made available to each subject, but few of them had the patience to read these instructions. Most participants wanted to begin immediately and attend other portions of the 2022 PSCR Stakeholder meeting. Therefore, the test proctor summarized the instructions

verbally to each subject without reading the script aloud. An attempt was made to provide similar instructions to all subjects.

The experiment did not involve a practice session. The test proctor remained in the room for a few minutes after each new subject began, to answer any questions they had.

Subjects signed a release form that described their rights in human testing. Most subjects provided the following demographic information: gender, age, and experience with public safety. This information is summarized in Table 4.

ITSnoise has an unrepeated scene experiment design [3], due to the differences impacting the photography (e.g., camera placement, scene framing, and potential movement of objects within the scene). These differences have a confounding impact on the main experiment variable, ICM.

Table 4. Subject Demographics

Age						
Age Range	20-29	30-39	40-49	50-59	60-69	70+
Number of Subjects	5	10	12	16	9	2
Gender						
Male	49					
Female	11					
Current or Retired First Responder						
Yes	29					
No	31					

4. MEAN OPINION SCORE ANALYSIS

The individual ratings from each subject are not presented in this memorandum. We have used the ratings to calculate mean opinion scores (MOSs) for each image in the test. These values are presented in Table A-1 in Appendix A. A visual representation of the MOSs is shown in Figure 1.

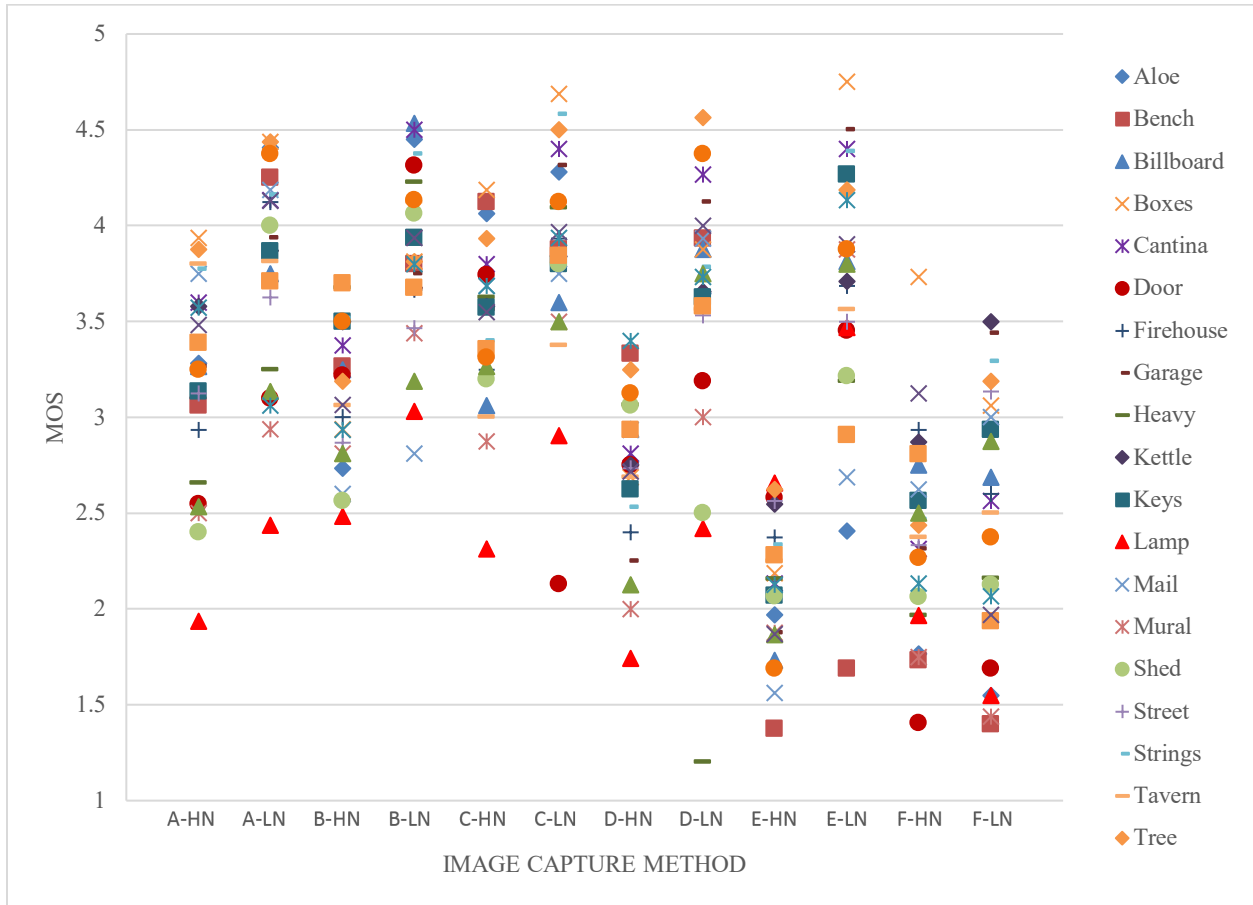


Figure 1. MOS values.

The first general observation we can make from this plot is that the LN ICMs tend to have higher MOS scores than the equivalent HN ICMs, as expected. It should be noted that this difference is not particularly large compared to the range of MOS values within a given ICM. This suggests that the sensor noise level contributes significantly to opinions of the test subjects but is far from the only factor that influences a quality rating. Moreover, while there is a tendency for some image content to generally score near the top and other image content to score near the bottom, there is significant variation among ICMs.

For example, for the SLR-based ICMs, “Lamp” is frequently rated as the lowest-quality image by a significant amount, but in the device-based ICMs, it can rate significantly higher than other content in the test. These variations suggest that there is a complex interaction between the scene content and the noise introduced by each ICM. This is not surprising based on our observations

during the test design. Certain content served to mask or obscure some of the sensor noise, while other content could serve to highlight it.

In summary, the relationship between sensor noise and image quality is not simple, but these data should be useful for studying that relationship and possibly developing automatic noise metrics.

5. REFERENCES

- [1] ITU-T Recommendation P.910, “Subjective video quality assessment methods for multimedia applications,” 2022.
- [2] M. A. Saad, M. H. Pinson, D. G. Nicholas, N. Van Kets, G. Van Wallendael, R. V. Jaladi and P. J. Corriveau, “Image Quality of Experience: A Subjective Test Targeting the Consumer’s Experience,” in *Proceedings of the International Symposium on Electronic Imaging 2016*, 2016.
- [3] M. H. Pinson, “ITS4S3: A Video Quality Dataset With Unrepeated Videos, Camera Impairments, and Public Safety Scenarios,” NTIA Technical Memorandum TM-19-538, U.S. Department of Commerce, National Telecommunications and Information Administration, April 2019. <https://its.ntia.gov/publications/3220.aspx>.

APPENDIX A: MOS VALUES AND NUMBER OF RATINGS

Table A-1. MOS Values

Name	A-HN	A-LN	B-HN	B-LN	C-HN	C-LN	D-HN	D-LN	E-HN	E-LN	F-HN	F-LN
Aloe	3.28	4.41	2.73	4.45	4.06	4.28	2.77	3.65	1.97	2.41	1.77	1.55
Bench	3.06	4.25	3.27	3.80	4.13	3.88	3.33	3.93	1.38	1.69	1.73	1.40
Billboard	3.27	3.75	3.25	4.53	3.06	3.60	2.94	3.88	1.73	3.81	2.75	2.69
Boxes	3.94	4.44	2.94	3.81	4.19	4.69	2.69	3.88	2.19	4.75	3.73	3.06
Cantina	3.60	4.13	3.38	4.50	3.80	4.40	2.81	4.27	2.13	4.40	2.31	2.56
Door	2.55	3.10	3.22	4.31	3.74	2.13	2.75	3.19	2.58	3.45	1.41	1.69
Firehouse	2.93	4.13	3.00	3.67	3.25	3.93	2.40	3.60	2.38	3.69	2.93	2.60
Garage	3.27	3.94	2.56	3.75	3.56	4.31	2.25	4.13	1.88	4.50	2.31	3.44
Heavy	2.66	3.25	3.68	4.23	3.63	4.09	3.06	1.20	2.16	3.19	1.97	2.16
Kettle	3.58	3.87	3.50	3.68	3.58	3.84	2.75	3.66	2.55	3.71	2.87	3.50
Keys	3.13	3.87	3.50	3.94	3.57	3.80	2.63	3.63	2.07	4.27	2.56	2.93
Lamp	1.94	2.44	2.48	3.03	2.31	2.90	1.74	2.42	2.66	3.47	1.97	1.55
Mail	3.75	4.19	2.60	2.81	3.69	3.75	2.93	3.93	1.56	2.69	2.63	3.00
Mural	2.50	2.94	2.81	3.44	2.88	3.50	2.00	3.00	1.88	3.88	1.75	1.44
Shed	2.40	4.00	2.56	4.06	3.20	3.80	3.06	2.50	2.07	3.21	2.06	2.13
Street	3.13	3.63	2.87	3.47	3.31	3.86	2.73	3.53	2.56	3.50	2.33	3.13
Strings	3.77	4.16	3.52	4.38	3.40	4.58	2.53	3.78	2.33	4.39	2.47	3.29
Tavern	3.80	3.81	3.06	4.13	3.00	3.38	2.69	3.56	2.25	3.56	2.38	2.50
Tree	3.88	4.44	3.19	3.81	3.93	4.50	3.25	4.56	2.63	4.19	2.44	3.19
Truck	3.39	3.71	3.70	3.68	3.35	3.84	2.94	3.58	2.28	2.91	2.81	1.94
Venus	2.53	3.13	2.81	3.19	3.27	3.50	2.13	3.75	1.87	3.80	2.50	2.88
Vinyl	3.48	4.13	3.06	3.94	3.55	3.97	2.72	4.00	1.87	3.90	3.13	1.97
Walk	3.57	3.06	2.93	3.80	3.69	3.94	3.40	3.73	2.13	4.13	2.13	2.07
Yemiats	3.25	4.38	3.50	4.13	3.31	4.13	3.13	4.38	1.69	3.88	2.27	2.38

Table A-2. Number of Ratings

Name	A-HN	A-LN	B-HN	B-LN	C-HN	C-LN	D-HN	D-LN	E-HN	E-LN	F-HN	F-LN
Aloe	32	32	30	31	31	32	30	31	32	32	30	31
Bench	16	16	15	15	16	16	15	15	16	16	15	15
Billboard	15	16	16	15	16	15	16	16	15	16	16	16
Boxes	16	16	16	16	16	16	16	16	16	16	15	16
Cantina	15	15	16	16	15	15	16	15	15	15	16	16
Door	31	31	32	32	31	31	32	32	31	31	32	32
Firehouse	15	16	15	15	16	15	15	15	16	16	15	15
Garage	15	16	16	16	16	16	16	16	16	16	16	16
Heavy	32	32	31	31	32	32	31	30	32	32	31	31

Name	A-HN	A-LN	B-HN	B-LN	C-HN	C-LN	D-HN	D-LN	E-HN	E-LN	F-HN	F-LN
Kettle	31	31	32	31	31	31	32	32	31	31	31	32
Keys	15	15	16	16	14	15	16	16	14	15	16	15
Lamp	32	32	31	31	32	31	31	31	32	32	31	31
Mail	16	16	15	16	16	16	15	15	16	16	16	16
Mural	16	16	16	16	16	16	16	15	16	16	16	16
Shed	15	15	16	16	15	15	16	16	15	14	16	16
Street	16	16	15	15	16	14	15	15	16	16	15	15
Strings	31	31	31	32	30	31	32	32	30	31	32	31
Tavern	15	16	16	16	16	16	16	16	16	16	16	16
Tree	16	16	16	16	15	16	16	16	16	16	16	16
Truck	31	31	30	31	31	32	31	31	32	32	31	31
Venus	15	15	16	16	15	14	16	16	15	15	16	16
Vinyl	31	31	31	31	31	31	32	32	31	31	32	32
Walk	14	16	15	15	16	16	15	15	16	15	15	15
Yemiats	16	16	16	15	16	16	16	16	16	16	15	16

BIBLIOGRAPHIC DATA SHEET

1. PUBLICATION NO.	2. Government Accession No.	3. Recipient's Accession No.
4. TITLE AND SUBTITLE ITSnoise: An Image Quality Dataset With Sensor Noise		5. Publication Date
		6. Performing Organization Code
7. AUTHOR(S) Joel Dumke and Margaret H Pinson		9. Project/Task/Work Unit No.
8. PERFORMING ORGANIZATION NAME AND ADDRESS Institute for Telecommunication Sciences National Telecommunications & Information Administration U.S. Department of Commerce 325 Broadway Boulder, CO 80305		10. Contract/Grant Number.
		12. Type of Report and Period Covered
11. Sponsoring Organization Name and Address National Telecommunications & Information Administration Herbert C. Hoover Building 14 th & Constitution Ave., NW Washington, DC 20230		
14. SUPPLEMENTARY NOTES		
15. ABSTRACT (A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here.) This technical memorandum provides details for the image quality subjective experiment ITSnoise. This experiment includes 24 scenes, each digitally photographed using 12 different Image Capture Methods (ICMs) for a total of 288 images. The ICMs were designed to reflect the way images are captured in the most common public safety applications and also to generate a range of different sensor noise levels in the resulting images, from very low to very high. This dataset was intended to provide training and testing data for potential no-reference sensor noise metrics that could automatically predict the impact on perceived quality of sensor noise within a given image. This dataset is freely available for research and development purposes.		
16. Key Words (Alphabetical order, separated by semicolons) image quality; public safety; subjective testing		
17. AVAILABILITY STATEMENT <input checked="" type="checkbox"/> UNLIMITED. <input type="checkbox"/> FOR OFFICIAL DISTRIBUTION.	18. Security Class. (This report) Unclassified	20. Number of pages 25
	19. Security Class. (This page) Unclassified	21. Price: N/A

NTIA FORMAL PUBLICATION SERIES

NTIA MONOGRAPH (MG)

A scholarly, professionally oriented publication dealing with state-of-the-art research or an authoritative treatment of a broad area. Expected to have long-lasting value.

NTIA SPECIAL PUBLICATION (SP)

Conference proceedings, bibliographies, selected speeches, course and instructional materials, directories, and major studies mandated by Congress.

NTIA REPORT (TR)

Important contributions to existing knowledge of less breadth than a monograph, such as results of completed projects and major activities.

JOINT NTIA/OTHER-AGENCY REPORT (JR)

This report receives both local NTIA and other agency review. Both agencies' logos and report series numbering appear on the cover.

NTIA SOFTWARE & DATA PRODUCTS (SD)

Software such as programs, test data, and sound/video files. This series can be used to transfer technology to U.S. industry.

NTIA HANDBOOK (HB)

Information pertaining to technical procedures, reference and data guides, and formal user's manuals that are expected to be pertinent for a long time.

NTIA TECHNICAL MEMORANDUM (TM)

Technical information typically of less breadth than an NTIA Report. The series includes data, preliminary project results, and information for a specific, limited audience.

For information about NTIA publications, contact the NTIA/ITS Technical Publications Office at 325 Broadway, Boulder, CO, 80305 Tel. (303) 497-3572 or e-mail ITSinfo@ntia.doc.gov.