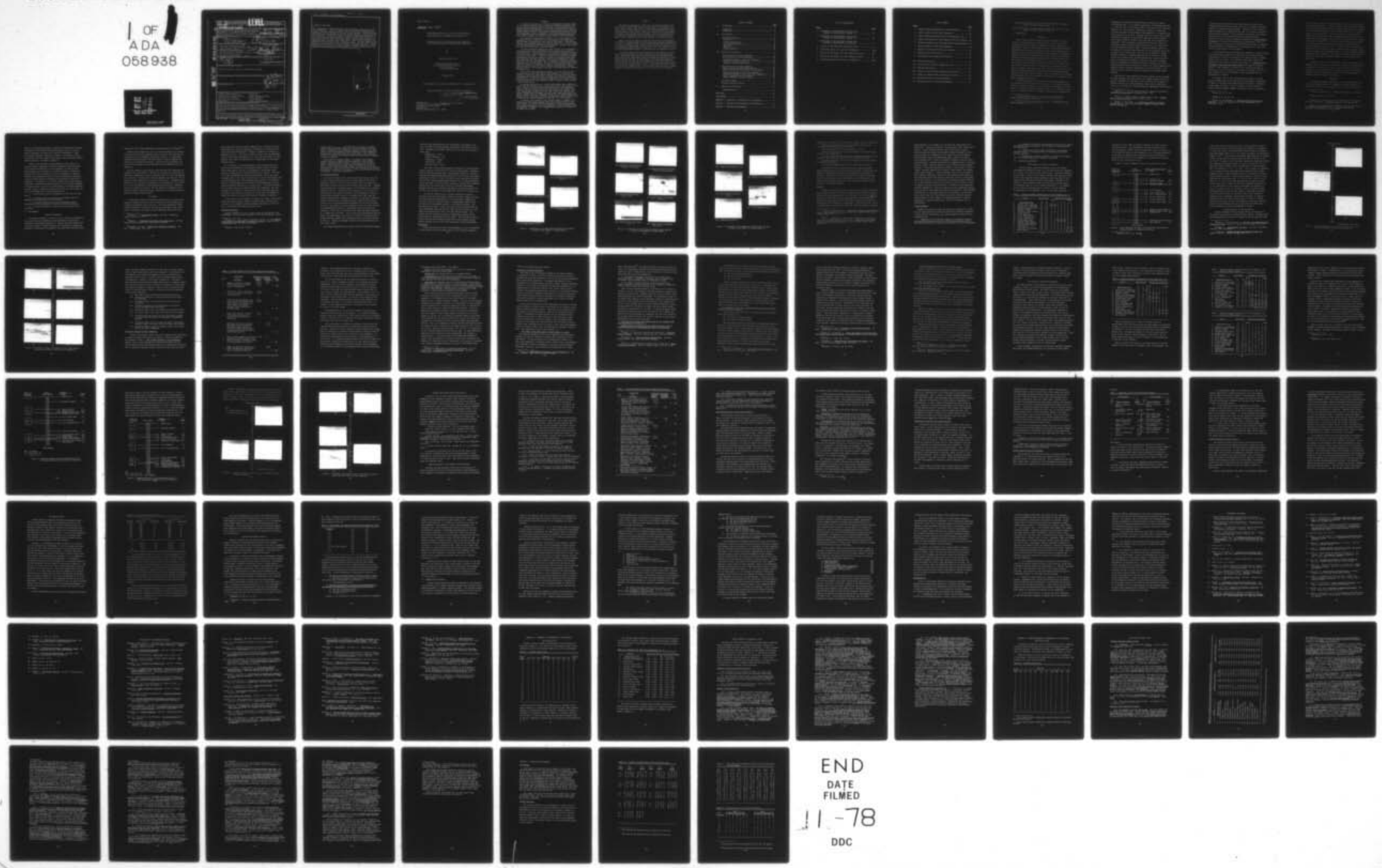
AD-A058 938 FLORIDA INTERNATIONAL UNIV MIAMI SCHOOL OF TECHNOLOGY F/G 17/5
HUMAN FACTORS ASPECTS OF LOW LIGHT LEVEL TELEVISION AND FORWARD--ETC(U)
JAN 78 S DE GROOT AFOSR-77-3242

UNCLASSIFIED AFOSR-TR-78-1237 NL

1 OF 1
ADA
058938



END
DATE
FILMED
11-78
DDC

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

LEVEL

AD A058938

DDC FILE COPY

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR-TR-78-1237	2. GOVT ACCESSION NO. / 3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): HUMAN FACTORS ASPECTS OF LOW LIGHT LEVEL TELEVISION AND FORWARD LOOKING INFRARED SENSOR DISPLAYS. I. A Feasibility Study of Scaled Subjective Complexity of Still Scenes Applied to Computer Image Generation.	5. TYPE OF REPORT & PERIOD COVERED: Final rept. February 1977 — January 1978
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Sybil de Groot Ph.D.	8. CONTRACT OR GRANT NUMBER(s): AFOSR-77-3242
9. PERFORMING ORGANIZATION NAME AND ADDRESS: The School of Technology, Florida International University, Miami, Florida 33199	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F, 2313D9
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research (NL), Air Force Systems Command, Bolling Air Force Base, D.C. 20332	12. REPORT DATE: 1 January 1978
	13. NUMBER OF PAGES: 87
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): 94p.	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Computer Image Generation
Electro-Optical Sensor Displays
Forward Looking Infrared
Low Light Level Television
Human Factors
perception
Sensor Operator Training
Sensor Simulation
visual displays
visual scene-complexity

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Initial research was conducted to investigate human responses to still E-O sensor displays so effects of simulated realism could be evaluated. Research objectives included identifying scenes at different levels of scene-complexity and relating major perceptual with physical factors amenable to computer image generation

Nine subjects psychometrically scaled for scene-complexity photographic sets of Low Light Level Television (L3TV) and Forward Looking Infrared (FLIR) displays

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

410 858

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(Block 20 continued)

of 16 target-areas. After debriefing, subjects matched sensor displays with color photographs. Analysis included tests for significant complexity differences, correlation between scales, and content analyses. Results included synthesized physical measures to quantify sensor displays and contrasting subject responses. Scaling FLIR scenes was reported a different and more difficult task than scaling L3TV displays, however, matching FLIR displays with color photographs was faster with fewer errors, implying FLIR perceptual cues comparable with the graphic detail displayed by L3TV. Findings support a CIG simulation using an optical array of surfaces, edges, and lines with trade-off parameters in the design of a Sensor Simulator determined empirically. Additional evidence indicated that visual parameters of FLIR displays may be more modified by mission assignment and atmospheric variables.

REPORT AFOSR-78-

AFOSR-TR- 78 - 1237

HUMAN FACTORS ASPECTS OF LOW LIGHT LEVEL TELEVISION AND FORWARD LOOKING INFRARED SENSOR DISPLAYS: I.

A Feasibility Study of Scaled Subjective Complexity of Still Scenes Applied to Computer Image Generation.

by

Sybil de Groot, Ph.D.

The School of Technology
Florida International University
The Tamiami Trail
Miami, Florida 33199

1 January 1978

Final Report for the Period 15 February 1977 - 1 January 1978

Approved for public release; distribution unlimited.

Prepared for
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
Directorate of Life Sciences
Building 410
Bolling Air Force Base, D.C. 20332

SUMMARY

The reported research is a heuristic investigation of human responses to still EO Sensor displays so that the effects of degrees of realism in simulated displays could be evaluated. Research objectives included scaling scene-complexity of Low Light Level Television (L3TV) and Forward Looking Infrared (FLIR) displays, identifying scenes at significantly different levels of complexity, determining major perceptual factors associated with sensor displays, and relating perceptual with physical factors amenable to computer image generation (CIG) simulation.

Nine subjects psychometrically scaled a set of photographs of L3TV displays and later a similar set of FLIR displays of 16 target-areas which had been simultaneously video taped at Pre-Sunset. After scaling, subjects were debriefed with their comments submitted to a content analysis from which physical measures were synthesized. Subjects also matched L3TV and FLIR displays with color photographs.

To permit trade-off comparisons of the influence of CIG parameters, analysis of L3TV scaling identified scenes at three and six levels of scene-complexity. Scaling analysis of FLIR displays identified three and four levels with lowered significance requirements increasing the number of scenes at each level. A content analysis of L3TV debriefing comments produced seven descriptors common to a majority of subjects from which four parameters and seven physical measures were synthesized by which each L3TV scene could be quantified. A technique to develop an L3TV/CIG "Property List" was identified. The FLIR content analysis produced seven descriptors common to a majority of subjects, three of which were equivalent to those of the L3TV content analysis and supported a limited quantification of FLIR displays using four of the seven physical measures derived earlier.

Scaling FLIR scene-complexity appeared to be more difficult and a different kind of task than scaling L3TV displays. The FLIR content analysis revealed that some subjects tended to display increased mission orientation and to redefine targets (perhaps due to blurring, smearing, and blending.) Despite extensive prior experience of the subject group with FLIR displays, initial responses tended to regress toward expecting a normal visual scene.

An average correlation of +.72 was found between scene-complexity scales with the variance in one accounting for 56% of the variance in the other. Results of the matching study indicated that most mismatches occurred with less complex scenes and that matching FLIR displays with color photographs was faster and with fewer errors than matching L3TV scenes, implying that perceptual cues and unique target signatures displayed by FLIR may be at a comparable level with the graphic detail displayed by L3TV. Additional evidence indicated that visual parameters may be modified by mission assignment, particularly for FLIR displays.

In general, the findings of this research appear to support a CIG simulation of L3TV sensor displays which uses an optical array of surfaces, edges, and lines but leaves simulation of FLIR sensor displays with unresolved questions about the influences of task variables as well as atmospheric variables. Five recommendations relative to training in EO Sensor use are made.

PREFACE

The reported research was initiated in the Advanced Systems Division of the Air Force Human Resources Laboratory in the summer of 1976 by Dr. Sybil de Groot, then a USAF - ASEE Summer Research Fellow sponsored by the Air Force Office of Scientific Research, under the technical monitorship of Dr. Lawrence Reed, a Task Scientist of the Personnel and Training Requirements Branch of AFHRL. The ten week summer period was sufficient to develop a conceptual framework, to collect data, to present some tentative results, and to organize a preliminary report.

During 1977 Dr. de Groot, an Associate Professor in the Department of Industrial Systems of The School of Technology, Florida International University in Miami, Florida, continued the research with the support of a follow-on mini-grant AFOSR 77-3242, from the Air Force Office of Scientific Research, AFSC, Bolling Air Force Base, D.C. Dr. Alfred R. Fregly was the Program Manager while Dr. Reed was again the Technical Monitor.

At Wright Patterson Air Force Base superb cross-laboratory cooperation and inspiration was provided by members of the Electro-Optical Sensor Laboratory (AFAL): Dr. Bill Lanich (AFAL/RWI) and consultants Dr. Bob Neitman and Mr. Mike Kompar (Mead Technology Laboratories), who provided stimulus materials, and individuals in the Advanced Simulation Branch of AFHRL: Mr. William Foley (AFHRL/ASM), Capt. Bob Orlando, Mr. Mike Nicol, 1st Lt. Mike Ingalls, and Mr. Jim Basinger, who participated in the research as members of a Computer Image Generation team. Thanks are also due Mr. Dwight Miller, an Ohio State University graduate student employed at AFHRL for the summer, for his participation as a subject.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

HUMAN FACTORS ASPECTS OF LOW LIGHT LEVEL TELEVISION AND FORWARD LOOKING INFRARED SENSOR DISPLAYS: I.

A Feasibility Study of Scaled Subjective Complexity of Still Scenes Applied to Computer Image Generation.

I. INTRODUCTION

Background

More is being demanded today of display technology for a wide variety of flight missions, particularly with respect to electro-optical (E/O) sensor displays and computer generated imagery, than ever before. Within a short period of time, Forward Looking Infra-red (FLIR) and Low Light Level Television (L3TV) sensor systems were installed in B-52s to enhance damage-assessment missions.[1] Recently, research was initiated to improve computer image generation techniques simulating E/O sensor displays for such tasks as flight control, reconnaissance, and navigation missions.[2] In line with this, the Air Force Human Resources Laboratory at Wright Patterson Air Force Base is developing a dynamic system for emulating sensor signatures which is based on computer image generation (CIG) techniques to present views of very low, very fast flights. Focusing on Forward Looking Infra-red (FLIR) and Low Light Level Television (L3TV) sensor systems, the planned Sensor Simulator will be functionally interactive with a perceiver, the Sensor Operator-in-training.

A fundamental issue in all simulator design and development is the determination of exactly what is required of a simulator to permit positive transfer of training to later operational environments by the trainee. Simulator development usually starts with a resolution of this issue by decisions based on experimental evidence or on commonly held

---

[1]"Electro-Optical Navigation System Planned for B-52 G/H" Aviation Week and Space Technology, May 10, 1976, pp.130-131.

[2]"Human Resources Lab Accelerates Research", Aviation Week and Space Technology, July 19, 1976, p. 225.

assumptions, such decisions being carefully evaluated for maximal cost-effectiveness. In the present application the unanswered question is: How much like actual airborne L3TV or FLIR displays must the Sensor Simulator be? Here, however, experimental evidence is lacking and opinions vary about the applicability and amenability of assumptions on which decisions can be based. Reynolds revealed a serious lack of human performance information by stating that despite plans for extensive use of ground trainers, getting "hard" data lags the new emphasis on simulators.[3] Stein illustrated the diversity of opinions about what E/O sensor simulators should include by quoting Lt. Col. Lacey of the Air Crew Simulation Division (AFSC), "the using commands would like everything in the visual scene simulated, but as a practical matter and in terms of cost effectiveness, this is just not possible.[4]" Beyond problems indigenous to display simulation per se, Rosell and Willson caution that maximum sensor system performance may be quite unrealizable in an operation environment.[5] Pointing to the lack of adequate models for atmospheric transmission, Rosell and Willson cite several display-degrading factors "beyond the control" of a sensor designer thus, possibly, beyond intelligent deciphering by a Sensor Operator. This raises additional questions about CIG simulation of degraded sensor displays.

Knowing that little empirical data exist relevant to CIG simulation of sensor displays, particularly about E/O display characteristics and operator performance at altitudes as low as 500 feet and at speeds in excess of 400 mph, the author performed a human factors feasibility study to "get a handle on" sensor displays and appropriate CIG simulation.

---

[3]Reynolds, P. "'76 AIAA Visual and Motion Simulation Conference." Astronautics & Aeronautics, July/August 1976, p. 68.

[4]Stein, K.L. "AFSC Stresses Increased Simulator Use." Aviation Week and Space Technology, July 19, 1976, p. 101.

[5]Rosell, F. and Willson, R. Performance Synthesis of Electro-Optical Sensors. Tech. Rep. AFAL-TR-104, WPAFB, Oh., April 1974, pp. 127-137, 141-142.

A long term goal was defined as the identification of those visual parameters which should be and which need not be included in the development of a FLIR/L3TV Sensor Simulator. Support for a human factors approach appeared in an address by Brig. Gen. B.K. Partin, Aeronautical Systems Division deputy.[6,7] While discussing E/O sensor cues, General Partin zeroed in on the research needs for sensor simulator by emphasizing that the effect on operator performance of varying degrees of realism in FLIR and L3TV displays must be established. Thus, in order to focus on the effects of varying degrees of realism in computer generated displays, the assessment of human responses to actual L3TV and FLIR scenes became a prior necessity.

CIG techniques store three dimensional information in terms of edges and surfaces which represent the composition of real-world features. Transformation of that 3D information can be made to produce a dynamic display upon a 2D screen. In a definitive 1975 study, Bunker and Heeschen of General Electric assert that realism is "a function of the complexity of the scene that can be shown" with computer generated imagery.[8] Scene complexity, thus held to be a primary determiner of realism, is also a major factor in specifying CIG hardware, the cost of which depends upon the total number of scene edges which must be programmed for a display.

The present level of CIG instrumentation available to the ASM Branch of AFHRL permits 2,000 potentially visible edges. Observations by members of the ASM Branch support Bunker and Heeschen's statement: as the number of edges in a scene increases, the more "realistic" computer generated imagery appears; conversely, the more "cartoonish"

---

[6] Stein, op. cit. p. 105.

[7] Reynolds, op. cit., p. 70.

[8] Bunker, M. and Heeschen, R. Airborne Electro-Optical Sensor Simulation. Tech. Rep. AFHRL-TR-75-35, AFSC, Brooks AFB, Tx., July, 1975, p. 89.

as the number of edges in a scene decreases.[9] Additionally, the Chief of Simulator Engineering (ASD/ENET), Arthur Doty, is reported as saying that operators experience difficulty detecting heights at low altitudes in the Advanced Simulator for Pilot Training which also has a display based on 2,000 CIG edges.[10] In line with Gibson's findings of the importance of texture gradients in everyday perception of three dimensional space, Doty further suggested that the introduction of texturing might improve imagery.[11,12] At the low altitude flight levels intended for simulation by the FLIR/L3TV Sensor Simulator, a difficulty in detecting heights could be critical. At those times of day when FLIR and/or L3TV sensor displays closely approximate normal daylight visual scenes, texturing would undoubtedly help if the task were TA/TF (terrain avoidance/terrain following). The value of texturing in FLIR or L3TV displays is presently undetermined, however, at other times or for other tasks.

Although recent state-of-the-CIG-art includes as many as 5,000 potential edges, an increase in the number of edges is achieved only at considerably and increasingly higher costs in terms of hardware and programming time. Obviously, some trade-offs must be made which will coincide with those visual parameters which are least important in providing positive transfer of training.

Objectives

To distinguish between appropriate and inappropriate CIG trade-offs in E/O sensor simulation, a comparison of human responses to actual L3TV and FLIR sensor displayed scenes (or target areas) with

---

[9] Foley, W. and Orlando R. Personal communication. July 1976.

[10] Stein, op. cit., pp. 106-107.

[11] Gibson, J.J. Optical Motions and Transformations As Stimuli For Visual Perception. Psychological Review, 1957, 64, 288-295.

[12] Gibson, J.J. Perception of Distance and Space In The Open Air, in Beardslee and Wertheimer (Eds.) Readings In Perception. Princeton: Van Nostrand, 1958, pp. 415-431

those to CIG simulated displays is needed for cost-effective decisions. In addition to the number of edges and surfaces, there are several other CIG parameters available for trade-off consideration. These include edge smoothing, contouring, curvature, the degree of detail, the addition of noise, number of grey levels, ground texturing, foliage masking functions, and weather algorithms.

Based on the conclusion already referenced that realism is a function of scene complexity, the visual parameters which contribute to the subjective experience of scene complexity of L3TV and FLIR sensor displays are the subject of the present study. Provided that differences in scene complexity can first be established, computer hardware and programming trade-offs could be determined by simulating a scene of known subjective complexity and deliberately varying CIG parameters. The adequacy of a simulated display for training purposes could then be evaluated by modeling important visual display parameters at different levels, even with a limited number of CIG edges available. In order to make a comparison between actual FLIR and L3TV displays and later CIG simulated ones, research objectives are:

1. To scale subjectively the complexity of still L3TV and FLIR scenes.

2. To identify scenes of significantly different levels of complexity for comparison and later simulation.

3. To determine empirically the major perceptual factors associated with scene complexity of L3TV and FLIR sensor displays.

4. To relate perceptual with physical factors amenable to CIG simulation.

II. THE APPROACH

Research Assumptions

The present research is, thus, an initial, heuristic investigation into perceived scene complexity of sensor displays with particular referenced to programable parameters in computer generated imagery held to increase realism. Prudence requires that assumptions incorporated regarding a scaling model be simplistic and permit wide latitude in data manipulation for as broad implications and applications as possible.[13,14]

Subjects were assumed to be able to follow directions and to produce interval scales of scene complexity based on L3TV and FLIR displays with variability in their judgements distributed normally. If individuals displayed extreme deviation from group norms, it would be assumed that they were not following instructions. Subjects were also assumed to be aware of the parameters which contributed to their judgements of scene complexity.

Barring evidence to the contrary, L3TV and FLIR scene complexity was assumed to be unidimensional. Subjects were assumed to be replicates of each other, permitting an overall complexity value or score to be computed for each scene. The mean scale value was assumed to be the best value of subjective scene complexity for a given stimulus scene. Assumptions were made that if a scale of scene complexity were developed for a set of L3TV or a set of FLIR displays, scenes differing significantly in complexity could be calculated and further, that the subjective complexity scales would be related to physical scales, measurable in the stimulus materials, producing monotonic relationships.[15]

The Method

In individual sessions nine subjects evaluated sets of sixteen photographs of displays of two E/O sensor systems which had previously been recorded simultaneously onto video tape. The first experimental method involved psychometric scaling of the stimulus photographs along a metathetic or qualitative continuum of scene complexity divided into seven

---

[13] Nunnally, J.C. Psychometric Theory. New York: McGraw-Hill Company, 1967, pp. 31-56.

[14] Edwards, A. Techniques of Attitude Scale Construction. New York Appleton Century Crofts, Inc., 1957, pp.5-9, 172-198.

[15] Biberman, L.M. (Ed.) Perception of Displayed Information. New York: Plenum Press, 1973, p. 43.

equal intervals.[16] This was followed immediately by a debriefing period during which each subject identified or described the criteria he had used to assign scale positions.[17] Subjects' comments are presented in Appendixes A and B. Initially all subjects scaled a set of photographs based on L3TV sensor displayed scenes. Using the same method, the same subjects later scaled an equivalent set of photographs of FLIR sensor displayed scenes.

As a follow-up, a second experimental method required subjects to match color photographs of more extensive daylight views of the same scenic areas with the two sets of photographs of video-taped sensor displays originally used as stimulus materials for scaling. The "real-life" color photographs included areas surrounding each of the sensor displayed scenes. The order of presentation of the two photographic sets of sensor displays were counterbalanced among subjects: five subjects matched L3TV photographs first followed by FLIR photographs while the remaining subjects matched sets in reverse order. The order of presentation of scenes within each matching session was randomized. The time required by each subject to match each set with color photographs was recorded by stop watch and the number of errors noted. Procedures and instructions used in this matching study are presented in Appendix C.

In the first or scaling method, the subject sat before a table across which seven equally spaced, numbered categories were indicated with a covered stack of randomly ordered photographs immediately in front of him. The following directions were read to him.

Scaling Instructions

Arranged randomly in a pile in front of you is a collection of 16 photographs. All of them were recorded at the same time period, Before

[16] Stevens, S.S. Ratio Scales of Opinion in Whitla, D. (Ed.) Handbook of Measurement and Assessment In Behavioral Sciences. Menlo Park, Ca.: Addison-Wesley Pub. Co., 1968, pp. 171-199.

[17] Edwards, A., op.cit., pp. 121-148.

Sunset, with a ___________ (low light level television or forward looking infra-red) sensor system and have been converted from video tapes. Without being too analytical, please sort these 16 pictures into one of the seven equally spaced categories of scene complexity which you see identified before you on the table. Try to find one picture at least for each category making the steps between groupings as equal as possible.

As you can see, Category number 1 is labeled "Least Complex"; Category number 4 is labeled "Average Complexity"; and Category number 7 is labeled "Most Complex". Review all of the pictures in this set before you begin to sort them, spreading them out if you wish, or putting them into simple groups, or however you wish to go about the job. Take as much time as you like, rearranging the pictures in different categories if you choose, until you are satisfied with your arrangement. Do you have any questions?

The Stimulus Materials

Late in the winter of 1976, the Electro-Optical Sensor Lab of the Air Force Avionics Laboratory began recording FLIR and L3TV sensor displays of shots of 16 local "target-scenes" taken from high in the AFAL Tower at Wright Patterson Air Force Base onto video tape. Although the FLIR system was a prototype model, the L3TV system was in standard production and use. This was the first time that 70 to 80 hours of simultaneously video-taped recordings of the performance of these two sensor systems had been taken on a scheduled basis under known and recorded environmental, atmospheric, weather, and sun-angle conditions. Stimulus materials for the present research consisted of two sets of sixteen 8" by 10" glossy-finished photographs of still video displays of the 16 target-scenes, one set consisting of L3TV, the other of FLIR. The photographs were made with a Hasselblad camera and tripod and represent a "first try" at recording stills from video tapes displayed on a "CONRAC QQA 21" monitor. Time constraints precluded maximizing the process for photographic quality and, as a result, the stimulus photographs contained slightly less information than was visually present on the monitor. However, photographic fidelity was judged sufficient for this initial study.

The stimulus photographs were based on L3TV and FLIR sensor displays which had been recorded during one video taping "run" between 19:15 and 19:22 during the time-enveloped classified as Pre-Sunset on May 11, 1976. Atmospheric and weather conditions at the time of the original taping were:

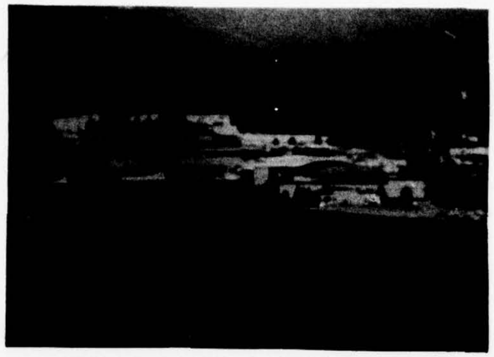
- -Temperature: 65.6^{o}F.
- -Relative Humidity: 32%
- -Visibility: 10 miles
- -Cloud Cover: Clear
- -Category of Day: Hazy

Both L3TV and FLIR sensor systems can present to a Sensor Operator either a wide or a narrow field of view of the environment. Depending upon the particular system, model, and operating conditions, the wide field of view subtends generally between 19^{o} and 23^{o} of arc horizontally and between 14^{o} and 17^{o} of arc vertically at the lense while the narrow field of view subtends approximately one-third of the linear dimensions of the wide field. Since the sensor systems at the AFAL Tower presented displays with the narrow field of view, the stimulus photographs captured views of the environment subtending at the lense between 6^{o}20' and 7^{o}40' of horizontal arc and between 4^{o}40' and 5^{o}40' of vertical arc.

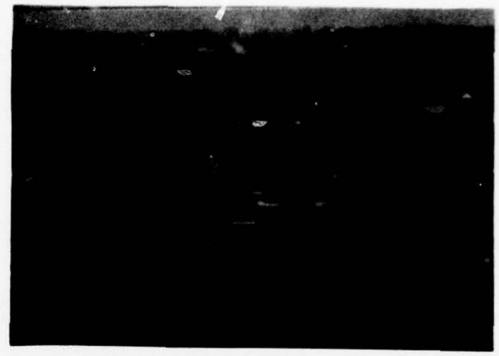
So that the reader can get a general impression of features within the L3TV and FLIR displays, Figures 1 through 3 (pages 15,16 and 17) present daylight photographs of each target scene, numbered and titled as during the original taping. The approximate distance from the sensor to an outstanding feature is cited where available. The pictures within Figures 1,2 and 3 are black and white reproductions of color photographs taken from the 11th floor of the AFAL Tower on a hazy summer morning which were used in the later Matching Study. Each picture has been cropped to present the approximate coverage of the narrow field of view displayed by each sensor. The fact must be kept in mind that during the Pre-Sunset period L3TV and FLIR sensors do <u>not</u> present displays which are the visual equivalent of these photographs.

Debriefings

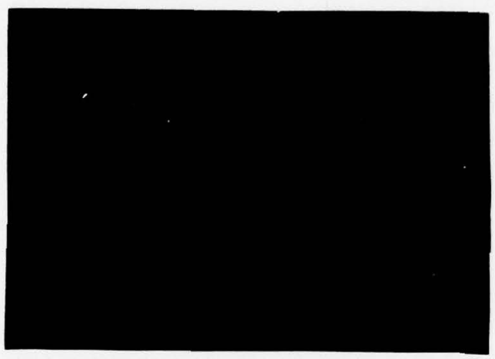
After the subject had scaled all photographs in a set, he was asked to remain seated with the pictures visible in his assigned categories.



#1. Patterson Field Hangars (4 mi)

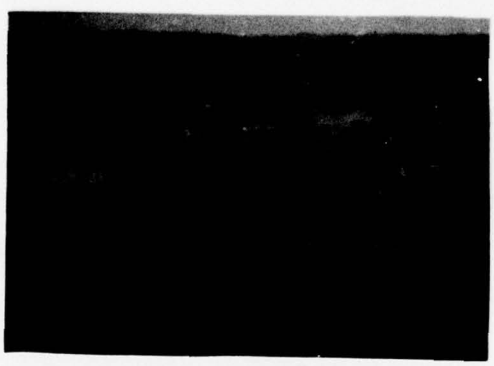


#2. The Steam Plant, Area C (4 mi)



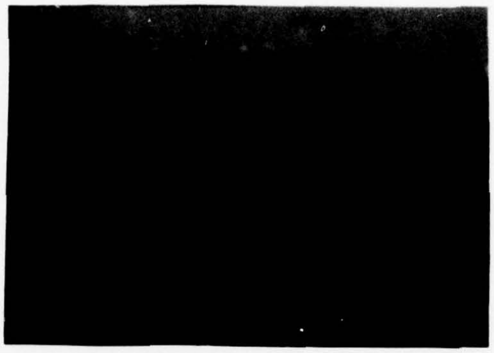
#3. Inactive Atlas Cement Plant (6 mi)

#4. Fairborn Grain Elevator (5 mi)

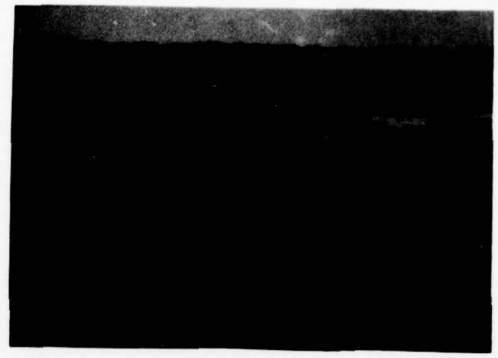


#5. The Base Hospital (3 mi)

Figure 1. Photographs Of The Approximate Display Areas Included In Target Scenes Number 1 Through Number 5

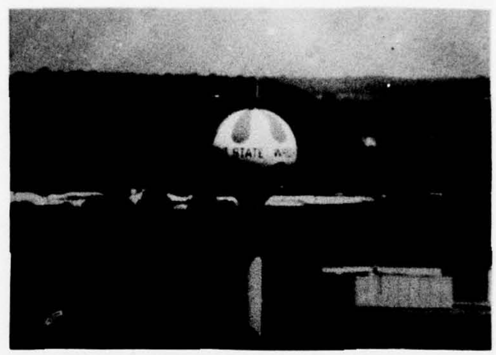


#6. Microwave and Water Towers (distant, on horizon)

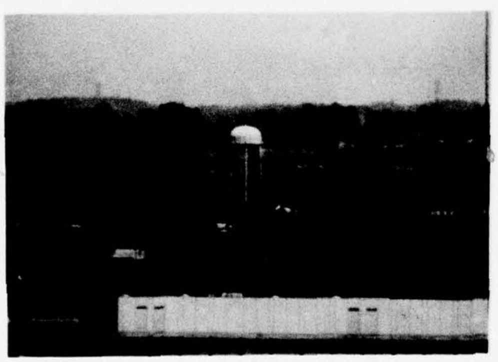


#7. Housing Area With Water Tower (distant, below horizon)

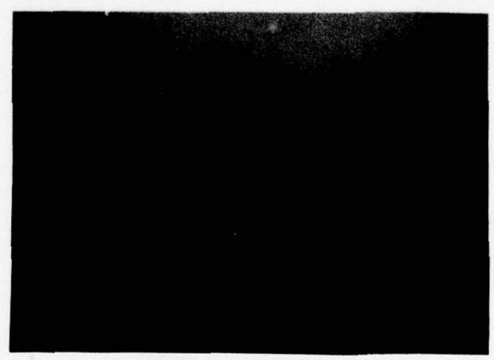
#8. Apartment Complex Among Trees (3 mi)



#9. Wright State U. Water Tower (1 mi)



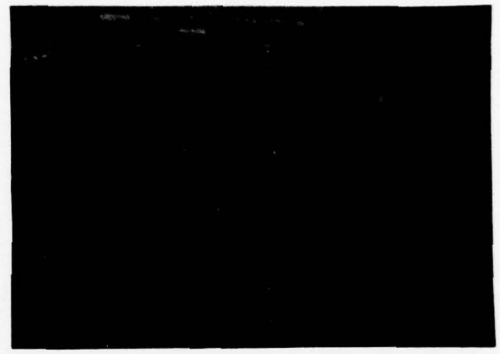
#10. Fairborn Green Water Tower (1 mi)



#11. Wright State U. Microwave Dish (1 mi)

Figure 2. Photographs Of The Approximate Display Areas Included In Target Scenes Number 6 Through Number 11

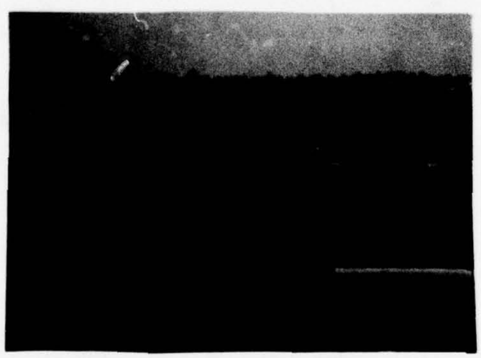
#12. Church In The Woods (1 mi.)



#13. Dayton P. & L. Sub-station

#14. Active Cement Plant Kiln (6 mi.)

#15. I-675 Overpass (2 mi.)



#16. Trebein Site (5 mi.)

Figure 3. Photographs Of The Approximate Display Areas Included In Target Scenes Number 12 Through Number 16

Comments were elicited from the subject by asking a series of questions:

What were you paying attention to or thinking about as you arranged the pictures into these categories?

How did you decide which picture should go where?

What made you decide this picture belonged, say, here (pointing) instead of here (pointing)?

Can you describe the differences you saw between the pictures you placed in lower categories from those in the middle or higher?

What made you decide that a picture belonged at this end instead of the other (pointing)?

Are there any comments you'd like to make about this task?

A pause being taken between each question, subjects usually started responding by the end of the second question. Rarely were all questions asked before the subject began reporting on his scaling criteria. Observations and verbatim notes were recorded of each subject's comments during scaling and debriefing. The essence of these comments are presented in Appendix A for the L3TV photographs and in Appendix B for FLIR photographs.

Subjects

Despite the findings of other researchers that subject groups made up of users, trainees, or trainers of given systems are often the best judges of training requirements, time, distance, and cost precluded the use of such individuals in the present feasibility-oriented study.[18,19] The subject group consisted of 9 individuals selected on the basis of availability and willingness to participate. Included were two engineering psychologists (S-1 and S-2), four ASM engineers engaged in computer

---

[18] Cream, B. and Woodruff, K. Functional Integrated Systems Trainer: Vol. I. Description and Evaluation. Tech. Rep. AFHRL-TR-75-6, WPAFB, Ohio, December 1975

[19] Smode, A. and Hall, E. Translating Information Requirements Into Training Device Fidelity Requirements. Proceedings of The Human Factors Society 19th Annual Meeting. Dallas, Texas, October 1975. pp. 33-36.

image generation (S-3 through S-6), and three E/O sensor experts (S-7 through S-9). Data were originally collected upon a 10th, totally naive subject. Because analysis revealed that he had failed to follow directions and had produced exceptionally divergent responses as displayed in Table A-1 of Appendix A, this subject's data were deleted. This individual scaled most photographs not in terms of increasing scene complexity but in terms of decreasing recognition stating during L3TV debriefing, "I changed my definition of complexity when I found four pictures I didn't know what they were and put them in the most complex category." During FLIR debriefing he stated, "I put scenes (in which) I couldn't recognize anything into category 7 (most complex) and things I could toward category 1." Scenes containing little to be recognized or having minimal information content were thus scaled as most complex by this individual while other subjects placed such scenes at the opposite end of the continuum. One of remaining nine subjects appeared similarly not to follow directions during his scaling of the FLIR scenes in which he assumed a specific tactical mission, as evidenced by his debriefing responses (Appendix B). While his FLIR data was not completely deleted, values found with and without this subject are presented. These subjects' interpretation of scene complexity will be considered in later discussions, as will the fact that although the subject group was visually representative of "user" groups, its representativeness in terms of expectancies, cognitive associations, and use of perceptual cues is indeterminant.

Content Analyses

Because the research included the objective of attempting to determine the major perceptual factors associated with subjective complexity of both L3TV and FLIR displays, two content analyses consisted of the following steps:

1. Examination of the verbatim comments and debriefing responses with preliminary classification into primary and secondary items or Descriptors stressed by the subject as influencing his scaling judgements.

2. Enumeration of primary and secondary scaling criteria adduced by combining similar Descriptors with the number of subjects making such references.

3. Synthesis of content analysis Descriptors into physical parameters useful in CIG applications and derivation of specific physical measures.

4. Consideration of means to develop a validated CIG "Property List" to determine appropriate simulation trade-offs.

III. RESULTS AND DISCUSSION

The Scaling of L3TV Scene Complexity

How each subject judged the complexity of the L3TV display photographs is presented in Table A-1, L3TV Raw Scaling Data (Appendix A), in which cell entries are the scale values or categories assigned to each scene. Noting that low values represent less complexity while higher values represent greater complexity, mere inspection of the table reveals substantial agreement among subjects included in the analysis. First analytical steps are taken when the scale value of the arithmetic mean, $\bar{X}$, and the standard deviation, s, are computed for each scene as displayed below in Table 1, Ordered Frequency Distributions Of L3TV Complexity, N=9. Entries under complexity

Table 1. Ordered Frequency Distributions of L3TV Complexity, N = 9.

Scenes		Scale Values		Complexity Categories						
				least						most
#	Abbrievated Title	$\bar{X}_9$	s_9	1	2	3	4	5	6	7
7	Housing Area (landscape)	1.1	0.3	8	1	0	0	0	0	0
16	Trebein Site (landscape)	1.7	0.9	5	2	2	0	0	0	0
8	Apartment Complex, Trees	1.9	0.8	3	4	2	0	0	0	0
12	Church in Woods	2.2	0.8	1	6	1	1	0	0	0
6	Microwave & Water Towers	2.3	1.2	2	4	2	0	1	0	0
15	I-675 Overpass	3.0	0.7	0	2	5	2	0	0	0
9	W.S.U. Water Tower	3.4	0.7	0	1	3	5	0	0	0
11	W.S.U. Microwave Dish	3.4	0.7	0	1	3	5	0	0	0
13	D.P.&L. Sub-station	3.8	1.3	0	1	4	1	2	1	0
14	Active Cement Plant	4.1	0.8	0	0	2	4	3	0	0
10	Fairborn Water Tower	4.4	0.9	0	0	1	4	3	1	0
4	Fairborn Grain Elevator	5.4	0.5	0	0	0	0	5	4	0
3	Inactive Cement Plant	5.4	1.0	0	0	0	2	2	4	1
2	Steam Plant	5.6	0.9	0	0	0	1	3	4	1
1	Patterson Field Hangars	6.6	0.5	0	0	0	0	0	4	5
5	Base Hospital	6.9	0.3	0	0	0	0	0	1	8

categories are the number of subjects assigning the column value to a particular scene. The distributions enclosed in boxes in Table 1 (and in subsequent tables) identify scenes which could be selected for Guttman-like steps of scene complexity, as in a Scalogram Analysis.[20] On the basis of completely non-overlapping response distributions, three distinct levels of scene complexity are represented by the L3TV scenes #7, #14, and either #1 or #5.

Figure 4, below, graphically depicts the scaling of all L3TV

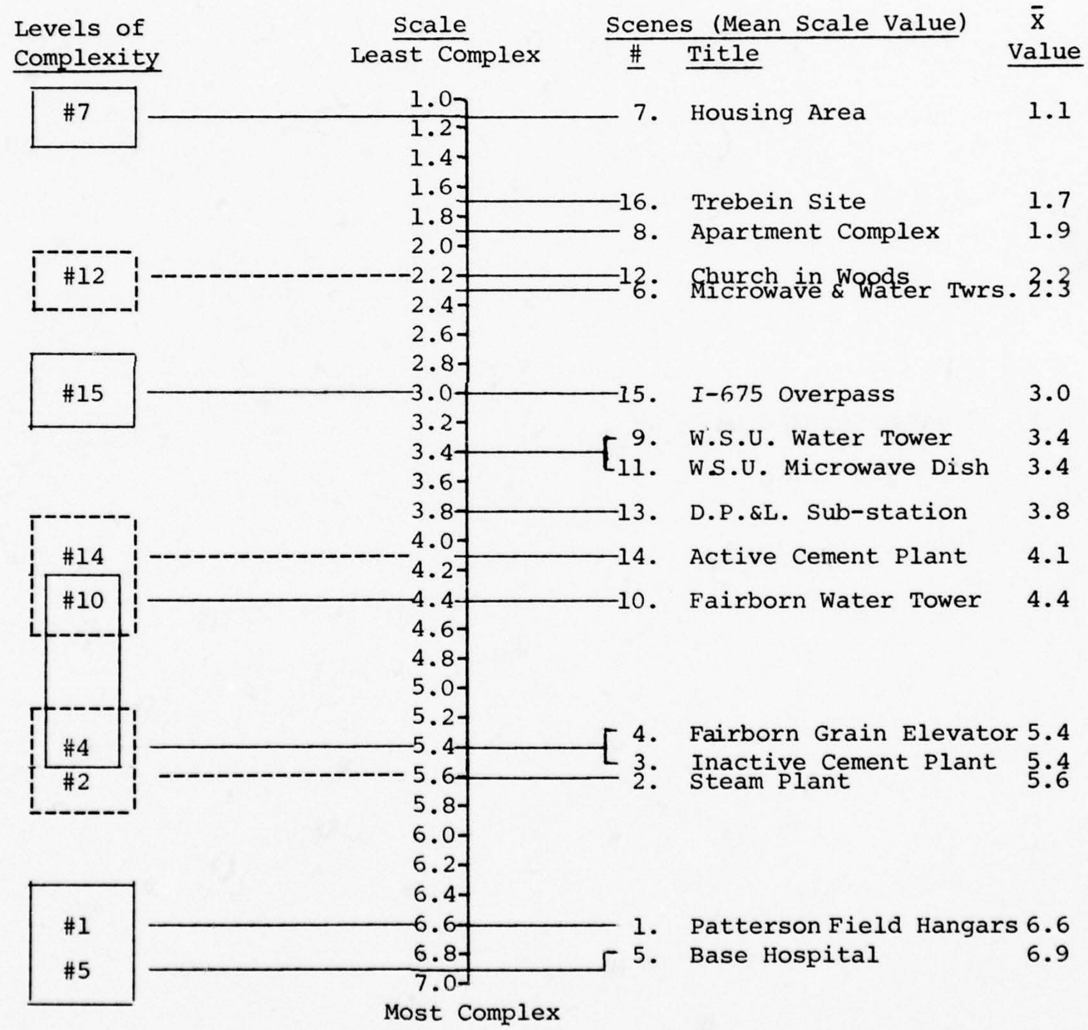


Figure 4. Scene Complexity Scaling of L3TV Scenes With Significantly Different Levels of Complexity Identified

---

[20]Edwards, op.cit., pp. 172-198

scenes with their mean values and titles. Also presented in Figure 4 is information based on statistics presented in Table A-2 of Appendix A. Two-tailed t-tests of the differences between means were performed upon scenes whose ranges of $\bar{X} \pm 1.96$ s barely overlapped, as illustrated in Table A-2.[21,22] With 8 degrees of freedom, four significantly different levels of complexity were distinguished with 99% confidence ($p < .01$; $t \geq 3.355$).[23] In Figure 4 solid line boxes identify by number those scenes which comprise these levels, namely: #7; #15; #10 or #4 and #1 or #5. Six significantly different levels of complexity were distinguished with 95% confidence ($p < 0.5$; $t \geq 2.306$). Thus in Figure 4 in addition to the solidly enclosed scene numbers which make up three of these complexity levels, three additional levels are indicated by enclosure within dashed-line boxes. Scenes typifying these six levels of scene complexity for L3TV displays are: #7; #12; #15; #14 or #10; #4 or #2; and #1 or #5. So that the reader may appraise the changes in scene content across the six levels of L3TV scene complexity significant at $p < .05$, photographs of the relevant target-areas are presented in Figure 5, on pages 23 and 24. Needless to say, the actual stimulus photographs scaled by the subjects, based on video-taped recordings of an L3TV sensor display at a different time of day, lacked the resolution, kind of shadows, figure-ground relations, and relative freedom from noise and edge distortion seen in Figure 5. However, under other sun-angle, weather, and operating circumstances L3TV and even FLIR sensor displays can appear quite like the normal visual scenes captured by a camera and presented in Figure 5.

Content Analysis of L3TV Debriefings

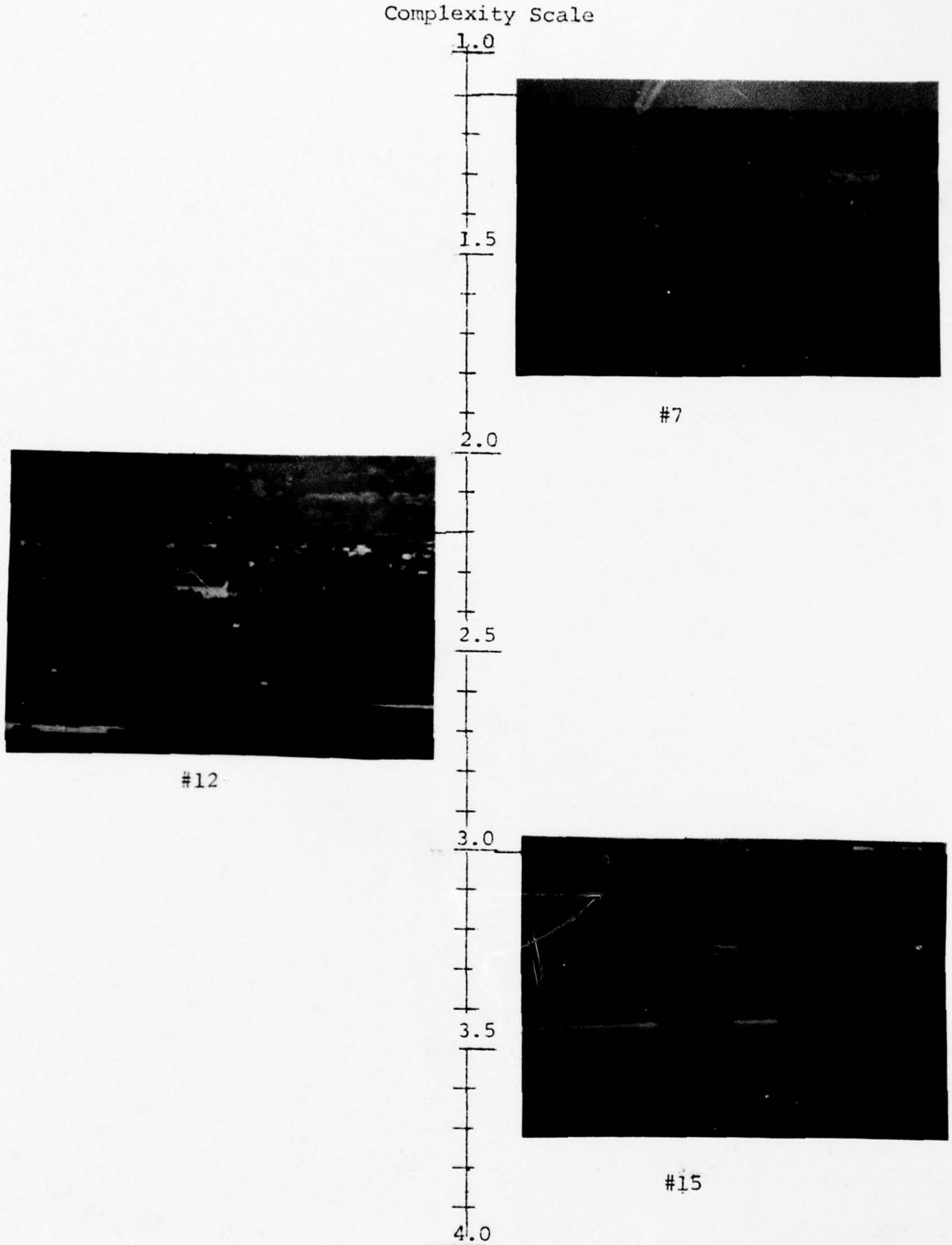
Subjects performed the scaling task soberly and quietly with few subjects volunteering any comments during scaling. Statements made during debriefing are presented in Appendix A with the expressed ideas,

---

[21] Peters, C. and Van Voohris, W. Statistical Procedures and Their Mathematical Bases. New York: McGraw Hill Inc., 1940, pp. 171-176.

[22] McNemar, Q. Psychological Statistics. New York: John Wiley and Sons, 1949, pp. 216-225

[23] Larsen, H. Rinehart Mathematical Tables, Formulas, and Curves. New York: Rinehart and Co., 1949, p. 156



(continued on next page)

Figure 5. Daylight Photographs of L3TV Target Areas at Six Significantly Different Levels of Scene Complexity.

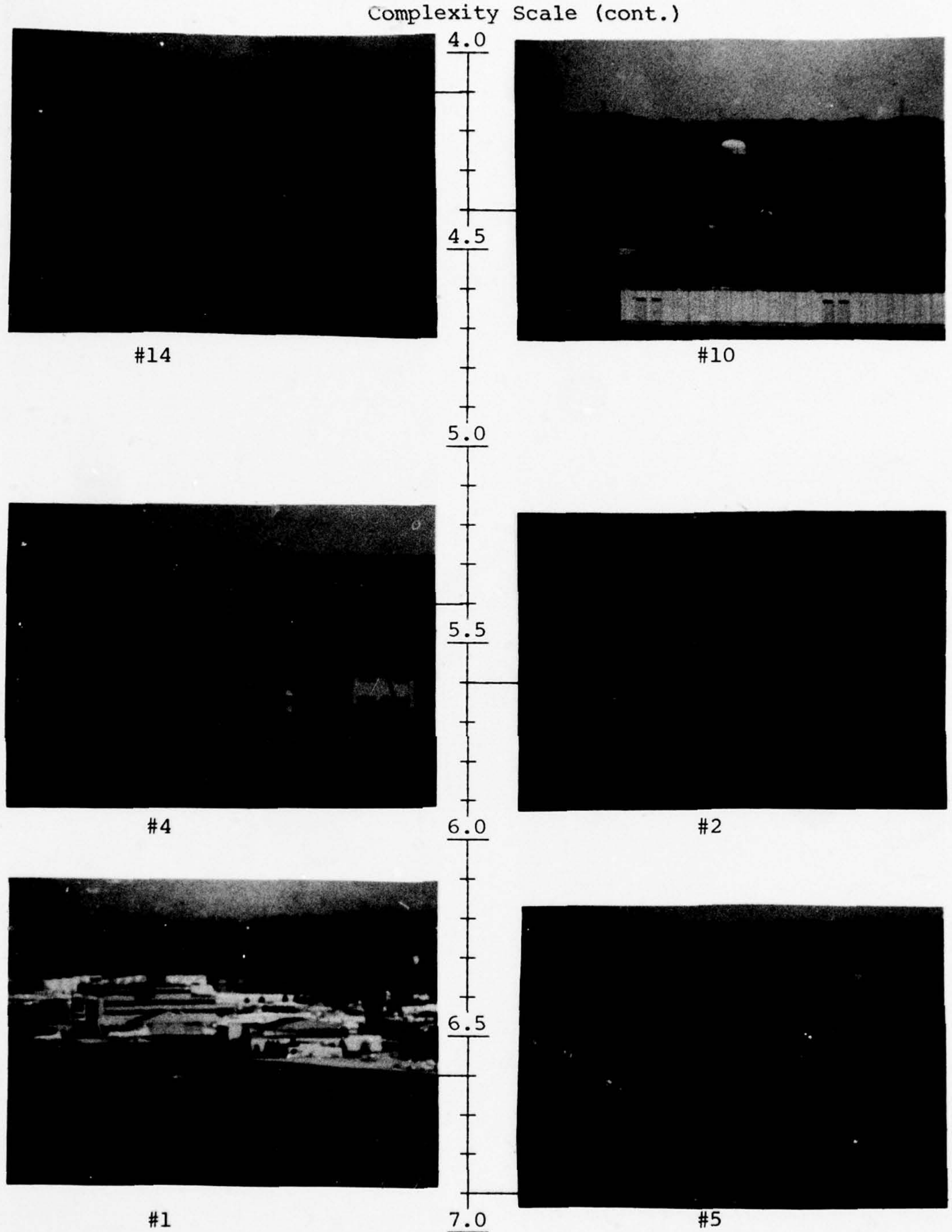


Figure 5. (Continued) Daylight Photographs of L3TV Target Areas at Significantly Different Levels of Scene Complexity.

items, opinions, judgments, and points of view used in scaling (identified in a preliminary content analysis) underlined. Such comments were then separated into those which appeared to serve as primary and as secondary scaling criteria. Primary criteria included concepts which were applied by a subject to the overall continuum or to the manner in which the subject differentiated between extreme ends. Secondary criteria were concepts secondly cited by the subject or mentioned as distinguishing between intermediate categories. From an overall perspective, each subject's reported general concept of increasing L3TV scene complexity can be briefly summarized:

S-1: More broken light distributions revealing more objects.

S-2: Increasing areal amount of complex man-made objects against the surrounds.

S-3: Increasing number and identifiability (revealed detail) of objects.

S-4: Increasing information on man-made objects; increasing percent man-made against the ground.

S-5: Increasing number, size, and complexity of man-made objects.

S-6: Increasing number and variety of man-made objects or shapes.

S-7: Increasing information which equalled increasing crispness (or sharp edges) and detail (or the number of discernable objects).

S-8: Increasing number and area of man-made objects (distinctive geometric shapes) with more finely structured items (detail).

S-9: Increasing number of targets, geometric shapes, or man-made objects of greater complexity.

Descriptors Related To Scene Complexity

Primary and secondary scaling criteria are summarized by Descriptors or similar key words mentioned by the subjects. These are enumerated and combined in Table 2. L3TV Content Analysis of Scene Complexity. Subjects using a given descriptor are identified by number with the last columns indicating the number of subjects, (N), mentioning the descriptor as a primary or secondary scaling criteria and the percent of total

Table 2. Content Analysis of L3TV Scene Complexity Debriefings

Item #	DESCRIPTORS OR KEY WORDS	CITATIONS AS CRITERIA: Primary Subject #s	CITATIONS AS CRITERIA: Secondary Subject #s	TOTAL: N	TOTAL: Per-cent
1	Number of, more, or frequency of man-made objects, cultural detail, (information)	1,2,3, 5,6,7, 8	4,9	9	100%
2	Vegetation, foliage, landscape, terrain, surrounds, background, light uniformity	1,2,4, 5,8,9	6	7	78%
3	Simple to complex geometric patterns, shapes, discriminable objects; distinctive shapes, variety of shapes, outstanding geometric forms, complexity of man-made objects.	4,5,6, 7,8,9		6	67%
4	Area, size, percent, or amount of man-made structures and cultural details relative to the ground.	2,4,5, 9	7,8	6	67%
5	Detailed or fine type structure; directions of lines of objects; detectable detail; supporting structures; sub-structures within primary structure; sub-objects; complex sub-structure	8	1,3,4, 5,6	6	67%
6	Black to white changes; clarity (of surrounds); clearest scenes; sharp edges; crispness, resolution; vague to distinct objects or targets.	4,7,9	1,2	5	56%
7	Number and different directions of lines, straight and curved lines; edges (of detectable details).	7	1,2,5, 9	5	56%

subjects. Only concepts mentioned by a majority of subjects are included in the analysis which is not a frequency count of how many times a given word was used, but rather, how many people in the subject group used it in describing how they performed the scaling task.

To all seven subjects who specifically mentioned it, the background (e.g. terrain, landscape, foliage, etc.) appeared equivalently unimportant. Only one subject implied a possible masking function. Unbroken lightish areas were merely trees to S-1 while two subjects, S's 4 and 6, described the vegetation/background as pretty much the same across all scenes (non-informational). Three subjects, S's 5, 8, and 9, clearly identified scenes visually comprised of only terrain or landscape as least complex while S-2 referred to the "garbage of the surrounds". The two remaining subjects used the word "clutter", but with opposite meanings! S-7 referred to possible targets lost in "clutter" in less complex scenes while S-3 referred to the "clutter" of objects in the most complex.

Unidimensionality of the L3TV scene complexity continuum was questioned only indirectly by two subjects: S-2, thought a different ordering might have resulted had he been asked to scale on his secondary criteria, "clarity", while S-7 mentioned using some conflicting criteria which might not exist on the same scale.

The relationship of a complexity continuum to visual aspects of mission-oriented tasks appeared to be questioned tangentially by two subjects. Pointing out that he focused on central portions and did not scale on background, S-4 suggested that perhaps his ordering should be reversed "if identification is not related to scene complexity." S-7 described his middle-scaled scenes as having "a good probability of hits if this is a recognition problem because only one thing stands out." Interestingly, despite the familiarity of all nine subjects with the ultimate training purposes of the Sensor Simulator under development, use of mission-oriented or task performance words occurred surprisingly infrequently during debriefing. For example:

"Target" was used by two subjects, #s 7 and 9, the former also referring to "hits" and "false alarms."

"Recognition" was stated only by S-7, as reported above.

"Identification" was mentioned by two subjects: S-4, as already cited, and S-9 describing the "identifiability" of mid-scaled target/objects (similar to S-4's point with recognition).

"Detection" was referred to by S-3 in describing mid-scaled scenes as having "objects I could pick out, that is, the detail I could detect." (Indirectly, the detection of detail in objects already visually established may have been implied by subjects #s 5,6, and 8 as "sub-objects", "sub-patterns", "sub-structure", "detailed type of structure", or "fine structure" in scenes scaled from middle to most complex).

If the author had held any preconceived notions, the first would have been that subjects would relate informational content to scene complexity. Yet "information" was used specifically by only three subjects (33%) in slightly different contexts. While S-7 clearly equated complexity with information mentally, he also wondered at scale unidimensionality stating that "information seemed sorta equal to crispness and detail, sharp edges and a number of discernable objects." While S-4 referred to less complex scenes as having "no particular information" with the exception of some simple geometric patterns in two of them (number of objects), S-6 in discussing differences among his mid-scaled scenes pointed out that scenes #9 and #11 did not have the "variety of information" as did scenes #2 and #4 with more complex sub-structures (detail). Thus, it is debatable for Table 2 whether "information" should be included in Item 1, referring to the number of objects (S's 7 and 4), or in Item 3, referring to object detail (S's 7 and 6). In either event, the net tallying results would have been the same. An unmet expectation supported by the literature was that brightnesses, the "luminous output pattern", or contrast levels would be mentioned frequently.[24] Instead, only one subject (S-1) spoke of "broken light uniformity" and another (S-7) referred briefly to resolutions (or, in general, to acutance). Other subjects immediately focused on concrete or real features with

[24] Snyder, H. Image Quality and Observer Performance. Ch. 3 in Biberman (Ed.) Perception of Displayed Information. New York: Plenum Press, 1973, pp. 87-118.

100% of the subjects mentioning objects

Conversion to Physical Measures

Each item or set of Descriptors adduced in the content analysis and presented in Table 2 could be converted directly into a physical parameter by which each photograph in the set of L3TV display photographs may be measured and quantified. For maximum applicability to the development of a L3TV/FLIR Sensor Simulator, however, physical parameters amenable to a CIG format and expressed in dimensions meaningful to programmers must be synthesized from the seven content analysis items. Synthesis into physical measures will, per force, be feasibility-oriented because the data here summarized were limited to a single stimulus set. A much greater data base would be required for a definitive conversion which would satisfy the training requirements of a variety of flight missions under varying circumstances.

In the entire stimulus set of L3TV photographs there were no natural terrain features or foliage objects which stood out as "figures" in the normal "figure-to-ground" relation. Similarly, with the exception of a few possibly leveled grassy patches, there were no alterations of the terrain or geographical man-made features such as cultivated fields, orchards, local roads, or fences visible with the exception of power lines and patches of an interstate highway. Had such elements been visible, it is probable that they would have been included as cultural features under "man-made" objects or shapes" as were the power/telephone lines and concrete roads.[25] Thus the first physical parameter, based on Item 1 and responses from 100% of the subjects, is:

I The number of man-made objects or object-areas in a scene.

This physical parameter involves a mere numerical counting process following an explicit definition of exactly what constitutes a "man-made object." Gestalt principles on the properties of figures (e.g. form, solid appearance, structured contour, surface color, and wholeness) plus "grouping" factors of incomplete or partially revealed patterns (proximity, similarity, good continuation, perseverance, etc.) would

[25] Reed, S.K. Psychological Processes in Pattern Recognition. New York: Academic Press, 1973, pp. 11-53, 223-226.

make a good beginning.[26] Each separate figure, no matter how large or small, even if partially obscured by interpositioned features would be counted as one.[27] Because figures are not necessarily the same as "man-made objects",[28] two derived measures are:

1) the number of separate, single entities, ignoring sub-structural aspects, attachments or features (such as windows which are smaller units contained within an identifiable figure).

2) the number of single entities including sub-structural aspects provided the smaller features can be recognized as completed units or distinct parts of the larger entity.

Item 4 (area of man-made objects) is related to Item 2 (terrain, foliage, etc.) but not necessarily inversely because the amount of sky visible in scenes varied. Somewhat curiously, the sky was never cited during debriefing as a scaling determiner. By overlooking the sky, it is probable that subjects in this experiment "lumped" or generalized the sky area in with foliage/terrain as background in this stimulus set. Yet it is also probable that with scenes containing important terrain avoidance features, the amount of visible sky would achieve greater significance, particularly in a TA task. To synthesize Items 2 and 4, incorporating responses from 89% of the subjects, the following two aspects of the same parameter are suggested, with the first limited to the present stimulus set while the second is expected to have broader applicability with a greater data base:

II <u>The ratio of the area of man-made objects to the combined areas of foliage/terrain and sky area</u>.

IIa <u>The percentages of total scene area comprised by the area of man-made objects; of the foliage/terrain area; of the sky area</u>.

---

[26] Graham, C.H. Visual Form Perception, in Graham (Ed.) <u>Vision and Visual Perception</u>. New York: Wiley & Sons Inc., 1965, pp. 548-567.

[27] Braunstein, M. <u>Depth Perception Through Motion</u>. New York: Academic Press, 1976, pp. 13-30, 41-56, 154-186.

[28] Posner, M. Coordination of Internal Codes, in Chase (Ed.) <u>Visual Information Processing</u>. New York: Academic Press, 1973, pp. 35-74.

The following three measures, necessary to compute the above related parameters, could be achieved simply by placing an appropriately fine transparent grid over a scene photograph and counting areal units of:

3) the area of man-made objects.

4) the area of terrain/foliage.

5) the area of sky.

Item 3 of Table 2 incorporates shape or outline complexity while Item 5 deals with interior structures or patterns. Both items deal with aspects of object detail and are related to Item 7 (number, kind, and direction of lines) not only because visual detail is graphically displayed by the presence of detectable lines, edges, or quick gradient changes but also because edge, slit, and line detectors monitor specific regions of the retina.[29] A third physical parameter, incorporating responses of 100% of the subjects, is a measure which can be achieved directly, most efficiently at the time measures #1) and #2), above, are being taken and recorded:

III-The number of straight and curved lines which make up visible portions of man-made objects.

Two measures, most meaningful to CIG programmers which reflect this third parameter are

6a) the number of external edges

6b) the number of internal lines

Several slightly different descriptors, including some adjective modifiers, compiled from 56% of the subjects are presented in Item 6 . The common thread running through these descriptors is that they appear to refer to scene quality or image quality. Subjectively, image quality of real-life scenes is influenced by many factors such as scene content, overall brightness, contrast levels, texture gradients, assigned task, noise, perceptual cues, and expectancy (to mention a few), while determining physical parameters have been identified as gray scale, modulation transfer functions reflecting resolution, signal to noise

---

[29] Lindsay, P. and Norman, D. Human Information Processing. New York: Academic Press, 1972, pp. 1-10, 94-113.

ratios, uniformity, geometric distortion, aliasing, luminance, image size, spread functions, etc.[30] Yet the traditional psychophysical approach tends to put first emphasis on resolution,[31] as do many current optical, photographic and electronic display approaches.[32] The human ability to resolove a visual target is often identified as visual acuity, which is measured in terms of the smallest resolvable angle subtended between disparate points, lines, or edges of symbols and patterns or in terms of the smallest linear extent of the lesser dimension of a resolvable shape.

Although stimuli in the present experiment were photographs of visual scenes transmitted through an L3TV system and recorded onto video tape, with concomitant diminution of resolution resulting from the connection of components or sub-systems in series, system variables remained fairly constant. That which varied most markedly across the stimulus set was scene content. Theoretically, one could expect the loss of resolution due to system factors to be fairly equivalent for each scene. Several researchers have pointed out, however, that both human resolution and system resolution will vary as a function of scene content. Human resolution, expressed as a minimum visual threshold modulation curve has been found to vary as a function of the geometry of the test pattern while varying system resolution has been noted as a function of the number and length of lines and different geometries of test charts.[34] Therefore the following fourth parameter is recommended:

---

[30] Biberman, L.M. (Ed.) Perception of Displayed Information. New York: Plenum Press, 1973, pp. 13, 35.

[31] Rosell, F. and Willson, R. Recent Psychophysical Experiments and the Display Signal to Noise Concept, in Biberman (Ed.) op.cit., pp. 167-231.

[32] Biberman, L., ibid., pp. 313-322.

[33] McCormick, E. Human Factors in Engineering and Design. New York: McGraw Hill Book Company, 1976, pp. 63-65

[34] Biberman, L., op.cit., pp. 21, 76-84.

IV-Scene quality as reflected by resolvable detail.

Since the visual acuity of the subjects exceeded system resolution as revealed in the set of stimulus photographs, a logical first approach to scene or image quality would be to measure the smallest extent of lines or objects in each scene which are recognized as such. Therefore, the following measures are recommended as a start:

7a) The shorter dimension of the smallest object counted in a scene during measure #2), above, or

7b) The length of the shortest line counted in measure #6), above.

Toward an L3TV/CIG Property List

Because the set of original stimulus materials were not available for this subsequent research, the above measures and the suggested analysis which follows could not be performed. However, due to the limited generalizability of present research findings, actual implementation of these concepts in this feasibility study is less important than the designation of means to achieve further ends in the design of a Sensor Simulator.

Rather than taking at face value the physical parameters and their constituent measures synthesized from the content analysis, each measure should be validated against L3TV scene complexity scale values derived earlier. This could be accomplished by pairing the value found for each scene on the seven derived measures with the mean L3TV scene complexity score for that scene and performing a correlational analysis.

Several correlational statistical techniques are available which range in complexity from simple two-dimensional plots (with average scene complexity on one axis, a physical measure on the other, and each plotted point representing a scene from the set of 16) which are "eyeballed" to determine the most successful measures; through correlation to regression analysis; to multiple correlation.[35,36] In general, only measures found to have moderate to high positive correlation coeffi-

---

[35] Winer, B. Statistical Principles in Experimental Design. New York: McGraw Hill Book Co., 1962, pp. 575-621.

[36] Hays, W. Statistics for Psychologists. New York: Holt, Rinehart and Winston, 1963, pp. 490-538.

cients or good predictive relationships should be retained. By such a culling, an initial L3TV/CIG "Property List" could be established indicating which CIG parameters are most important for simulation and, hence, must be included in the Sensor Simulator to maximize transfer of training.

The Scaling of FLIR Scene Complexity

The same procedures were followed as during L3TV scaling for scene complexity, however two features distinguished the FLIR experimental sessions from the earlier ones. The first, which will be discussed later, was that three times as many subjects as previously made voluntary comments while scaling the FLIR photographs. The second feature was that one subject, S-7, provided two different sets of scale values for the FLIR photographs. After debriefing when the Experimenter observed that the Subject had scaled the photographs as if he were assigned a particular flight mission, the Subject stated that although he had assumed the tactical mission of "reconnaissance strike utility", he could rescale the FLIR photographs. The Experimenter expected the Subject to re-assume the posture taken during earlier L3TV scaling; however, the subsequent debriefing revealed that during his second scaling of FLIR photographs S-7 had assumed a strategic mission, that of "low level ingress". At least for this subject, FLIR scene complexity appeared to be a function of specific flight missions. Questions of whether this subject was following instructions or if his data should be excluded in the analyses resulted. Since six other subjects had volunteered comments which generally reflected a different nature of the scaling task, both of S-7's scalings were included in initial examination. Table B-1 in Appendix B presents the complexity category into which each subject placed photographs of the 16 FLIR displays, with S-7's first scaling (tactical mission) and second scaling (strategic mission) identified under columns S-7a and S-7b respectively.

Ordered frequency distributions of FLIR scene complexity judgments, their means, and standard deviations, are presented in three tables which follow. Omitting S-7's data entirely, Table 3 is based on data of the remaining eight subjects. The distributions enclosed in boxes in Table 3 reveal three Guttman-like levels of scene complexity based on non-overlapping distributions.

Table 3. Ordered Frequency Distributions of FLIR Complexity, N=8

Scenes		Scale Values		Complexity Categories						
				least						most
#	Abbreviated Title	$\bar{X}_8$	s_8	1	2	3	4	5	6	7
7	Housing Area (landscape)	1.0	.0	8	0	0	0	0	0	0
8	Apartment Complex, Trees	1.6	.7	4	3	1	0	0	0	0
12	Church in Woods	2.5	.8	0	5	2	1	0	0	0
16	Trebein Site (landscape)	2.9	.8	0	3	3	2			
6	Microwave & Water Towers	2.9	1.0	1	1	4	2	0	0	0
10	Fairborn Water Tower	3.0	.8	0	2	4	2	0	0	0
9	W.S.U. Water Tower	3.4	.7	0	1	3	4	0	0	0
15	I-675 Overpass	3.8	1.3	0	2	1	2	3	0	0
11	W.S.U. Microwave Dish	4.1	1.0	0	0	2	4	1	1	0
13	D.P.&L. Sub-station	5.1	1.4	0	0	1	2	1	3	1
3	Inactive Cement Plant	5.5	1.3	0	0	1	0	3	2	2
2	Steam Plant	5.6	1.2	0	0	0	2	1	3	2
14	Active Cement Plant	5.9	.8	0	0	0	0	3	3	2
4	Fairborn Grain Elevator	6.0	.8	0	0	0	0	2	4	2
1	Patterson Field Hangars	6.1	1.1	0	0	0	1	1	2	4
5	Base Hospital	6.1	1.1	0	0	0	2	1	3	2

Table 4 incorporates S-7's first or "tactical mission of reconnaissance strike utility" scaling data which appear in parentheses. In Table 4 two sets, each with two non-overlapping levels of complexity, are identified. One set is boxed-in by solid lines, the other by dashed lines. Discrepancies between S-7's first scaling values and those of other subjects, while noticeable on some scenes, did not appear sufficiently large or often to discard S-7's data arbitrarily at this point in the analysis.

Table 5 includes S-7's second or "strategic mission of low level ingress" scaling data with his judgments, again, in parentheses. No

Table 4. Ordered Frequency Distributions of FLIR Complexity, N=9
With S-7's First Scaling.

Scenes		Scale Values		Complexity Categories						
				least						most
#	Abbreviated Title	$\bar{X}_T$	s_T	1	2	3	4	5	6	7
7	Housing Area (landscape)	1.1	.3	8	(1)	0	0	0	0	0
8	Apartment Complex, Trees	1.7	.7	4	(4)	1	0	0	0	0
12	Church in Woods	2.4	1.1	0	5	2	1	(1)	0	0
10	Fairborn Water Tower	2.9	.8	0	(3)	4	2	0	0	0
16	Trebein Site (landscape)	2.9	.8	0	3	(4)	2	0	0	0
9	W.S.U. Water Tower	3.1	1.1	(1)	1	3	4	0	0	0
6	Microwave & Water Towers	3.1	1.2	1	1	4	2	(1)	0	0
15	I-675 Overpass	3.6	1.3	0	(3)	1	2	3	0	0
11	W.S.U. Microwave Dish	4.1	.9	0	0	2	(5)	1	1	0
13	D.P.&L. Sub-station	5.2	1.3	0	0	1	2	1	(4)	1
2	Steam Plant	5.3	1.4	0	0	(1)	2	1	3	2
5	Base Hospital	5.6	2.0	(1)	0	0	1	1	2	4
3	Inactive Cement Plant	5.7	1.3	0	0	1	0	3	2	(3)
14	Active Cement Plant	5.7	1.0	0	0	0	(1)	3	3	2
1	Patterson Field Hangars	5.8	1.5	0	0	(1)	1	1	2	4
4	Fairborn Grain Elevator	5.8	1.3	0	0	0	(1)	2	4	2

Table 5. Ordered Frequency Distributions of FLIR Complexity, N=9
With S-7's Second Scaling

Scenes		Scale Values		Complexity Categories						
				least						most
#	Abbreviated Title	$\bar{X}_S$	s_S	1	2	3	4	5	6	7
7	Housing Area (landscape)	1.7	2.0	8	0	0	0	0	0	(1)
8	Apartment Complex, Trees	2.2	1.9	4	3	1	0	0	0	(1)
12	Church in Woods	3.0	1.7	0	5	2	1	0	0	(1)
6	Microwave & Water Towers	3.1	1.2	1	1	4	2	(1)	0	0
10	Fairborn Water Tower	3.2	.8	0	2	4	(3)	0	0	0
16	Trebein Site (landscape)	3.2	1.9	0	3	3	2	0	(1)	0
9	W.S.U. Water Tower	3.4	.7	0	1	3	(5)	0	0	0
15	I-675 Overpass	4.0	1.4	0	2	1	2	3	(1)	0
11	W.S.U. Microwave Dish	4.2	.9	0	0	2	4	(2)	1	0
3	Inactive Cement Plant	5.1	1.7	0	(1)	1	0	3	2	2
13	D.P.&L. Sub-station	5.2	1.3	0	0	1	2	1	(4)	1
14	Active Cement Plant	5.3	1.8	(1)	0	0	0	3	3	2
2	Steam Plant	5.4	1.2	0	0	0	(3)	1	3	2
4	Fairborn Grain Elevator	5.6	1.5	0	(1)	0	0	2	4	2
1	Patterson Field Hangars	5.7	1.7	0	(1)	0	1	1	2	4
5	Base Hospital	5.8	1.5	0	0	(1)	1	1	2	4

Guttman-like steps of scene complexity can be found among the ordered distributions in Table 5. Additionally, S-7's second scaling values frequently appear markedly different from those of the other eight subjects. Hence, data from S-7's second scaling were eliminated from further analysis.

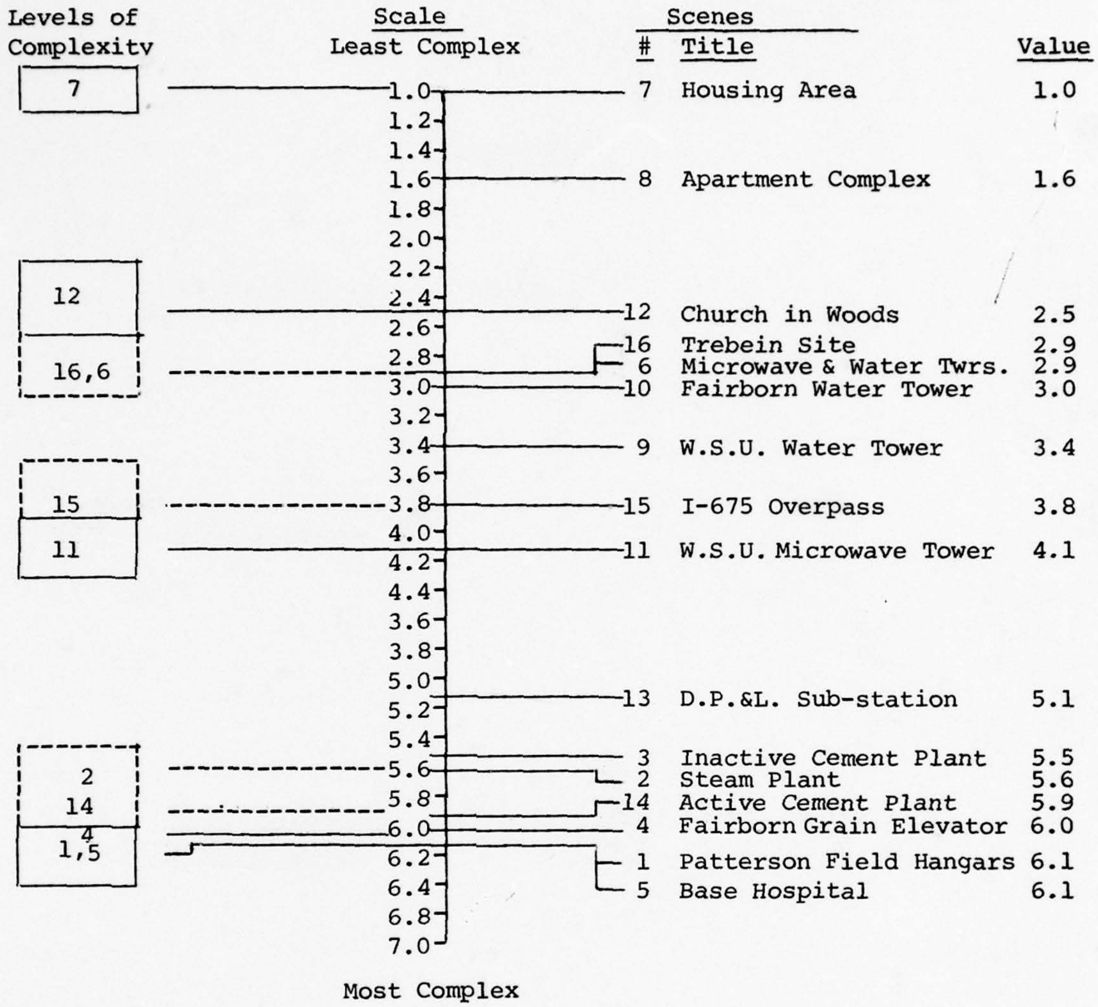
Statistics for between-scene comparisons of scene complexity are presented in Table B-2 in Appendix B for the N = 8 data presented in Table 3 with subject S-7 excluded and for the N = 9 data presented in Table 4 with S-7's first scaling judgments included. As in the earlier treatment of L3TV scenes, two-tailed t-tests of the differences between mean complexity were performed upon scenes whose ranges of $\bar{X} \pm 1.96$ s barely overlapped. Based on the data of the N = 8 group (with S-7 excluded), Figure 6 graphically displays the scaling of FLIR scene complexity with four significantly different levels of complexity identified at the 99% level of confidence by the solid line enclosure of scene numbers. The addition of scenes differing significantly at the 95% level of confidence, identified in Figure 6 by enclosure within dashed lines, did not produce an increased number of FLIR complexity levels, as it did with earlier L3TV data. As can be seen in Figure 6, the addition of scenes which were significantly different at $p < .05$ merely provided a few alternative scenes at the four levels of FLIR scene complexity already established.[37,38]

Figure 7, based on an N of 9 including S-7's first FLIR scaling judgments, similarly displays graphically a scaling of scene complexity of FLIR displays. The order of the scenes scaled in Figure 7 is much

---

[37] Larsen, op. cit., p. 156.

[38] McNemar, op. cit., pp. 216-225, 352.



Key: for Boxes
—— = Sig. at $p < .01$
--- = Sig. at $p < .05$

Figure 6. Complexity Scaling of FLIR Scenes Based on N = 8 With Significantly Different Levels of Complexity.

like that of Figure 6 with the exception of some inversions. Noticeable changes affected by the addition of S-7's data include movement of Scene #5 (The Base Hospital) toward less complexity and Scene #6 (Microwave and Water Towers) toward increased complexity, as well as reducing the overall range of mean scale values from 5.1 to 4.7. Again, four levels of complexity are identified based on t-tests of the differences between the means significantly different at $p < .01$. As with the smaller group (N=8), the addition of scenes significantly different at $p < .05$ does not increase the number of scene complexity levels for FLIR, but does provide a few more alternative scenes at an established level.

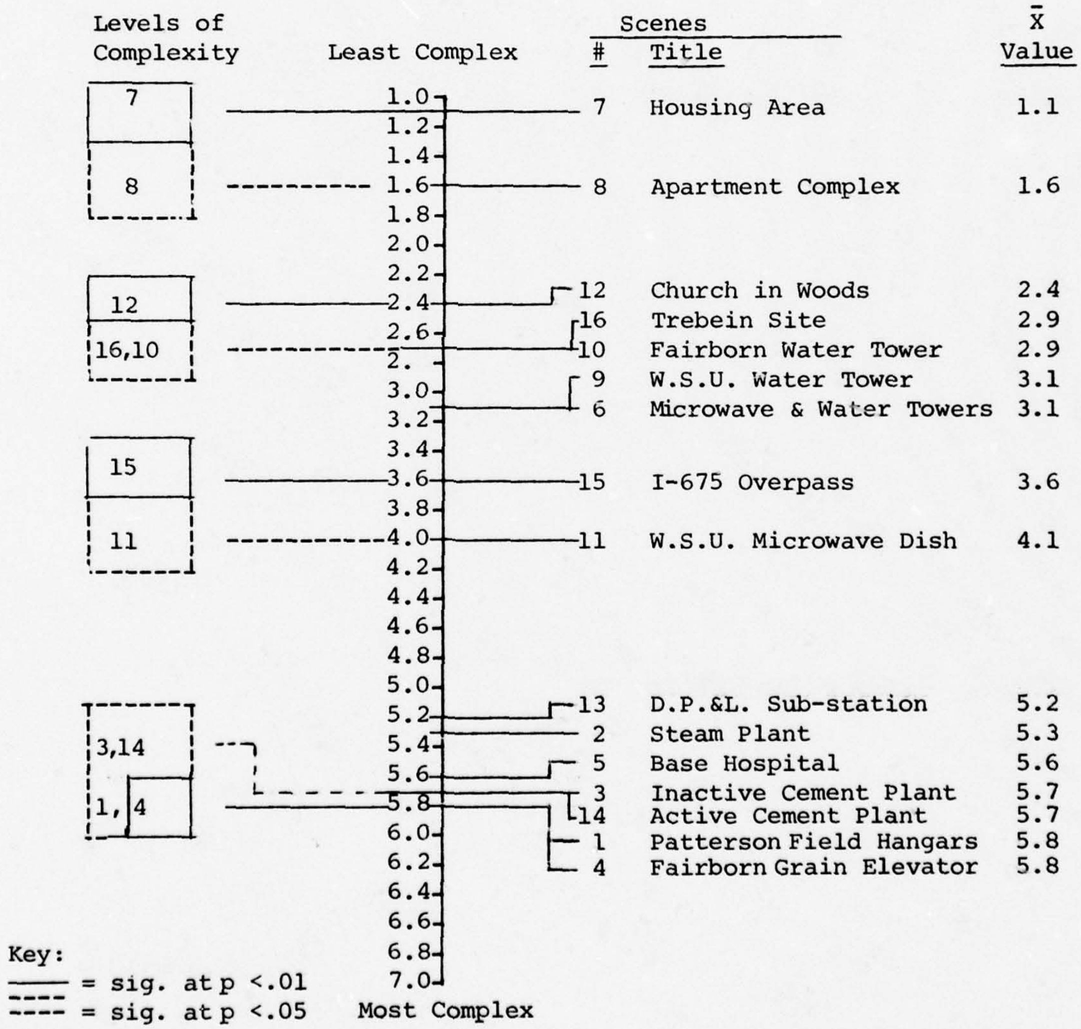


Figure 7. Complexity Scaling of FLIR Scenes Based on N = 9 With Significantly Different Levels of Complexity

Figure 8 illustrates by daylight photographs some of the differences in scene content between FLIR displays found to be significantly different at a 95% level of confidence or better among both the N = 8 and N = 9 groups. Again, the reader must be cautioned that FLIR displays at Pre-Sunset are <u>not</u> the visual equivaltent of these photos.

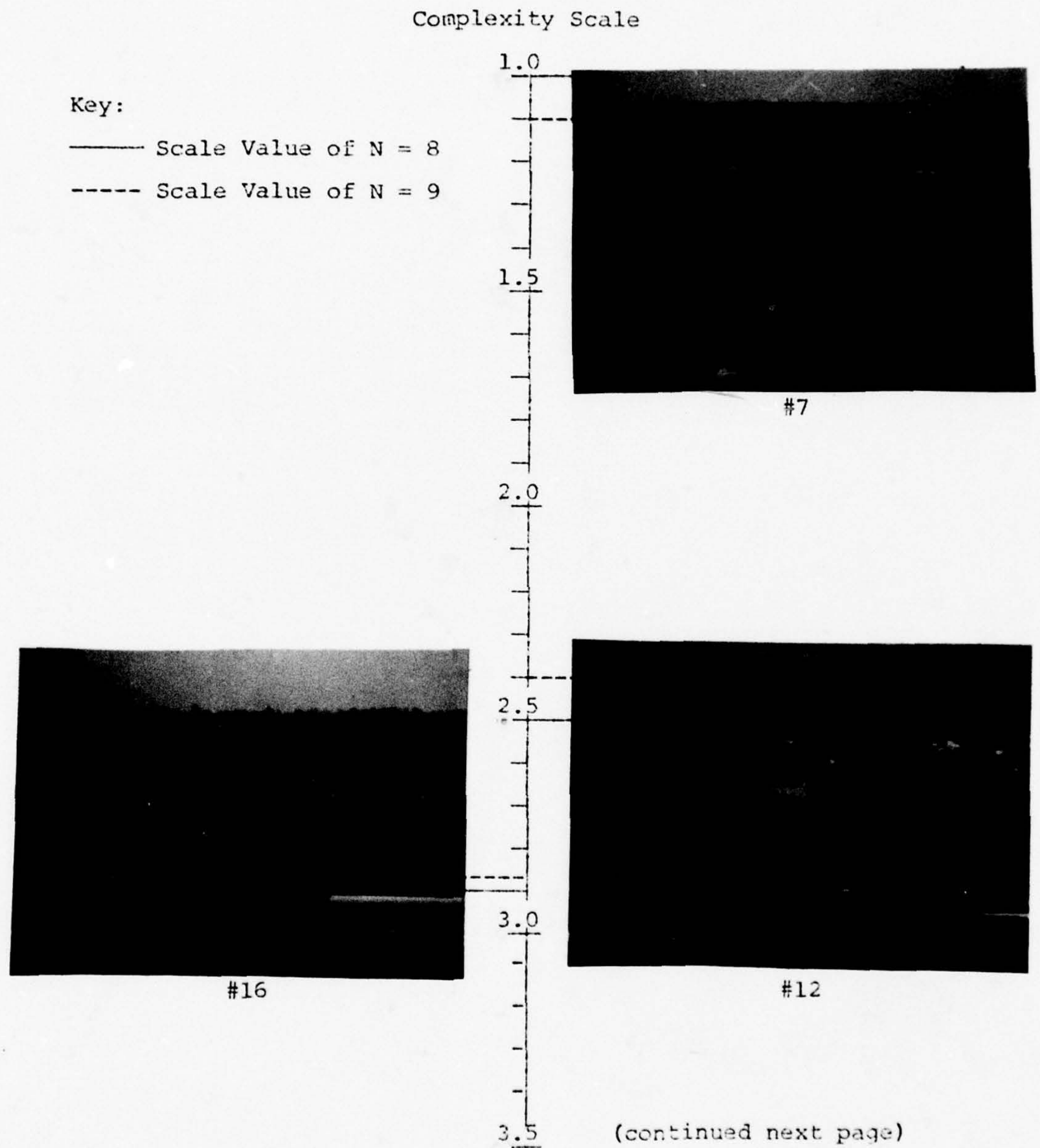


Figure 8. FLIR Target Areas at Four Significantly Different Levels of Scene Complexity.

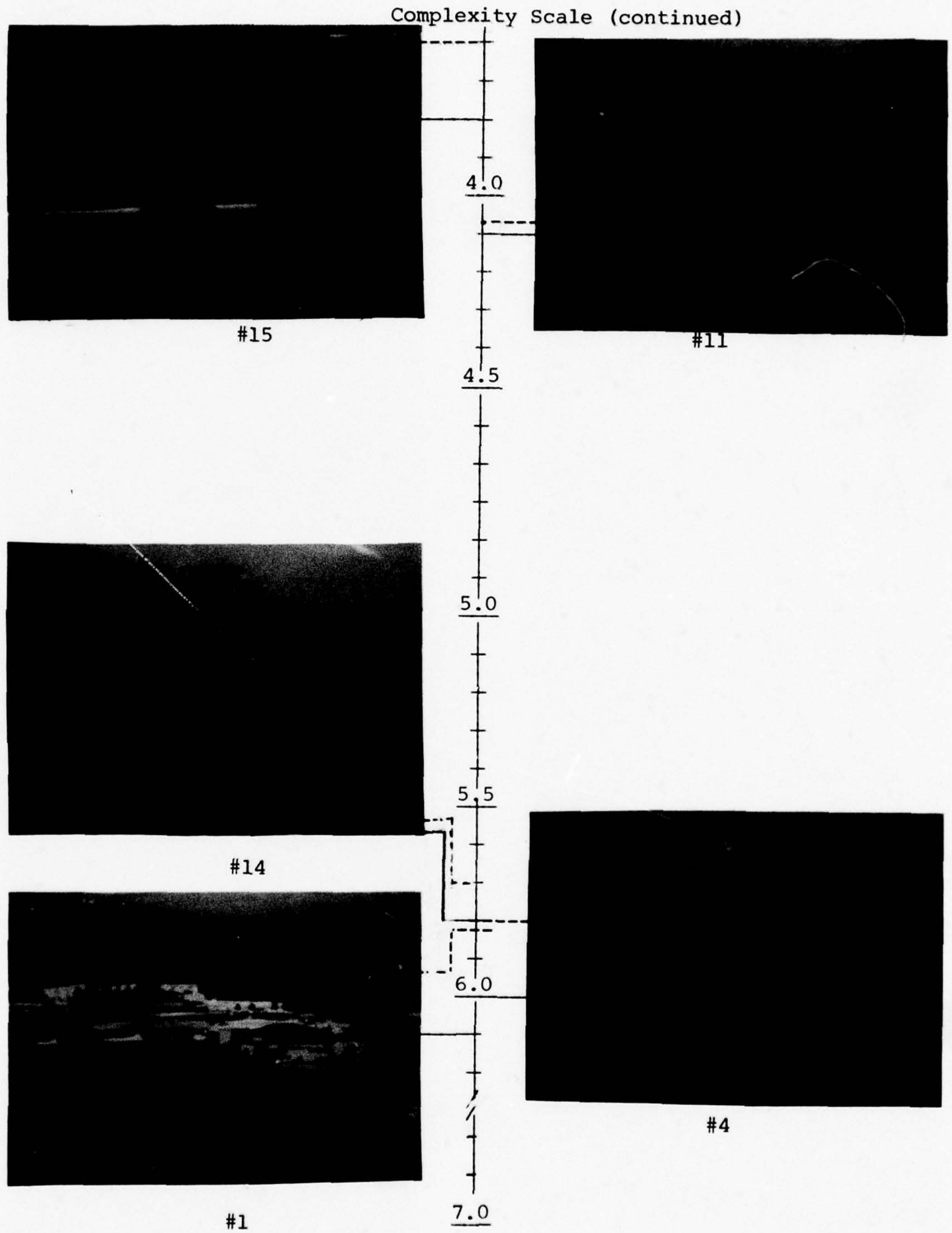


Figure 8. (Continued) FLIR Target Areas at Four Levels of Significantly Different Levels of Scene Complexity.

Comments Volunteered During FLIR Scaling

In addition to S-7's twice scaling FLIR photographs based on the assumption of two, different, specific flight missions, the second distinguishing feature of FLIR scaling sessions was initial subject response to the task. While scaling L3TV photographs, only two subjects made comments which were of a desultory nature, as reported in Appendix A. Early in the scaling process of FLIR photographs, however, six of the nine volunteered comments pertinant to the nature of their tasks. Since the volunteered comments were included in the content analysis which follows, the complete commentary is presented in Appendix B. The excerpts cited below demonstrate that the subjects appeared somewhat surprised, despite their general knowledge and familiarity with FLIR displays and their earlier scaling experience:

"...Things are quite different. This is considerably harder..."

"Holy Cow! This is harder...(it is) a different ball game..."

"This seems much harder...it...is not the same job...(I have) trouble differentiating..."

"In some (scenes)...the brightness is too high! ...detail masked by brightness...the field of view has changed...is different...more information available...quality better than L3TV..."

"More...difficult...than earlier...(some scenes) are overexposed, too white..."

"...(This) isn't the same task...looking (at it)...in a different light..."

On the basis of the above alone one can see that many, if not most, subjects considered the scaling of FLIR photographs for scene complexity not only to be harder than L3TV scaling but also considered it a different kind of task, in fact. More than one subject complained about the brightness level.

Content Analysis of FLIR Comments and Debriefings.

Following the same procedures used earlier, the subjects' debriefing responses made after scaling photographs of FLIR displays for scene complexity are presented in Appendix B with primary and secondary scaling criteria identified and underlined as previously. These criteria are enumerated and combined into similar Descriptor items in Table 6. As in Table 2, which resulted from a content analysis of debriefing responses following L3TV scaling, Table 6 identifies the subjects using a descriptor contained within an item by the subject's number while the last two columns indicate the number (N) and percent of subjects contributing to each item. Included in the content analysis, identified as "[Recon:...]" are the debriefing comments of S-7 following his first scaling in which he assumed a reconnaissance strike mission.

Although two subjects reported that they used the same criteria in scaling FLIR scene complexity as with L3TV, one of these expressed dissatisfaction with his scaling. Several subjects appeared less certain than with the earlier scaling assignment about what criteria they had actually used. Two subjects reported that their criteria seemed to change as they moved up the scale toward increasing scene complexity while S-7 again reported using conflicting criteria. From an overall perspective, each subject's general concept of increasing FLIR scene complexity can be briefly paraphrased:

S-1: Rather hard to say. My criteria seem to have changed moving up the scale. (From few) objects or white areas on a dark ground (to more) which are clearer and more separate.

S-2: (I used) the same criteria as before: the amount of man-made versus the surround in both area and number of objects.

S-3: How many objects...I could recognize as objects and how clear or how saturated objects were.

S-4: I guess the density per unit of man-made target is increasing as I go up the scale...coupled with...what I could distinguish.

S-5: I think the area, the amount of the photo that has terrain information or something in the foreground that was not washed out white was a criteria. (From) few cultural features to many well defined cultural details.

S-6: I just looked at the amount of cultural information, the amount of man-made objects. I seemed to have used the same criteria as before.

Table 6. Content Analysis of FLIR Scene Complexity Debriefing

Item #	DESCRIPTORS OR KEY WORDS	CITATIONS AS CRITERIA Primary Subject #	Secondary Subject #	TOTAL N	Per-cent
1	Number of recognizable objects, targets, white or heat-producing areas. [Recon: number of target like things]	1,2,3, 5,6,7, 8,9		8	89%
2	Standout from ground; more distinct, clearer, more saturated objects; clarity, contrast; more distinguishable as objects; less blended-in; smeared, blur. [Recon: threat, false alarm]	1,3,5	2,4,7 9	7	78%
3	White; dociles & density of white; blobs; bright objects; lighter or whiter man-made objects, heat emitting objects. [Recon: targets, defenses]	7,4	1,5,6, 8,9	7	78%
4	Highest percent area or amount of man-made objects, targets, white areas to surrounds; amount of sky & washed-out foreground buildings; density man-made/unit area, distribution bright objects.	2,4,6, 8,9	5	6	67%
5	Washed-out information, obliterated by brightness; masked detail; too high brightness; too white; poor quality, over-exposed; long-differentiating. [Recon: time consuming]	3,4,5, 7,8,9		6	67%
6	(Some) well defined cultural details; more detail; more different, distinguishable types of things; more specific objects & shape of buildings; significant object size & complexity. [Recon: Unmasked, differentiated targets]	5,9	3,4,7 8	6	67%
7	Homogeniety of background (darker), trees, or scene (lighter); surrounds (foilage), blank foreground; uniform distribution of light. [Recon: masking = foliage, clutter, & unimportant buildings.]	7	1,2,5, 8	5	56%
8	Recognizable patterns; reverse recognition task; curiosity vs necessity. [Recon: target=unique shape or man-made geometry & patterns, like LOCs]	4,7	2	3	33%

S-7: There were three conflicting criteria: A. When I couldn't find a target because of masking; B. (Detection) of a target and also many other target-like things; and C. Targets (which) have unique shapes or patterns.

S-8: (For some) I ignored the lower portion detail (washed-out nearby buildings); the most complex have the highest percentage of man-made objects in the picture or area of the picture.

S-9: I ranked them on what targets of man-made objects I could get out of the picture on the basis of the amount of the picture (which) showed heat.

Descriptors Related to FLIR Scene Complexity

The discussion which follows of FLIR Descriptors based on Table 6 parallels the earlier discussion of the results of the L3TV content analysis. As previously, items presented in Table 6 reflect the comments of a majority of subjects with one exception (Item 8) for later comparison purposes.

The background of FLIR photographs was not treated as unimportant, perhaps because the subjects appeared to have difficulty distinguishing where the background left off and objects began. Several subjects mentioned masking features and the smearing, blending, or blurring of objects into the background while some made a deliberate choice to treat whited-out portions of buildings in the immediate foreground of a few scenes as background. At least one subject (S-7) similarly ignored large or prominent objects because they would not have been defined as targets in his assumed missions. This time, also, one subject mentioned the sky and specifically referred to the area taken up by the sky in some displays.

Because many subjects reflected increased uncertainty about the criteria they had used in scaling, the unidimensionality of a scene-complexity continuum appears more questionable with FLIR scenes. While one subject (S-4) wondered at the appropriatness of "necessity vs curiosity", another (S-2) reported that if the task had been target recognition, he'd flip the complexity scale over. S-7 pointed out that his scale could be "folded over", implying a circular scale while a few subjects seem to imply the same about parts within the scale.

A task orientation was most fully manifested by S-7 for whom FLIR scene complexity could not be disassociated from the assumption of a specific flight mission. This association appeared to influence other subjects because the use of mission-oriented and task-performance words appeared with greater frequency than they had in the L3TV debriefings, for example:

"Target" was used this time by four subjects (S's 2,4,7,&9), twice as many as earlier.

"Recognition" previously mentioned by only one subject was referred to by five subjects, two of whom (S's 1 & 3) discussed the difficulty in recognizing objects due to an inability to see exact shapes while another (S-4) mentioned looking for recognizable patterns.

"Identification" was mentioned only by one of the two subjects who used it earlier, but this time with reference to the identification of "checkpoints" and "centers of man-made somethings."

"Detection", while used during L3TV debriefings by one subject, was not used by any subject after scaling FLIR photographs--but neither were there any references to "sub-objects", "sub-structures", or "sub-patterns." Instead, subjects discussed "distinct" things in contrast with the ground (S-1), "visible" detail (S-5), or "easily distinguished" objects (S-9), while a couple mentioned the length of time required to "differentiate" targets or objects (S's 3 & 7).

"Information" was used by five subjects, two of whom had been among the three subjects mentioning it with L3TV scenes. Two subjects (S's 3 & 4) appeared to use information as synonymous with "object detail" in discussing information which was not present, e.g. washed-out or obliterated. Another (S-5) referred to terrain information while S's 2 and 6 referred to more global concepts as "displayed information" or "cultural information."

The "luminous output pattern"[39] or scenic pattern of contrast levels (pointed out in earlier discussion as an unmet expectation of L3TV responses), while not cited as such directly, appeared to give many subjects some difficulty. All subjects were intellectually prepared to see heat patterns while scaling photos of FLIR displays, yet at least initially several subjects failed to recognize as

---

[39] Snyder, op. cit., pp. 89-96

information-carrying cues the whiteness or brightness of some display areas (e.g. relatively hot) or the fuzzy edges of some objects (e.g. heat emissivity). While a couple of subjects reminded themselves of what light areas meant, more frequently subjects mentioned the washing-out, smearing, or obliteration of figural detail by brightness or over-exposure. In short, infrared information was often treated as noise. Despite an intellectual understanding of the meaning and use of FLIR displays plus prior experience, both of which can reasonably be expected to be greater than a typical trainee would have, the predominant "gut-level" response of the subject group was as if they had expected a display of a normal daylight visual scene where brightness meant light and not relative heat.

Comparison of FLIR and L3TV Content Analyses

Contrasting the FLIR content analysis of Table 6 with that of L3TV's Table 2, certain similarities can be noted in that three items appear to be directly comparable. In both analyses, the item mentioned by the greatest number of subjects, Item 1, refers to an increasing number of objects in a scene as complexity increases. In the FLIR analysis 89% of the subjects made such a reference while 100% did so in the L3TV analysis. Similarly, both Items 4 refer to the percent or proportion of object area in a scene relative to the area of the surrounds as contributing toward scene complexity. In both analyses this was mentioned by 67% of the subjects. Item 7 in the FLIR content analysis contains descriptors referring to the nature or appearance of the background or surrounds, similar to Item 2 in the earlier L3TV analysis. While a greater number of subjects mentioned the background in the L3TV study (78%), the 56% who mentioned it in the FLIR study appeared to attribute a greater importance to its nature.

Several items in the FLIR content analysis refer to aspects of scene content, yet many of these have no direct equivalent in the earlier analysis. This is illustrated in Table 7 which presents a summarized topic or title for each item in both analyses as well as the percentage of subjects mentioning descriptors contained within the item. With the exception of the three items discussed above which are connected by solid line arrows, Table 7 reveals that of the eight items adduced from the FLIR content analysis, the remaining five have no clear-cut or single equivalencies with those of the L3TV analysis. Connected by dashed line arrows, Item 2 of the FLIR analysis appears to be partially covered by Items 6 and 7 of the L3TV analysis and conversely, Item 3 of the L3TV analysis appears to be partly covered by Items 6 and 8 of the FLIR analysis. Thus, some portions of items from one analysis appear to be matched by more than one item in the other, but a careful review of the descriptors involved indicates that the matching is quite partial.

This lack of correspondence between items from the two analyses indicates that the visual parameters and synthesized measures adequate for quantification of L3TV scenes will probably not be adequate when applied to FLIR scenes other than those based upon L3TV Items 1,2, and 4, namely:

-Parameter I with only the derived measure 1) since sub-structural objects were not mentioned in the FLIR content analysis (presented on page 29).

-Parameters II and/or IIa since sky was mentioned (presented on page 30). These parameters, however, may not be adequate as illustrated in the discussion which follows.

General Comments About FLIR Displays

Specification of the means of deriving a FLIR/CIG Property List comparable to the L3TV/CIG Property List defined earlier is not warranted based on the present research. This is supported by the fact that comments made in response to FLIR photographs reveal several issues which either are in conflict with or did not emerge in the earlier L3TV analysis.

Table 7. Comparison of FLIR and L3TV Content Analyses

FLIR Analysis			L3TV Analysis		
Item #	Topic or Subject	Responses	Item #	Topic or Subject	Responses
1	Number of man-made objects or	89%	1	Number of man-made objects	100%
2	Contour/edge gradient descriptors	78%	2	Background	78%
3	White = hot = man-made	78%	3	Object detail: pattern & shape (outline) complexity	67%
4	Proportion of object area to surrounds	67%	4	Proportion of object area to surrounds	67%
5	Detail: masked or washed-out	67%	5	Object detail: interior complexity	67%
6	Cultural detail, object shape	67%	6	Scene or image quality (resolution)	56%
7	Nature of Background	56%	7	Lines and Edges	56%
8	Patterns	33%			

For example:

1. Not all man-made buildings were classified as objects or figures of interest in the FLIR analysis. At least one subject tended to ignore some large, central objects on the ground that they could not possibly be targets while several subjects appeared to classify as background large areas of the top portions of near-by buildings in the immediate foreground, a distinction they did not appear to make with L3TV.

2. Some subjects saw small, amorphous bright patches as ground or clutter, others saw them as undifferentiable objects, while at least one subject interpreted such light patches as possible "threats" or evidence of "defenses."

3. Although object shape was an important cue in the L3TV analysis appearing as a primary criterium for 67% of the subjects, only one subject used the word "shape" in distinguishing "targets" from "non-targets", and only three subjects even mentioned the word "pattern", principally in reference to high power lines, roads, poles, or trees with the FLIR displays.

4. In the FLIR analysis there appeared to be a greater divergence among subjects as to what constituted objects or "targets" as well as a tendency toward a greater task or mission orientation than in the earlier L3TV analysis. It is here suggested that graphics, which may be important with L3TV, are less important in FLIR displays where target recognition can reasonably be expected to depend more heavily on a particular pattern of light or "signature" cues.

5. Items 3 and 5 in Table 6 of the FLIR content analysis reveal the lack of figural detail which subjects had attributed to complex figures in L3TV displays. As such, these two items: "White = Hot = Man-Made," and "Detail: Masked and Washed-Out," imply that the subjects may not have been viewing the FLIR scenes for maximal visual efficiency and support the above suggestion.

Subject Responses to Sensor Displays

The preceding discussion and analyses indicate that to some extent the content of FLIR displays and subjects' responses to that content differred from those of simultaneously recorded L3TV displays and/or normal visual scenes. At some "time envelopes" during daylight hours under certain weather and atmospheric conditions (such as a Spring day, clear skies, low relative humidity, etc.) L3TV and FLIR sensors can present displays which are nearly equivalent, visually. Further, both displays can look much like the normal, "real-world" visual scene and at times present even an "enhanced" view of that in which environmental features are seen with greater clarity and distinction.

However, again depending upon weather and atmospheric conditions, at other time-envelopes FLIR and L3TV sensors present displays with predictably different characteristics which are quite unlike daylight visual scenes. For example, FLIR sensors, responsive to relative temperature changes which occur around sunset as the heating properties of the sun are withdrawn, regularly displayed a period of transition during which environmental features which have been reflecting heat from the sun cool down and the natural pattern of heat emissivity of objects in the environment is registered. Depending upon climate factors, this transition period can be but is not necessarily brief, nor is it of constant duration. Additionally, some environmental features which generate no heat of their own cool quickly while others dissipate heat stored during the day quite slowly, sometimes over a period of hours. On the other hand, the standard L3TV sensors utilized in the present research can be, at times, notoriously over-sensitive to some light sources in the night environment displaying the phenomenon of "blooming" in which small lights "bloom" or enlarge to flood the entire display if corrections are not made.

It is important to note that the displays used in the present research were recorded on a May 11th between 1915 and 1923 hours (7:15 to 7:23 P.M.) during the time envelope of Pre-Sunset. Of the two sensors, the L3TV displays more closely approximated an early dusk visual scene with several features typical of the time period present as long shadows and rounded, opaque foliage. No lights were blooming. The FLIR displays revealed the beginning of the transition period in that concrete objects were still warm while some foliage was beginning to show its natural heat emissivity pattern (e.g. lightish trunks and branches) and in some water towers the level of the water could be faintly distinguished within the tank. The differences between the L3TV and FLIR displays occasioned one subject to summarize during the FLIR debriefings: "One very valid thing I have learned from this is you have to treat FLIR as very different from L3TV."

The Matching Study

Because debriefing responses indicated the possibility that a limited correspondence between visual parameters related to FLIR and L3TV scene complexity might result from further analysis, but primarily because so many subjects reported that scaling FLIR scene complexity was a more difficult or harder task, this matching study was performed as a heuristic after-thought. Had any formal hypothesis been expressed at the time of data collection, it would have been that matching normal daylight photographs would be faster and more accurate with the photographs of the L3TV displays.

Subjects, in counterbalanced order, matched a set of 16 small (3 x 5½") daylight color photographs of views containing within them the FLIR and L3TV target-areas with both sets of photographs which they had previously scaled for complexity. Following instructions presented in Appendix C, subjects confronted a covered table on which the 16 FLIR or L3TV photographs were arranged in a 4 x 4 display and were handed a stack of small color photographs. The randomized arrangements of FLIR and L3TV photographs and the randomized sequences of color photographs for each subject are presented in Appendix C. The subjects' task was to match the sensor display photographs as quickly as possible by placing the small color photo containing the scene on top of it. At a given signal, the cover was removed and a stop watch started which was stopped when the subject indicated he had finished. While the subject stepped out of the room, the Experimenter set up the second arrangement, covered it, and rearranged the order of the color photographs. The process was then repeated so that all subjects matched daylight color photos with both FLIR and L3TV display photographs. At the end of the two matching sessions, the subject was asked if he had any comments.

Subject performances were scored and times as presented in Table 8.

Table 8. Results of the Matching Study

	L3TV		FLIR	
Subjects	Time (sec.)	# of Errors	Time (sec.)	# of Errors
S-1	102	0	147	0
S-2	262	3	384	4
S-3	155	0	276	3
S-4	125	4	195	2
S-5	199	2	165	3
S-6	188	2	438	6
S-7	237	0	227	0
S-8	72	0	134	0
S-9	100	0	78	0
$\bar{X}$	160.00	1.22	227.11	2.00
s	65.76	1.56	119.21	2.19
$s_{\bar{X}}$	23.25	.55	35.07	.77

As illustrated in Table 8, matches of color photographs with the L3TV photographs occasioned more errors and required longer times, on the average, than those with the FLIR photographs. However, t-tests of the differences between the means of both measures failed to reach a significant level, principally because of high with-in group variance. None-the-less, the direction of the differences was opposite to that which had been expected. Despite the fact that FLIR scene-complexity scaling had been cited as more difficult by many subjects, the matching study revealed a greater average percent correct responses per subject (92.4% versus 85%) with shorter average matching time per picture (10 seconds versus 14.2 seconds) for the FLIR displays than for the L3TV displays.

Out of a total of 144 matches with each set, there were 18 total errors or 12.5% with the L3TV displays and 11 total errors or 7.6% with the FLIR displays. Table C-3 in Appendix C presents the matching errors. No matching errors occurred with either set for Scenes numbered 1,2,3,4, 5,9,10,11 and 14. In addition, Scene #12 in the set of photographs of FLIR displays had no matching errors.

With the L3TV displays all but one of the 18 matching errors occurred among the first six scenes scaled at the lower end of the scene-complexity continuum. A similar tendency, but to a considerably lesser degree, appeared in the matching errors associated with the FLIR displays. Thus, the occurrance of an error in matching daylight color photographs with photographs of L3TV and FLIR displays appears to be related to a lack of or lessened scene complexity, particularly for L3TV displays.

Correlations Between Scalings

Evidence presented thus far indicates that there is some correspondence between the visual parameters of FLIR and L3TV displays. Evidence also indicates that the correspondence is not total or complete, albeit the subject group was small, the data were derived from essentially one stimulus set of scene contents, and displays of a prototype FLIR sensor were compared to those of a standard L3TV sensor.

One method to evaluate the degree of that correspondence between visual parameters is to correlate for each subject his scaling judgments of the two sensor displays for each scene or target-area. The Pearson product moment correlation coefficients under the column headed $\underline{r}$ are presented in Table 9 with an estimate of the proportion of variance in one scene complexity scale which is due to the variance in the other.[40] This variance overlap or common variance between the two scene-complexity scalings was found by squaring each subject's correlation coefficient and is expressed as a percentage in the column $\underline{r^2}$.

Inspection of Table 9 reveals that all correlations are positive and fall between a low value of +.31 to a very high value of +.92. The average or overall correlation is r = +.72, significantly different from zero with greater than a 99% level of confidence.[41] Although the variance of one scale accounted for by the variance of the other varied from 10% to 85% among subjects, the average is about

---

[40] McNemar, Q., op. cit., p. 116.

[41] Edwards, A. Statistical Analysis. New York: Rinehard and Co., 1949, p. 331.

56%. Thus, a moderately good overall linear relationship is found to exist between the two scales with the error of estimate from one to the other reduced by about 30%.

Table 9. Correlations and Common Variances Between Complexity Scales

Subjects	r	r^2
S-1	+.75	56%
S-2	+.72	52%
S-3	+.48	23%
S-4	+.92	85%
S-5	+.70	49%
S-6	+.82	67%
S-7 (1st FLIR Scaling)	+.31	10%
S-8	+.89	79%
S-9	+.89	79%

The above indicates that there is a very good probability that scenes from one sensor display scaled as high or low in scene-complexity will be similarly scaled when from the other sensor display. These statistics attach an added importance to the three common Items found in the two content analyses performed earlier and to the first two parameters and the physical measures derived from those items. In brief paraphrasing recapitulation, these were:

I. The number of man-made objects or object-areas in a scene.

1) The number of separate, single entities ignoring sub-structural aspects, attachments, or sub-features.

2) (not applicable to FLIR)

II. The ratio (or percent) of the area of man-made objects to the combined or separate areas of foliage/terrain and sky area.

3) The area of man-made objects.
4) The area of terrain/foliage.
5) The area of sky.

However, it is interesting to note that the subject who displayed the greatest mission orientation during FLIR scaling, S-7, also displayed the lowest proportion of common variance, 10%. With his lowered correlation of +.31 in Table 9 the indication is strong that task assignment has the ability to modify visual parameters and perceptual cues. This is brought more clearly into focus when the two complexity scalings S-7 performed upon photographs of the FLIR displays are correlated. As the reader will recall, this subject first scaled FLIR displays for scene complexity assuming the tactical mission of air reconnaissance utility and then rescaled the same displays assuming a strategic mission, that of low level ingress. The correlation coefficient between his two sets of scale values for each scene was $r = -.13$, or essentially zero. When the scale values from this subject's second FLIR scaling (the strategic mission of low level ingress) are correlated with his original scene complexity scale values of the L3TV displays, a correlation coefficient of $r = -.68$ results which is significantly different from zero in a negative direction at $p < .01$.

Supported by some comments of other subjects which, at the time, appeared to be equivocating, this subject serves to remind us that how pilots actually use sensor displays, most particularly FLIR displays, when under specific flight assignments, must be taken into account when designing a simulator for Sensor Operator training which incorporates computer generated imagery if positive transfer-of-training is to be effected.

IV. SUMMARY AND CONCLUSIONS

A human factors feasibility study was performed on the perceived scene-complexity of E-O sensor displays with particular reference to programable parameters in computer generated imagery. The reported research is a heuristic investigation in which human responses to still L3TV and FLIR scenes were assessed so that the effects of varying degrees of realism in CIG displays, held to be a function of scene complexity in terms of real-world edges and surfaces, could later be evaluated for cost effective trade-off decisions in the development of a Sensor Simulator.

Research objectives were: 1) To scale subjectively the complexity of still L3TV and FLIR scenes; 2) To identify scenes of significantly different levels of complexity for comparison and later CIG simulation; 3) To determine empirically the major perceptual factors associated with L3TV and FLIR sensor displays; and 4) To relate perceptual with physical factors amenable to CIG simulation.

Along an equal interval continuum of scene complexity, subjects psychometrically scaled first a set of photographs of L3TV displays subtending a narrow field of view and later scaled a similar set of photographs of FLIR displays of the same 16 target-areas at Pre-Sunset which had previously been simultaneously recorded onto video tape. After the scene-complexity scaling of each set of sensor-displays, subjects were debriefed as to the criteria used in scaling. Debriefing comments were submitted to a content analysis in which primary and secondary scaling criteria were identified and combined into descriptors from which visual parameters and parallel physical measures were synthesized. In an add-on study, subjects matched in counterbalanced order the randomized photographs of L3TV and FLIR displays with small daylight color photographs containing the same target-areas. Time and matching errors were recorded. With one subject omitted for failure to follow directions, the remaining subject group consisted of nine individuals with previous experience with L3TV and FLIR displays. Although the subject group was too small and the stimulus sample too restrictive for any conclusions to be "cast in concrete", several inferences may be drawn from the results.

Analysis of L3TV scene-complexity scaling data identified scenes for later computer modeling which were located at three levels of complexity scale values at $p \leq .05$. Computer modeling would permit trade-off comparisons of the influences of various CIG parameters such as the number of edges and surfaces, edge smoothing, curvature, contouring, noise, number of grey levels, and weather algorithms among others for cost effective Sensor Simulator design decisions to optimize positive transfer-of-training.

A content analysis of L3TV debriefing responses revealed: 1) apparent acceptance of the assumed unidimensionality of the scene-complexity scale; 2) while subjects focused on real-world objects or features the background appeared relatively unimportant, although mentioned; 3) relatively infrequent use of task or mission related words; and produced seven descriptors common to a majority of subjects. The L3TV descriptors, albeit limited to a single stimulus set of 16 scene contents, dealt primarily with geometries of the visual array can be briefly summarized with the percentage of subjects included as:

1.	Number of objects	100%
2.	Background	78%
3.	Object detail: Outline shape	67%
4.	Proportion of object area to ground area	67%
5.	Object detail: Sub-structure, interior complexity	67%
6.	Resolution	56%
7.	Lines and edges	56%

Four parameters and seven physical measures were synthesized by which each photograph of an L3TV display could be quantified. Although the original stimulus materials could not be released for a validating study of the derived measures, it was hypothesized that the best measures would show a monotonic relationship with scene-complexity scale values and lead toward the development of an L3TV/CIG "Property List". The parameters and associated physical measures recommended for L3TV displays were:

I. The number of man-made objects or object areas in a scene.

1) the number of separate, single entities, ignoring sub-structural components of such entities.

2) the number of separate, single entities including sub-structural aspects provided the smaller features can be recognized as

completed units.

II. The ratio of the area of man-made objects to the combined (or separate) areas of foliage/terrain and sky.

3) the area of man-made objects (%).
4) the area of terrain/foliage (%).
5) the area of sky (%)

III. The number of straight and curved lines which make up visible portions of man-made objects.

6a) the number of external edges.
6b) the number of internal object lines.

IV. Scene quality as reflected by resolvable detail transmitted.

7a) the shorter dimension of the smallest object in 2) above or 7b) the length of the shortest line in 6) above.

A new dimension was introduced during the scaling of FLIR scene-complexity in that while in the process of scaling a majority of subjects made comments reflecting surprise at the difference from L3TV complexity scaling, increased task difficulty, and the brightness of some areas in some scenes. The situation was further compounded by one subject, apparently unable to scale FLIR scene-complexity without assuming a specific flight mission, who scaled the same FLIR displays twice first assuming a tactical mission and secondly a strategic mission. Omitting that subject's data, FLIR scenes were identified for later computer modeling at three levels of scene-complexity, based on non-overlapping response distributions, and at four complexity levels, based on t-tests of differences between the means, $p < .01$. The less stringent requirement of significance at $p < .05$ increased the number of FLIR scenes included at each level but did not increase the number of levels of scene-complexity as earlier in the L3TV analysis. Inclusion of the highly mission-oriented subject's first scaling data (tactical mission) produced two levels of FLIR scene-complexity based on non-overlapping response distribution and identified scenes at four complexity levels by t-tests at $p < .01$ and $p < .05$ levels of significance. Since the inclusion of this subject's second scaling data (strategic mission) produced no response distributions which did not overlap, those data were omitted from further analysis.

A content analysis of comments made while scaling and during debriefing revealed: 1) despite the subjects' intellectual mastery and prior experience with FLIR sensors and displays, generalized initial responses appeared to regress toward the expectancy of a normal visual scene; 2) scaling FLIR was reported to be more difficult and a different kind of task ("a new ball-game") by a majority of subjects; 3) less certainty was displayed about scaling criteria with indications that a FLIR scene-complexity continuum might not be unidimensional; 4) apperception of the background appeared more difficult and some subjects tended to redefine targets (or objects) apparently because of increased blending, blurring, smearing, and whiting-out of portions of scenes; and 5) subjects displayed increased mission-orientation using more task words.

The FLIR content analysis produced seven descriptors common to a majority of subjects, only three of which had a direct equivalency with those produced earlier from the L3TV content analysis. With the three equivalent descriptors underlined and the percent of subjects citing each, these are briefly summarized as:

1. Number of objects	89%
2. Edge gradients	78%
3. White=hot=man-made	78%
4. Proportion of object area to ground area	67%
5. Missing detail: masked and washed out	67%
6. Distinguishable cultural detail	67%
7. Background	57%
8. Patterns	33%

A moderate linear relationship between L3TV and FLIR scene-complexity scale values was indicated by an average correlation of +.72 with the variance in one scale accounting for about 56% of the variance in the other, in general. Results of the matching study indicated that 1) for both sensor displays the most mis-matches were made with scenes judged as less complex and 2) although scaling FLIR scene-complexity might be harder for many subjects, matching FLIR displays with daylight color photographs was faster and accomplished with fewer errors. This implies the perceptual cues and unique signatures of targets as displayed by FLIR sensors to be at least at a comparable level with the graphic detail displayed by L3TV sensors.

The three descriptors evolved in the FLIR content analysis, identified above, which were equivalent to L3TV descriptors undoubtedly account for a great deal of similarity between the L3TV and FLIR scene-complexity scales and support quantification of FLIR displays using the physical measures 1), 3), 4) and 5) derived from Parameters I and II. However, further analysis with extension toward devlopment of a FLIR/CIG "Property List" is not presently recommended because additional evidence indicated that visual parameters associated with FLIR displays can be modified by cognitive factors and different mission or task assignments, particularly during the transition period involving thermal inversion during Pre-Sunset.

In general, the findings of this research appear to support a CIG simulation of an L3TV sensor display which stresses the optical array of surfaces, edges, and lines; but simulation of FLIR sensor displays is left somewhat open to question about the influence of task variables upon human perception as well as the influence of atmospheric variables including the time-of-day and sun angle upon sensor performance and, hence, human performance and training. For example, no finding denied the possibility that the use of FLIR sensors in conjunction with other E-O sensors may render as inadequate models of target acquisition which assume a hierarchiacal ordering or sequencing of the visual processes of target detection, recognition, and identification.

Recommendations

Since the experimental methods and analytical techniques herein reported have manifested the feasibility of deriving CIG sensor simulation information from subjectively scaled L3TV and FLIR displays in terms of scene-complexity, it is natural that the first recommendations should extend and refine these methods and techniques.

1. A fundamental goal of this research was the identification of L3TV and FLIR displayed scenes at various levels of scene-complexity

so that programmers could model such scenes with CIG techniques. With scene-complexity levels now identified for both L3TV and FLIR sensor displays, a minimal next step should be performed. Target-areas found to be significantly different in complexity for both sensors (such as Scenes 7. Housing Area With (Distant) Water Towers; 15. I-675 Overpass; and 1. The Patterson Field Hangars, for example) should be modeled. Presently, perception of CIG simulated imagery in the Sensor Simulator under development is defined by size, shape, gray shade (16), contrast, noise, and curved surface shading. Additional features such as texture, level of detail, complexity, fewer gray shades, foliage masking, the weather algorithm (simple contrast reduction), the kind of noise, and perhaps a smearing algorithm to simulate air speed are CIG parameters which are amenable to trade-off consideration. These should now be systematically and progressively varied. Comparison of the resulting programmed scenes simulated for each sensor, particularly if made dynamic, would provide the basis for adequate cost-effective decisions regarding the hardware and programming needs of the Sensor Simulator.

2. The recommended procedures of quantifying L3TV displays in terms of the physical measures derived in this study (pp. 30-33) with a subsequent analysis toward the establishment of a L3TV/CIG "Properties List" (pp. 33-34) should be tested out on the original stimulus materials. Quantification of the FLIR stimulus materials using the recommended measures (p. 48) should also be initiated.

3. Because of built-limitations, replication of this study is strongly recommended using a larger and different group of subjects which are, perhaps, more representative of the present user group of sensors and/or of potential trainees. A wider variety of still photographs of L3TV and FLIR displays during different time-envelopes and other weather and atmospheric conditions is also recommended.

4. In line with the above, to improve CIG sensor simulation of L3TV and FLIR sensor displays, techniques need to be developed to

improve the general understanding of the visual relationships between and within L3TV and FLIR sensor parameters as observed on sensor displays. For example, L3TV and FLIR sensors are considered influenced by some 18 atmospheric variables including sun angle (time-of-day). In terms of the displayed information for a variety of flight missions, L3TV and FLIR parameters need to be catalogued, identified, verified, and measured (if possible) which:

a. have essentially the same or very similar visual effects upon sensor displays and, hence, could reduce CIG or training complexity without sacrificing confidence in ultimate positive transfer of training.

b. have important and differing effects upon visual sensor displays and, therefore, must be included in ground-based sensor simulation and training.

5. The influence of cognitive factors in conjunction with mission orientation in the present study, most particularly with reference to FLIR displays, leads to the recommendation that prior to full implementation of the Sensor Simulator, intermediate briefing or training materials be developed for Airmen which addresses their expectancies and perceptual sets. Such briefing materials should include descriptions and illustrations of FLIR sensor displays for a variety of flight missions under regular conditions involving transition periods or thermal inversions which occur as a function of sun angle, season, weather, etc., and less frequent but often occuring conditions which could disrupt operations, as well as typical FLIR displays. A separate topographical taxonomy may need to be developed for training of various flight missions (such as TA/TF, Long range ingress, Air to ground attack, Battlefield survelliance, Fly in landing, etc.) with subcategories of flight variables known to affect sensor displays such as weather and atmospheric influences; altitude and speed of aircraft, and so forth.

REFERENCES (FOOTNOTES)

1. "Electro Optical Navigation System Planned for B-52 G/H", Aviation Week and Space Technology, May 10, 1976, pp. 130-131.

2. "Human Resources Lab Accelerates Research", Aviation Week and Space Technology, July 19, 1976, p. 225.

3. Reynolds, P. "'76 AIAA Visual and Motion Simulation Conference." Astronautics & Aeronautics, July/August 1976, p. 68.

4. Stein, K.L. "AFSC Stresses Increased Simulator Use." Aviation Week and Space Technology, July 19, 1976, p. 101.

5. Rosell, F. and Willson, R. Performance Synthesis of Electro-Optical Sensors. Tech. Rep. AFAL-TR-74-104, WPAFB, Oh., April 1974, particularly pp. 50-56, 127-137, 141-142, 175-176, and 193-194.

6. Stein, op.cit., p. 105.

7. Reynolds, op.cit., p. 70.

8. Bunker, M. and Heeschen, R. Airborne Electro-Optical Sensor Simulation. Tech. Rep. AFHRL-TR-75-35, AFSC, Brooks AFB, Tx., p. 89.

9. Foley, W. and Orlando, R. Personal Communication. July, 1976.

10. Stein, op.cit., pp. 106-107.

11. Gibson, J.J. Optical Motions and Transformations As Stimuli For Visual Perception. Psychological Review, 1957, 64, 288-295.

12. Gibson, J.J. Perception of Distance and Space in The Open Air, in Beardslee and Wertheimer (Eds.) Readings in Perception. Princeton: Van Nostrand, 1958, pp. 415-431.

13. Nunnally, J.C. Psychometric Theory. New York: McGraw-Hill Co., 1967, pp. 31-56.

14. Edwards, A. Techniques of Attitude Scale Construction. New York: Appleton Century Crofts, Inc., 1957, pp. 5-9, 172-198.

15. Biberman, L.M. (Ed.) Perception of Displayed Information. New York: Plenum Press, 1973, p. 43.

16. Stevens, S.S. Ratio Scales of Opinion, in Whitla, D. (Ed.) Handbook of Measurement and Assessment in Behavioral Sciences. Menlo Park, Ca.: Addison-Wesley Pub. Co., 1968, pp. 171-199.

17. Edwards, A. op.cit., pp. 121-148.

18. Cream, B. and Woodruff, K. Functional Integrated Systems Trainer: Vol. I. Description and Evaluation. Tech. Rep. AFHRL-TR-75-6, WPAFB, Oh., December 1975.

19. Smode, A. and Hall, E. Translating Informational Requirements Into Training Device Fidelity Requirements. Proceedings of The Human Factors Society 19th Annual Meeting, Dallas, Tx., October 1975, pp. 33-36.

20. Edwards, op.cit., pp. 172-198.

21. Peters, C. and Van Voohris, W. Statistical Procedures and Their Mathematical Bases. New York: McGraw-Hill Inc., 1940, pp. 171-176.

22. McNemar, Q. Psychological Statistics. New York: John Wiley and Sons, 1949, pp. 216-225

23. Larsen, H. Rinehart Mathematical Tables, Formulas, and Curves. New York: Rinehart and Co., 1949, p. 156.

24. Snyder, H. Image Quality and Observer Performance, Ch. 3 in Biberman (Ed.) Perception of Displayed Information. New York: Plenum Press, 1973, pp. 87-118.

25. Reed, S.K. Psychological Processes in Pattern Recognition. New York: Academic Press, 1973, pp. 11-53, 223-226.

26. Graham, C.H. Visual Form Perception, in Graham (Ed.) Vision and Visual Perception. New York: Wiley & Sons Inc., 1965, pp. 548-567.

27. Braunstein, M. Depth Perception Through Motion. New York, Academic Press, 1976, pp. 13-30, 41-56, 154-186.

28. Posner, M. Coordination of Internal Codes, in Chase (Ed.) Visual Information Processing. New York: Academic Press, 1973, pp. 35-74.

29. Lindsay, P. and Norman, D. Human Information Processing. New York: Academic Press, 1972, pp. 1-10, 94-113.

30. Biberman, L.M. (Ed.) Perception of Displayed Information. New York: Plenum Press, 1973, pp. 13, 35.

31. Rosell, F. and Willson, R. Recent Psychophysical Experiments and the Display Signal to Noise Concept, in Biberman (Ed.), ibid., pp. 167-231.

32. Biberman, L., ibid., pp. 313-322.

33. McCormick, E. Human Factors in Engineering and Design. New York: McGraw-Hill Book Co., 1976, pp. 63-65.

34. Biberman, L., op.cit., pp. 21, 76-84.

35. Winer, B. Statistical Principles in Experimental Design. New York: McGraw-Hill Book Co., 1962, pp. 576-621.

36. Hays, W. Statistics for Psychologists. New York: Holt, Rinehart, and Winston, 1963, pp. 490-538.

37. Larsen, op.cit., p. 156.

38. McNemar, op.cit., pp. 216-225, 352.

39. Snyder, op.cit., pp. 89-96.

40. McNemar, Q., op.cit., p. 116.

41. Edwards, A. Statistical Analysis. New York: Rinehart and Co., 1949, p. 331.

BIBLIOGRAPHY (SUPPLIMENTARY READING)

Anderson, N. and Leonard, J. The Recognition, Naming and Reconstruction of Visual Figures as a Function of Contour Redundancy, J. Exper. Psychol., 1958, 56, 262-270.

Bartley, S.H. Principles of Perception. New York: Harper and Row, Pub., 1969, pp. 51-109, 214-220.

Bennett, D. "EVS Distraction", Combat Crew, August, 1975, p. 6-7.

Biberman, L. Fallacy and Fact of Sampled Imagery Displays. Human Factors, 1974, 16, 286-299.

Broadbent, D.E. Perception and Communication. New York: Pergamon Press, 1964.

Brumaghin, S. Transmittal of Study Report: B-52 Pilot EVS Familiarization Guidelines Element 4170 Systems Integration of the B-52 Retrofit Program Part II. Coordination Sheet to C.A. Miller, Boeing Wichita, May 31, 1972.

Caro, P. Simulator Training Reconsidered: Alternative Concepts of Transfer. Proceedings of The Human Factor Society 19th Annual Meeting, Dallas, Tx., Octover 1975, pp. 29-32.

Casperson, R. The Visual Discrimination of Geometric Forms, J. Exper. Psychol., 1950, 40, 668-681.

Chase, W.G. Visual Information Processing. New York: Academic Press, 1973.

Cloud, G. Lasers in Engineering Education, Engineering Education, May 1976, 837-840.

Cream, B. Training Requirements and Simulator Utilization, Paper presented to the Society for Applied Learning, Washington, D.C., July 22, 1976.

Cream, B., Eggemeier, F. and Klein, G. Behavioral Data in the Design of Aircrew Training Devices. Proceedings of the Human Factors Society 19th Annual Meeting, Dallas, Tx., October 1975.

DeGreene, K.B. Systems Psychology. New York: McGraw-Hill Book Co., 1970.

Fantz, R. The Origin of Form Perception, Scientific American, May 1961, reprint.

Fitts, P., Weinstein, M., Rappaport, M., Anderson, N. and Leonard, J. Stimulus Correlates of Visual Pattern Recognition: A Probability Approach, J. Exper. Psychol., 1956, 51, 1-11.

Forgus, R.H. Perception. New York: McGraw-Hill Inc., 1966.

Gibson, J.J. The Information Available in Pictures, Leonardo, 1971, 4, 27-35.

Greening, C.P. Mathematical Modeling of Air-to-Ground Target Acquisition, Human Factors, 1976, 18, 111-147.

Grebs, M. Scanning Patterns in Real Time FLIR Displays. Proceedings of the Human Factors Society 19th Annual Meeting, Dallas, Tx., October 1975.

Hand, L.M. Recent Research on Avionics System Concepts for Low-Level Night Helicopter Operations, Aircrew Performance in Army Aviation, Proceedings of a Conference At U.S. Army Aviation Center, Ft. Rucker, Al., November 1973, pp. 89-106.

Hansen, O., Humes, J. and Offenstein, R. Multi Sensor Operator Training Phase I. Analysis of Operator Tasks and Target Recognition Skills. Tech. Rep. AFAL-TR-70-4, WPAFB, Oh., 1970.

Harshbarger, J. and Gill, A. Development of Techniques for Evaluation of Visual Simulation Equipment. Tech. Rep. AMRL-TDR-64-49, WPAFB, Oh., 1969.

Hawkins, J. and Valverde, H. Introduction to Multisensor Reconnaissance. Tech. Rep. AFAL-TR-71-43, WPAFB, Oh., 1971, pp. 54-56.

Howell, W. and Goldstein, I. (Eds.) Engineering Psychology. New York: Appleton-Century-Crofts, 1971.

Hudson, R.D. Infrared System Engineering. New York: John Wiley and Sons, Inc., 1969.

Hycon Aerial Photography Handbook. Monrovia, Ca.: Hycon Co., 1969.

Johnston, D.M. Target Recognition on TV as a Function of Horizontal Resolution and Shades of Gray. Human Factors, 1968, 10, 201-210.

Johnson, S.L. Training Devices: Physical Versus Psychological Simulation. Proceedings of the Human Factors Society 19th Annual Meeting, Dallas, Tx., October 1975, pp. 24-28.

Leonard, J., Clarke, H., and Staats, S. Factors Contributing to General Versus Specific Perceptual Learning, J. Exper. Psychol., 1957, 53, 324-329.

Leonard, J. and Wierwille, W. Human Performance Validation of Simulators; Theory and Experimental Verification. Proceedings of the Human Factors Society 19th Annual Meeting, Dallas, Tx., October 1975, pp. 446-456.

Leonard, K., Sowal, L., and Waller, D. The Impact of Weather on the Performance of Electro-Optical Sensors in Europe. Tech. Rep. AFAL-NVA-698CF (contract F33615 70C 1461) WPAFB, Oh., November 1972.

Magnusson, D. Test Theory,. Palo Alto, Ca.: Addison-Wesley Pub. Co., 1967.

Marley, A.A. Conditions for the Representation of Absolute Judgment and Pair Comparison Isosensitivity Curves by Cummulative Distribution, J. of Mathematical Psychol., 1971, 8, 554-590.

Neisser, U. Visual Search, Scientific American, June 1964, reprint.

Nunnally, J.C. Introduction to Psychological Measurement. New York: McGraw-Hill Book Co., 1970.

Pollack, I. Detection of Changes in Spatial Position: Short Term Memory or Motion Perception? Perception and Psychophysics, 1972, 11, 17-27.

Pollack, I. Visual Noise Filtering By Human Observers: I. Sequentially Encoded Information. Tech. Rep. AFCRC-TR-54-57, Bolling AFB, Wash., D.C., 1955, AD #64490.

Potempa, K., Lintz, L., and Luckew, R. Impact of Avionic Design Characteristics on Technical Training Requirements and Job Performance. Human Factors, 1975, 17, 13-24.

Potter, J. Velocity as a Cue to Segmentation, IEEE Transactions on Systems, Man, and Cybernetics, 1975, 5, pp. 390-394.

Rappeport, M. The Role of Redundancy in the Descrimination of Visual Forms, J. Exper. Psychol., 1957, 53, 3-10.

Ratliff, F. Contour and Contrast, Scientific American, 1972, 226, reprint.

R.C.A. Electro-Optics Handbook, Technical Series EOH-11.RCA Commercial Engineering, Harrison, N.J. 1974.

Reed, L. Snyder, M., Baran, L. and Curtin, J. Development of a Prototype Human Resources Handbook for Systems Engineering. Tech. Rep. AFHRL-TR-75-64, WPAFB, Oh., December 1975.

Richmond, J. Survey of Image Quality Criteria for Passive Night Vision Devices, Law Enforcements Standards Program Report, LESP-RPT-0301.00, U.S. Department of Justice, Washington, D.C., June 1974, 18 pp.

Schroder, H., Driver, M., and Streufert, S. Human Information Processing. New York: Holt, Rinehart, and Winston, Inc., 1967.

Sidowski, J.B. (Ed.) Experimental Methods and Instrumentation in Psychology. New York: McGraw-Hill, Inc., 1966.

Sinaiko, W. (Ed.) Selected Papers on Human Factors in The Design and Use of Control Systems. New York: Dover Pub., Inc., 1961.

Snyder, H. Image Quality and Face Recognition on a Television Display Human Factors, 1974, 16, 300-307.

Stich, K. and Helm, D. Research on Night Vision Devices to Enhance Helicopter Pilot Performance, Aircrew Performance In Army Aviation, Proceedings of Conference at U.S. Army Aviation Center, Ft. Rucker, Al., November 1973, pp. 107-110.

Wagner, R., Hogg, D., Richmond, D., Stauffer, D. and Youngling, E. Hybrid Simulation of an Imaging Infrared Target Acquisition System, Proceeding of the IRIS IR Specialty Group, Institute for Defense Analysis, Arlington, Va., May 1975, pp. 1-9.

Wierwille, W. and Fung, P. Comparison of Computer-Generated and Simulated Motion Picture Displays in a Driving Simulation. Human Factors, 1975, 17, 577-590.

APPENDIX A: RESPONSES TO PHOTOGRAPHS OF L3TV DISPLAYS

L3TV Scaling Data

Original subject responses to photographs of L3TV displays are presented in Table A-1. All entires are the category values of scene complexity into which each subject placed each of the 16 photographs

Table A-1. L3TV Raw Scaling Data

Target	Subjects									Deleted
Areas	S-1	S-2	S-3	S-4	S-5	S-6	S-7	S-8	S-9	S-10
1.	6	7	6	7	7	6	7	7	6	4
2.	6	4	5	5	7	5	6	6	6	4
3.	5	4	4	6	6	4	7	6	5	5
4.	5	5	5	6	6	5	5	6	6	2
5.	7	7	7	7	7	7	6	7	7	6
6.	2	3	3	1	2	1	5	2	2	7
7.	1	1	1	1	1	1	2	1	1	7
8.	1	2	2	2	2	1	1	3	3	7
9.	4	4	4	2	4	3	3	3	4	1
10.	5	6	5	3	5	4	4	4	4	3
11.	4	4	4	3	4	3	3	2	4	3
12.	2	2	2	2	2	2	1	3	4	7
13.	4	3	3	3	3	2	6	5	5	7
14.	3	5	3	4	5	4	4	5	4	3
15.	3	3	3	2	3	2	4	4	3	1
16.	1	1	2	1	1	3	2	1	3	7

of video-taped L3TV displays. The photographs are identifed by the number assigned as a "target-area" during the original video recording. Category 1 was identifed as "Least Complex"; Category 4 as "Average Complexity"; and Category 7 as "Most Complex". Included in Table A-1 are the data of a tenth subject which were deleted for reasons cited in the text, supported by comments made by the subject during debriefing (p. 75).

The category number assigned by a subject while sorting photographs was treated as a score or scale value along an equal-interval, unidimensional continiuum of scene complexity in deriving the statistics presented in Table A-2.

Table A-2 Statistics for L3TV Scene Comparisons, N = 9.

Scenes		Statistical Values & Significance Range			
#	Abbreviated Title	$\bar{X}$	s	$s_{\bar{X}}$	$\bar{X} \pm 1.96$ s
1.	Patterson Field Hangars	6.556	0.527	.186	5.53 - 7.0+
2.	Steam Plant	5.556	0.882	.312	3.83 - 7.0+
3.	Inactive Cement Plant	5.444	1.014	.358	3.44 - 7.0+
4.	Fairborn Grain Elevator	5.444	0.527	.186	4.41 - 6.47
5.	Base Hospital	6.889	0.333	.117	6.24 - 7.0+
6.	Microwave & Water Towers	2.333	1.225	.433	0.00 - 4.73
7.	Housing Area	1.111	0.601	.212	0.46 - 1.76
8.	Apartment Complex, Trees	1.889	0.782	.276	0.36 - 3.42
9.	W.S.U. Water Tower	3.444	0.726	.257	2.02 - 4.86
10.	Fairborn Water Tower	4.444	0.882	.372	2.71 - 6.17
11.	W.S.U. Microwave Dish	3.444	0.726	.257	2.02 - 4.86
12.	Church in Woods	2.222	0.833	.295	0.59 - 3.85
13.	D.P.& L. Sub-station	3.778	1.302	.460	1.23 - 6.33
14.	Active Cement Plant	4.111	0.782	.276	2.58 - 5.64
15.	I-675 Overpass	3.000	0.707	.250	1.61 - 4.39
16.	Trebein Site	1.667	0.866	.306	0.00 - 3.37

The data of the tenth individual being excluded, statistics of Table A-2 were computed on a subject group of 9. Scenes are identified by abbreviated titles and numbers correspond to the "target areas" listed in Table A-1.

Verbal Comments and Responses, L3TV

The quotations which follow are based on verbatim notes transcribed during experimental sessions. Extraneous verbalizing, pauses, and incorrect grammar have been deleted for the most part. In some instances incomplete sentences have been completed and some phrases and sentences transposed for greater clarity. Brackets, [ ] , indicate an insertion by the experimenter for clarification or explanation. Phrases selected during preliminary content analysis are underlined. A solid (or unbroken) underlining indicates later classification as a primary scaling criterium while a broken (or dashed) underlining indicates later classification as a secondary scaling criterium.

Comments and Observations During Scaling

S-4: [Spread out all 16 photos and sorted first into crude piles, then started selecting pictures for the least complex end.]

S-5: "I was going to mention that I am an amateur photographer and that might introduce a bias. [We] could have better lighting in this room. There is a bit of glare [and] I'm having trouble with categories 3 and 4."

S-6: "I'm looking through a lot of stuff. That is, the quality of the photographs has to be 'looked through'."

Comments During Debriefing

S-1: "I looked at the overall picture and without trying to consciously interpret what I was looking at -- that is [looking at pictures in Category 7], I think I sorta integrated things like black to white changes, the number of lines, and the different directions of lines of objects or things I could see [looking at the lower categories]. If the trees had about the same amount of light or bright uniformity over a big area, it was simple. But if the light distribution was broken or as the picture began to show more objects, it got more complex."

S-2: "I think I used two general ideas: the amount of complex man-made objects against the surround, that is, the most man-made area equals the most complex. Then I looked at the clarity of the surround. If you had asked for me to scale on clarity, I think I would have given a slightly different order. Maybe you have too many categories -- maybe there should be 5 instead of 7." [Asked what he meant by "clarity"] "I also looked at the number of straight lines -- visible straight lines -- against or versus the sort of garbage of the surrounds."

S-3: "Mostly I arranged them first by the number of objects. Then by the objects I could pick out, that is, by the detail I could detect looking at middle-ranged pictures . The most complex has a lot of stuff -- a clutter of objects in the picture -- picking out one thing here would be difficult."

S-4: "Mostly I looked for what could be seen of man-made structures that were centered or near the central portion. The background portions looked pretty much the same for vegetation for all of them -- did not sort on the background, but how much of the picture was man-made increased the complexity. The lowest complexity seemed to have no particular information -- well, some simple geometric patterns in two of them. Category 3 seemed to have some supporting structures. Category 4 had more substructure within the primary structures. Categories 5 and 6 seemed to contain increasing information on man-made objects. Category 7 seemed to have the most information content -- also I think they are the clearest. If identification is not related to complexity, then maybe they should be reversed. I guess my criteria was the percent of man-made visible against the ground."

S-5: "Under the least complex category there is very little man-made or cultural detail -- it is mostly terrain, at least that's all I can see. Category 2 was basically because there are only 1 or 2 man-made objects in each scene, and they have simple shapes. Category 3 pictures have one main, larger feature and the shapes are more complex, in Category 2 the shapes were like rectangles, in Category 3 the shapes have more sides and lines. In Category 4 man-made objects are larger and the shapes are also larger. Category 5 contains complex shapes -- there are many sub-patterns, that is, sub-objects make up the central man-made object...usually only one. Category 6 appears more complex because these pictures contain two or three -- in fact, several man-made objects. Category 7 is the most complex because it has the greatest number of man-made objects."

S-6: "Really I looked at the man-made objects in the scene -- I noticed that the background seemed to be the same -- what differed was the number of man-made objects, and the variety of shapes." [Asked how he differentiated between which objects should go into Category 3, say, as opposed to Category 5:] "There wasn't the variety of information in Category 3 as in 5, Category 5 had more complex sub-structures."

S-7: "I seemed to use a number of criteria, some of which conflicted and some of which didn't. The criteria may not be along the same scale. That is, the most complex seemed to equal the most information which seemed sorta equal to crispness and detail, sharp edges and a number of discernable objects. High possibility of a false alarm here because there are so many possible targets [looking at the more complex end of the continuum]. At the low end, the pictures appear to have poorer resolution and fewer and simpler objects. If there are targets in these scenes [looking at less complexity], they look like clutter." [Looking toward the central categories:] "Good probability of "hits" if this is a recognition problem because only one thing stands out."

S-8: "For the least complex items I chose the most foliage content in the scene. The more complex scenes not only had foliage but also a larger number of man-made objects which had distinctive shapes. I graded them as to increasing amounts of complexity of man-made objects--which had more detailed type of structure. Category 1 had almost nothing [man-made visible] in them; Category 2 has a small amount; Category 3 had relatively simple geometric structures--circles, squares, rectangles--geometric shapes. Category 4 started having a larger number of these geometric shapes. The question between Category 4 and Category 5 was that 5 has fewer geometric shapes, but has finely structured items. Categories 6 and 7 show an increase in geometric shapes and fine structure encompassing larger parts of the scene."

S-9: "The first thing I used, really, for these black and white photos were the outstanding geometric forms. Prior knowledge did influence me. Categories 6 and 7 had the most picture content. I tended to look at these photos in terms of targets. Category 1 had no visible targets, it was total landscape, no man-made objects. Category 2 has an object visible on the horizon and maybe some houses lower down -- is very vague. If targets, they are not distinct. Category 3 -- each scene has 1 particular and primary target -- two have bridges, or what look like bridges. In Category 4 targets are becoming more distinct -- geometric objects known to man, easily identified. Category 5 scenes have more than one target or man-made object which is identifiable. In Category 6 the objects are more distinct -- there are more number geometric shapes and less foliage -- more straight lines in one scene, more round ones [lines] in the other. Category 7 is to me the most complex geometric shape because I know the steam generating plant."

Edited comments of original S-10: "...First of all... I looked for an abundancy of man-made items mostly. ...And then I realized I was taking advantage by knowing what pictures should go into categories as you would want them. When I found 4 pictures I didn't know what they were, I changed my definition of what was complex because I didn't know even what they were supposed to be. Because I couldn't identify anything, I put them in the most complex category. I realized that if I were a pilot, I would consider what I have put into category 7 the most complex." [Subject did not follow directions, changing the continuum from increasing scene complexity to decreasing identifiability of objects, performing as he presumed a pilot would.]

APPENDIX B. SUBJECT RESPONSES TO PHOTOGRAPHS OF FLIR DISPLAYS.

FLIR Scaling Data

All entries in Table B-1 are the category value of scene complexity into which each subject placed each of the 16 photographs of video-taped FLIR displays. Identification number os the photographs are the same as in Appendix A and target areas displayed in Figures 1,2 and 3.

Table B-2 presents two sets of statistics used in the comparison of FLIR scenes. One set (N=8) excludes data of S-7 while the other (N=9) includes S-7's first scaling judgments.

Table B-1. FLIR Raw Scaling Data

Target Areas	Subjects									
	S-1	S-2	S-3	S-4	S-5	S-6	S-7a[1]	S-7b[2]	S-8	S-9
1.	7	7	7	6	4	5	3	2	7	6
2.	6	4	4	7	7	6	3	4	5	6
3.	5	6	3	7	7	5	7	2	6	5
4.	7	7	6	6	6	6	4	2	5	5
5.	6	6	4	7	5	7	1	3	7	7
6.	2	3	3	3	3	4	5	5	4	1
7.	1	1	1	1	1	1	2	7	1	1
8.	1	2	1	1	1	3	2	7	2	2
9.	4	3	4	2	4	4	1	4	3	3
10.	3	4	2	4	2	3	2	4	3	3
11.	4	4	6	4	5	4	4	5	3	3
12.	2	2	2	2	3	2	5	7	4	3
13.	6	7	6	4	6	3	6	6	4	5
14.	7	6	5	5	7	5	4	1	6	6
15.	5	5	5	2	3	4	2	6	4	2
16.	3	3	4	2	2	3	3	6	2	4

[1] S-7's first scaling, assuming the tactical mission of "Reconnaissance Strike Utility."

[2] S-7's second scaling, assuming the strategic mission of "Low Level Ingress."

FLIR Content Analysis Data

Comments Volunteered During Scaling

S-1: Hmmmmmmmmm. Things are quite different. This is considerably harder than last time.

S-2: Holy Cow! This is harder than the other time. It doesn't look like the same thing. There is more to go on. This is much harder than the last time and I know they are the same scenes but a different ball game, and this is my first go. But Holy Cow, I can't use the same criteria as before and it's bothering me. I shift back and forth. There's sort of an overlap around or between the middle categories 3,4, and 5. I could just about turn these around. I'm suspicious about one picture--it doesn't fit. Especially categories 3 and 4 are giving me trouble. I could lump them together.

S-3: Last time I think I put in too much of the computer image generator point of view. For some reason this seems much harder than last time. It is and is not the same job because last time I did not have such trouble differentiating between scenes and complexity groups.

S-4: (Spreading out all 16 pictures) I'm looking to see what common features I might find within scenes, like tree trunks vs telephone poles. In some of these pictures the brightness is too high. The detail has been masked by brightness in the hospital and steam plant, for instance. There seem to be some similarity among the more complex scenes...Actually, these look as if the field of view has changed. I see some old friends, but the field of view is different. It seems to me they are capturing more real estate or a higher vantage point or maybe the horizon has been attenuated. There just seems to be more information available. The quality is better than the L3TV shots, the atmosphere was attenuated so the horizon appears closer.

S-5: These pictures are more difficult to classify than the earlier L3TV shots, Photographically, these are over-exposed or too white.

S-9: This isn't the same task as before. I'm looking at it in a different light.

Responses During Debriefing, Edited.

S-1: My criteria are rather hard to say. The least complex end has homogeniety of the scene, that is, the least amount of objects or white areas on a dark ground. Middle categories seem to have a fair amount of background homogeniety but with a few recognizable objects. By recognizable I don't mean that I know exactly what they are, but

Table B-2. Statistics for FLIR Scene Comparisons for N=8 and N=9 Groups.[1]

Scenes		Values for N=8 Group				Values for N=9 Group			
#	Abbreviated Title	$\bar{X}$	s	$s_{\bar{X}}$	$\bar{X} \pm 1.96s$	$\bar{X}$	s	$s_{\bar{X}}$	$\bar{X} \pm 1.96s$
1.	Patterson Field Hangars	6.125	1.126	.426	3.93-7.0$^+$	5.778	1.481	.524	2.88-7.0$^+$
2.	Steam Plant	5.625	1.188	.449	3.30-7.0$^+$	5.333	1.414	.500	2.56-7.0$^+$
3.	Inactive Cement Plant	5.500	1.309	.495	2.93-7.0$^+$	5.667	1.323	.468	3.08-7.0$^+$
4.	Fairborn Grain Elevator	6.000	0.756	.286	4.52-7.0$^+$	5.778	1.253	.443	3.32-7.0$^+$
5.	Base Hospital	6.125	1.126	.426	3.93-7.0$^+$	5.556	2.007	.685	1.63-7.0$^+$
6.	Microwave & Water Towers	2.875	0.991	.375	0.44-4.82	3.111	1.167	.413	0.82-5.40
7.	Housing Areas	1.000	0.0	.00	1.0 -1.00	1.111	0.333	.118	0.46-1.76
8.	Apartment Complex, Trees	1.625	0.744	.281	0.17-3.09	1.667	0.707	.250	0.28-3.06
9.	W.S.U. Water Tower	3.375	0.744	.281	1.92-4.84	3.111	1.054	.373	1.04-5.18
10.	Fairborn Water Tower	3.000	0.756	.286	1.52-4.48	2.889	0.782	.276	1.26-4.42
11.	W.S.U. Microwave Dish	4.125	0.991	.375	2.19-6.07	4.111	0.928	.328	2.29-5.93
12.	Church in Woods	2.500	0.756	.286	1.02-3.98	2.778	1.093	.386	0.64-4.92
13.	D.P. & L. Sub-station	5.125	1.356	.513	2.47-7.0$^+$	5.222	1.302	.460	2.67-7.0$^+$
14.	Active Cement Plant	5.875	0.834	.484	4.25-7.0$^+$	5.667	1.000	.354	3.71-7.0$^+$
15.	I-675 Overpass	3.750	1.281	.315	1.24-6.26	3.556	1.333	.471	0.95-6.17
16.	Trebein Site	2.875	0.834		1.25-4.51	2.889	0.782	.276	1.36-4.42

[1] S-7a is included in the N=9 group and excluded from the N=8 group

S-1 continued:
that there are things which seem to stand out from the background, which seem to be trees. At the more complex end there just seems to be more of those objects which stand out.

Retrospectively, my criteria for complexity at the more complex end seems to have changed from what it was moving up the scale. Pictures in category 6 are equal to those in category 7 in terms of white or dark objects against the background, but category 7 seems to have more distinct things. That is, I can see the cultural objects clearer and as more separate from the background while category 6 also has a lot of cultural objects but they are more blended-in or smeared.

S-2: [Expressed dissatisfaction with his sorting, but decided to go with his first try.] I guess I used pretty much the same criteria as before with L3TV. The criteria was the amount of man-made versus the surround. By amount I mean both the area and the number of objects. For example, in the most complex category this picture (#4) has a lot of man-made objects in it. They are scattered but they take up more of the room of the picture. The less complex categories have more foliage. Middle complexity has increasing man-made both in size and quantity, and the most complex are even more so.

If the task were target recognition, I would in fact I could, reverse the ordering of these pictures. I could flip the whole scale right over. In a target recognition task those pictures in category 1 would be most complex in terms of displayed information.

S-3: My criteria were a combination of how many objects which I could recognize as objects and also how clear objects were or how saturated. Because of the blur, you can't see an exact shape. With less complex pictures there is no object I can pick out and be pretty certain what it is. Scenes in category 3 have towers and unique things that don't look natural, they are man-made. Category 4 shows increasing complexity: there are more specific objects I can pick out and make a guess as to what they are in terms of the shape of buildings. The most complex picture (#1, Patterson Field Hangars) has the most objects about which I can guess what they are. That's the clearest picture in the entire bunch. It took me so long differentiating pictures with losts of information from the ones that had information which was washed out.

S-4: These [lower categories] seem to be least complex because I'm not sure what is there; there seems to be more natural things than man-made things. I guess I'm hanging in on 'white = hot'. Therefore anything white is man-made. Not much that is white is in the least complex. My criteria in category 2 was stocks that look like trees vs poles, two bridges in here. I guess I looked for patterns I can recognize. The picture in category 3 has quite a number of poles or trees and a blob in the center I can't identify. [Otherwise] things

S-4 continued:
look much like other [less complex] pictures. There's quite a bit or quite a number of dociles of white in category 4. [These] must be something, but I can't make out what they are. I guess the density per unit area of man-made target is increasing as I go up the scale. There must be more information here [category 5] but the picture is almost obliterated by brightness. If you were looking for a cement plant, it doesn't have to hit you over the head. It is identifiable as a check point although detail is masked by brightness setting. I guess I don't need to see all the detail for a checkpoint. Category 6 has pictures of land, one is busier than the other, but there are more buildings. They are clustered in one picture as opposed to the other in which they are spread out. Whatever the small localized blobs are, they are just too bright. They can be identified as centers of man-made somethings. These are mostly at the foot of the scene. What I put into the most complex are perhaps more distant; the predominance of man-made features in the upper part of the scene. It seems as if there is more to be seen with these most complex figures. An operator might be scanning such a scene with knobs, if held in a fixed position mode.

This is difficult. I'm not sure that what I have done is correlatable: the density of the white spots coupled with the idea of what I could distinguish, as if white is something that ought to be seen. I don't know if curiosity should be a point of view or necessity should be a point of view. It is the task as before in a mechanical way, but the problem is more complex because there are more contrasting things.

S-5: I think the area, the amount of the photo that has terrain information or something in the foreground that is white and washed out was a criterium. Looking at the most complex pictures this picture has foreground information, building features off to the middle-left. It is not washed out. In this picture I could definitely make out a couple of water towers and several buildings. In another there are a couple of silos (being a farm boy I know thats what they look like). But that's the trouble. In my own mind they may not belong in the most complex category, maybe lower. One very valid thing I have learned from this is that you have to treat FLIR as very different from L3TV.

At the least complex end, I used as criteria very limited or relatively few cultural features, man-made objects, or buildings. Then [in category 2] there were a few more features; one had a water tower fairly well defined, and the other had a greater number of man-made objects. A man-made object, in contrast with the background, is lighter or whiter because of emissivity. There was a lot of sky, a good 1/3rd is sky or the bottom was whited out so you didn't get as much information [in less complex categories] . In category 3 although there was not so much cultural detail, the natural objects, like trunks of

S-5 continued:
trees, stood out more, particularly in one picture. There was a more uniform distribution of bright areas and very little sky.

The main reason these scenes are in category 4 is it seems like there is one cultural detail which is of significant size. One picture is a structural testing facility and in the other there is a water tower and horizontal buildings. Category 5 again shows large cultural features and more detail to these features that are visible. This I assume is the hospital and here I can make out the structure of the radar or microwave dome. There are also background features you can make out.

These two pictures [in category 6] contain many well defined cultural details. Its evident that electric power poles are there, on the other you can tell detail of some of the smaller buildings: gables, smoke stacks, a couple of water towers, and some sort of information about a large figure. I can see windows in some of the buildings in the foreground. The most complex scenes contain primarily many man-made objects, but they also appear a little clearer than in category 6.

S-6: I just looked at the amount of cultural information, the amount of man-made objects. The least complex have the fewest number because I can't see much detail. The more complex scenes have the most white. I seemed to have used the same criteria as when we sorted the L3TV, I didn't seem to have any reason to change. Having done the other task did seem to influence me because I used, or tried to use, the same criteria. The quality of the reproductions isn't very good.

S-7: I kept asking myself 'What is the least complex?' There were three contrary criteria I was using. These were:

A. When I couldn't find a target because of masking. (Masking= foliage or a building of no particular importance. Take this picture for example: these trees and structures might mask a target. I guess I was assuming a tactical target). [Subject discussing scene #9 which had the Wright State University Water Tower distinctly in the foreground. S-7 obviously saw the tower, decided it was not a target, and placed the picture in the least complex category.]

B. A picture is complex if when I detect a target there are also many other things that are target-like. Its sort of a false alarm problem with complex scenes, like this picture. [Subject pointing to #3, Inactive Atlas Cement Plant, judged to be the most complex of all scenes.]

C. If I'm looking for something large like a structure, it ought to be easily different from regular types of buildings. That is, targets have unique shapes or patterns. This picture has a well defined target. It is a LOC (line of communication) which could be a target, yet areas of possible defense of that target are clear except for

S-7 continued:
large masked areas that I can't do anything about anyway. [S-7 was discussing scene 15, the I-675 Overpass, which he judged as less complex, placing in category 2.]

I found myself setting up a different task than last time. This could be used to do a task that I am familiar with. With the early categories there was no job that I could do. As I went up the scale there may be a job to do. By the most complex categories I was going to have to spend a lot of time either doing a job or finding out if I had a job to do. Anything that takes time is complex. [When asked what was the task he had assumed, the subject responded "AT: Ground Reconnaissance Strike." That the subject took his task quite seriously was evidence during scaling by an apparent anxiety response to some scenes, as perspiration, breathing, and muscular or postural changes.]

[Describing the least complex category] Here is a big image with a lot of clutter and masking. I said to myself "I can't use that so I will ignore it." [Describing The Base Hospital scene #5.] This one [looking at #9] has no undifferentiated targets which are not masked, or if masked, I can't find them. In category 2 [already partly covered by discussion of #15] this picture [#8, Apartment Complex Among Trees] has no value to be gained. It is almost total masking. This one [#10, Fairborn Green Water Tower] is garbage again. A water tower couldn't be a target. This picture in category 3 is clearly an airdrome with parked aircraft and no evidence of defenses seen. If that's my job, O.K.

In category 4, one picture is obviously an active industrial center with outlying buildings. The reason it is not ranked as simple is that many areas here could contain threats. That applies to all of the pictures here. The general arrangement [of pictures in category 5] indicates a man-made geometry, like L.O.C.s (lines of communication). There is also some masking but I feel I can partially see through, or should try to see through. These are probably the nastiest images to give a guy [#6 & #12]. They are interesting pictures, too interesting. They occupy too much time, which is dangerous on a recon strike mission. [While discussing the one picture in category 6:] there is an obvious target in the center, but I have a TA problem (terrain avoidance problem). If making a pass, I may tangle up in wires. Category 7 could almost be folded over into category 1. This is chock full of fascinating things: a bridge, LOCs, strange structures. They look like important things. But it is also full of threats. This looks like a prime target area.

S-8: Well, let's see. For group 3 I have given a slightly below average complexity, but I have somewhat ignored the lower portion detail. These big white blobs I know are parts of Wright State University buildings. The least complex pictures are almost totally terrain. The

S-8 continued:
most complex have the highest percentage of man-made objects in the picture or area of the picture. That is, except for those scenes in category 3 where the bottom of the picture is essentially blank. If there had been greater detail on the pictures of group 3 I might have classified them differently. Category 4 is middle ground. Between categories 3,4, and 5 I guess I personally decided... influenced by the numbers and amounts [of objects] rather than the complexity of the objects.

S-9: I ranked them on what targets of man-made objects I could get out of the picture. On the basis of the amount of the picture that showed heat. When I started out I really began like I was looking at television. Then I realized I was looking at FLIR, at heat producing objects, and I sorted them on the complexity of the object that was providing the heat. Then I had to re-sort a little to separate what I know about the heat producing capability of targets.

The least complex had few targets. The next group have distant targets with some secondary targets. Category 3 has fairly obvious closer-in targets which are more easily distinguished, with further out targets also recognizable. I don't know about the middle category, I just picked [the picture there] as a medium. It has different types of man-made objects. Category 5 has quite a few different things, more distinguishable things to look at. The picture quality is bad but more interesting in category 6. There is more detail here. In this one I can see the water level in a tower. In another there are three aircraft parked on the ground and you can see that there are hangars and one hangar door is open. The third picture [#6] has a plume on the horizon. [The picture in] category 7 I know is the base hospital with stack and a lot of things, but very poor picture quality.

S-7b: [After second sorting in which "Complexity for Low Level Ingress, a strategic mission" was the assumed task. These data were not included in the content analysis.]

The key example is this picture [#6] in category 5 which has a blob of cultural detail. [I] would have made this more complex except for the unique, identifying tower on the horizon. Unique=easily identified [by salient cues]. In pictures in the most complex category there are so many cultural details which are evidences of man-made things, a number of them, but they are not distinguishable. That is, [they] could be [found] anywhere. That's different compared to the scene in category 3 [#5, Base Hospital]. That's an easily distinguishable, memorable bunch of objects. [Hence, it is less complex.]

Category 1 [pictures] have exceptionally big, special, high isolated landmarks. Unique objects imbedded in the background. [Category 2 has] unique objects and here is something on the horizon. But category 2 is not quite like 1: There is an airdrome [which] might

S-7b continued:
be special, might not. In the other pictures, one had two pointy water towers and water towers are not usually pointy, and a grain elevator that is a bit different.

The roundess of the water towers [in category 4] leads to increasing complexity because there are so many round water towers [that] it is hard to tell which, while the pointy one [#2 is peculiar. The one thing which helps [make scene #6 more easily identifiable] was the cylindrical tower in this picture in category 5 on the horizon. The rest of the area was undifferentiatable from many other such areas. With this picture [#11], my only problem was which wing I wanted to loose. [Category 6 shows] a highway overpass, radar dish, power station. There are a lot of these things scattered about the countryside. How do you know it is which one? Notice the possibility of foldover again! The most complex category could have been the simplest or least complex, yet it is complex because I can't easily find what I am looking for............

[Here the subject scaled more like the 10th subject where recognition=simple and non-recognition=complex].

APPENDIX C. MATCHING STUDY ADDENDUM

Instructions

Covered over in front of you is a display of the 16 FLIR or 16 L3TV photographs which you have already scaled for complexity. This set is ________________. They are arranged in a 4 by 4 random layout. To your left, covered by a sheet of paper, is a randomly arranged stack of 3½ by 5 inch colored photographs. Your job will be to match the color photographs up with the E/O Sensor pictures by placing the color photographs up with the E/O Sensor pictures by placing the color photograph on top of it's match. I must caution you that the color photographs have a much wider field of view than do the FLIR or L3TV reproductions. When we get started you should work as fast and as accurately as possible for you will be both timed and scored. When you have finished the job, shout "Done!" and I will stop the clock. When we finish with this job, you will do the same with the other set of E/O Sensor photos.

Now, please close your eyes and keep them closed while I remove the masking papers. Nod when I ask if you are ready, keeping your eyes closed. Then when I say "go", open your eyes and go to work. Are there any questions?

Stimulus Materials

The eighteen different 4 x 4 arrangements or layouts of the 16 photographs of L3TV and FLIR displays presented to the subjects were established by random card drawing and are displayed in Table C-1. The 3½ x 5 inch color photographs placed in the subjects' hand were randomly sequenced using a table of random numbers. The order of these small photographs, identified by the target area encompassed within the scene, is presented in Table C-2 for each matching session for each subject.

Table C-1. Layouts of Display Photos and Presentation Order

Sub./ Order	1st Layout	2nd Layout	Sub./ Order	1st Layout	2nd Layout
	5 7 14 4	3 5 1 16		16 12 1 6	8 15 14 16
S-1	11 8 10 13	15 14 2 4	S-6	9 15 3 4	11 4 9 6
F:L[1]	6 2 9 12	11 9 13 8	L:F	7 14 13 2	3 13 10 1
	1 16 3 15	10 7 12 6		5 8 10 11	12 5 2 7
	8 11 6 12	5 6 13 8		12 8 5 11	6 9 13 8
S-2	9 4 3 14	4 12 16 7	S-7	2 13 9 7	12 5 4 11
L:F[2]	15 7 1 10	15 8 11 9	F:L	16 10 6 1	3 1 15 14
	5 16 2 13	10 14 3 14		14 15 4 3	2 7 10 16
	14 1 2 11	13 6 10 4		13 11 5 9	14 10 15 13
S-3	15 6 10 9	1 16 9 15	S-8	3 1 12 7	16 2 3 11
F:L	8 13 12 3	11 8 2 7	L:F	10 8 2 6	6 4 1 9
	4 7 16 15	5 14 3 12		14 16 15 4	5 8 7 12
	2 14 13 6	15 11 10 3		9 2 4 11	7 13 1 15
S-4	16 8 15 4	6 8 14 9	S-9	3 8 14 5	12 4 10 3
L:F	11 7 12 3	12 5 16 2	F:L	15 13 7 12	16 11 2 9
	4 1 16 5	13 1 4 7		1 16 6 10	14 8 6 5
	1 11 9 15	14 10 6			
S-5	2 5 13 12	15 5 4			
F:L	8 3 7 6	2 11 8			
	16 4 10 14	12 13 7			

[1] The FLIR Set was presented first followed by the L3TV Set.

[2] The L3TV Set was presented first followed by the FLIR Set.

Table C-2. First and Second Presentation Orders of Color Photographs For Each Subject

S-1		S-2		S-3		S-4		S-5		S-6		S-7		S-8		S-9	
1[1]	2[2]	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
11	7	3	16	4	12	10	5	6	4	15	6	13	14	5	12	3	7
2	1	1	14	6	9	12	9	12	12	12	8	10	9	3	7	10	2
4	6	7	11	7	11	5	3	11	1	13	4	14	7	9	15	4	9
6	16	11	8	9	13	11	15	1	11	4	16	12	8	11	4	6	3
10	2	4	10	15	7	5	6	4	7	3	12	2	3	16	3	5	4
13	10	6	15	2	1	7	2	3	14	9	13	1	11	12	2	12	16
3	9	9	4	12	3	14	10	5	5	10	2	3	16	10	9	15	14
1	14	5	13	5	15	4	1	15	9	14	1	6	13	2	10	13	10
8	5	15	7	16	8	16	8	10	10	16	7	16	2	7	14	1	1
14	13	10	1	8	4	13	11	14	15	2	9	7	1	13	8	7	5
15	4	13	5	1	2	8	4	7	13	8	14	4	12	6	16	8	8
12	15	8	9	11	10	2	13	8	6	7	5	11	6	1	6	14	11
5	12	14	2	14	5	1	14	9	3	1	10	5	10	4	5	9	13
7	3	16	3	10	6	15	16	16	8	6	11	15	15	8	11	11	12
9	11	2	12	3	16	3	12	2	2	11	15	9	4	15	1	16	6
16	8	12	6	13	14	9	7	13	11	5	3	8	5	14	13	2	15

Table C-3. Matching Errors For L3TV and FLIR Displays

Scene Presented	L3TV Mistakenly Matched With:							FLIR Mistakenly Matched With:					
	6	7	8	12	13	15	16	6	7	8	13	15	16
6	-	1	0	0	0	0	1	-	2	0	0	0	0
7	1	-	1	0	0	0	1	0	-	0	0	1	0
8	1	0	-	0	1	1	1	2	0	0	0	0	0
12	0	1	0	-	0	0	0	-	-	-	-	-	-
13	0	0	0	0	-	1	0	0	0	-	-	0	1
15	0	0	0	0	0	-	2	1	0	0	0	-	1
16	0	1	3	1	0	0	-	0	0	1	1	1	-

[1] Column presents the first presentation order for this subject

[2] Column presents the second presentation order for this subject