P. G. Schwarz
Mapping Division
Survey Department of Rijkswaterstaat
Delft, Netherlands

A Test for Personal Stereoscopic Measuring Precision

The H-test has been developed for determining the personal precision and reliability of an observer in stereoscopically measuring elevations.

Introduction

Recently such matters as precision, reliability, and error detection have become extremely important in the geodetic world. In photogrammetry, especially within the field of aerial triangulation, considerable effort has been invested in implementing practical applications of modern statistical theories. These applications will induce improvements, such as increasing the precision and reliability of photogrammetric networks, that are better attuned to the goal of the measurements.

Abstract: *During the last decades several tests have been developed, aimed at the determination of a person's natural stereoscopic acuity. However, experience has shown that a sufficient stereoscopic acuity does not guarantee a sufficient stereoscopic measuring capability. Therefore, a test has been developed, especially aimed at the determination of an observer's personal stereoscopic measuring capability. This so-called H-test is as close to a photogrammetrist's daily work as possible: i.e., to perform the test, a candidate is asked to do just that which is, or will be, his main task. The H-test proved to be a helpful tool in training new employees. Furthermore, the differences in personal stereoscopic measuring precision being considerable, it enables the organization to make allowance for these differences when distributing work. Moreover, two years' experience with the H-test led to the recognition of a specific cause of a weak or even bad stereoscopic measuring capability: i.e., a difference in light sensitivity of the eyes.*

However, the greater part of the photogrammetric observations, especially in photogrammetric detail measurements, will not be affected by these theories, simply because only single measurements are involved.

In measuring cross-sections, Digital Elevation Models (DEM), and also in plotting, points are only measured once, so that only a rough check on large errors is possible. Therefore, one has to rely completely on the precision and reliability of the photogrammetric observer. Consequently, thorough training and a scrupulous working method are absolutely necessary. However, even this has been insufficient. The quality of the eyes and factors of cerebral nature also influence the quality of photogrammetric measurements.

Apart from the fact that not everybody has the ability of stereoscopic vision, there are considerable differences in personal stereoscopic acuity. Several tests for the determination of the acuity of stereoscopic vision have been designed. One should, however, be cautious in applying these tests, because quite often someone performing badly on test A may get a good score in test B. At the same time experience has shown that satisfactory or even good results do not guarantee that the candidate will indeed make a good observer.

Frequently, even after years and after an extensive training, a new employee may not acquire a sufficiently precise and reliable stereoscopic measuring capability. As will be obvious, in this case the consequences for the employee as well as for the organization are rather unpleasant. And it is obvious, too, that there are several, extremely

Photogrammetric Engineering and Remote Sensing,
Vol. 48, No. 3, March 1982, pp. 375-381.
0099-1112/82/4803-0375$02.25/0

practical reasons for asking for further research within the area of stereoscopic measuring capability.

The H-Test

For this reason, at the Survey Department of Rijkswaterstaat (Ministry of Transport and Public Works) a test has been developed for determining the stereoscopic measuring capability of the individual photogrammetrist, in particular, for determining the personal precision and reliability of an observer in stereoscopically measuring elevations.

This test (the H-test) is very close to representing the daily work of a photogrammetrist. The observer is asked to measure heights of some hundred points in a model. The model consists of black-and-white exposures of a strip of dunes along the Dutch coast, taken with a wide-angle camera at the scale 1:6000 (Figure 1). The photography is of a reasonable quality.

After relative orientation and leveling, using five ground control points, the model is fitted to a map at a scale of 1:2000. The hundred points are situated on eight cross-sections, and all points are marked on the map. Therefore, the operator recovers the points with the aid of the map, because the points are neither signalized in the terrain nor marked on the photos.

The influence of limited precision in identifying a point is eliminated by choosing all points in flat terrain. The points are of different degrees of difficulty, a number of them being very difficult to measure, due to a lack of detail for correlation. This is especially true of the points on the beach. The measurements take one to one-and-one-half hours.

The analysis of the results provides the information listed in Table 1 (the analysis method is explained in an appendix). The first group of parameters (index 1) originates from calculations that include all points. The second group (index 2) originates from calculations that exclude the 13 points on the beach which are very difficult to measure. All variables are given in micrometres at the scale of the exposures.

Standard Deviation

A group of 73 persons—consisting of mostly (more or less) experienced photogrammetrists (54), some operators still in training (11), and some totally inexperienced observers (8)—participated in the H-test in 1979. Their standard deviations are shown in the Figures 2 and 3. Interpretation of these distributions resulted in the following: Apart from one exception, the group with a standard deviation $10 < \sigma_1 \leq 20$ μm (Figure 2) appeared to consist of observers (46) having at least one year, but mostly several or many years of experience. The group with a standard deviation $\sigma_1 > 20$ μm consisted for 70 percent of participants those having only a few weeks experience, or of totally inexperienced observers.

Leaving the points on the beach out of the calculations (Figure 3) has only little influence on the results of the experienced group. Their measurements of the difficult points are only slightly less precise. The same holds for the inexperienced participants, with only little difference between σ_1 and σ_2. They measure all points with the same lack of precision. However, considerable differences, up to 12 μm, were found in the results of the group having at least some training. They are able to measure the less difficult points, but on the beach they still get into trouble. This means that the H-test inherits great power from these points. Exclusively considering the experienced group—which statistically means that the other participants are considered to belong to another population—the distributions approximately conform to the normal distribution. The expectation is about 15 μm, similar to 0.10‰ of the flying height, while apart from a few series, the complete histogram is limited between 10 and 20 μm, similar to 0.07 to 0.15‰ of the flying height.* The personal results σ_1 and σ_2 are highly reproducible. In 1979 and 1980 many operators participated in the H-test for a second time, a number of them more times (five, six, or seven times). Differences seldom amounted to more than 15 percent of σ_1 and σ_2, respectively.

Mean Error

The opinion that many observers measure with a certain constant mean error is often heard in the photogrammetric world. According to this opinion, this should in no way affect the results because all points are measured with the same systematic error. Only when continuing the work of another observer in the same model must this phenomenon be considered. The H-test, thus far not specifically aimed at testing this statement, gave no arguments to prove the correctness of this opinion, nor sufficient grounds for negation.

The H-test, however, did prove the existence of a second type of systematic error, which is not harmless at all. As a result of the chosen measurement and calculation method, the stochastic errors in the observations of the signalized ground control points exercise a considerable influence upon the parameters $\bar{x}_1$ and $\bar{x}_2$. Therefore, it is impossi-

* Using wide-angle exposures with 60 percent overlap, it can be stated roughly that $\sigma_{x\text{-parallax}} \approx 3/5.\sigma_{\text{height}}$. With an optical enlargement of 6×, σ_{height} corresponds with

$$s = 206265\left(\frac{3/5.\sigma_{height}}{250000/6}\right)$$

In this formula, s is the standard deviation of the observers' measuring precision, expressed in seconds of arc. It can be seen that $10 < \sigma_1 < 20$ μm corresponds to $30 < s_1 < 60$ sec of arc.

FIG. 1. Exposure used in the H-test.

TABLE 1. TEST RESULT PARAMETERS

	all points	beach points excluded
standard deviation	σ_1	σ_2
mean error (systematic error)	$\bar{x}_1$	$\bar{x}_2$
number of outliers >40 μm	ua_1	ua_2
number of outliers >80 μm	ub_1	ub_2

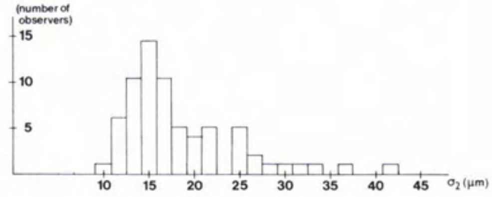

FIG. 3. A histogram of standard deviations versus number of observers when points on the beach are excluded.

ble to base reliable statements on outcomes of the parameters $\bar{x}_1$ and $\bar{x}_2 < 10$ μm. Nevertheless significant systematic errors up to 20 to 50 μm were found in the results of four observers.

This type of systematic error really means trouble, since it appears only in terrain height measurements, as long as no well-defined points are measured. In measuring well-defined points, for instance signalized points, the observer obviously is able to eliminate his personal systematic error. A plausible explanation for this phenomenon might be that x-parallaxes find much better expression when measuring well-defined terrain points. Regarding the personal systematic errors, a number of questions remain. One of them is especially interesting; that is, Is it possible to redress a systematic error by specific training?

OUTLIERS

Together, the parameters ua_1, ua_2, ub_1, and ub_2 give an indication of the reliability of an observer's measurements. An occasional exception left out, $ub_1 \leq 2$ and $ub_2 \leq 1$, holds for an experienced photogrammetrist. Candidates having some, but yet insufficient, experience are easily recognized. Their ub_1 will be much bigger than ub_2.

CORRELATION TO AGE

Although one would expect at least some correlation between the results of the H-test and the age of the observers (for instance, slightly less precise results for observers older than 50 or 55 years), thus far no significant correlation was found. However, for that matter the test group consisted of an insufficient number of observers of that age.

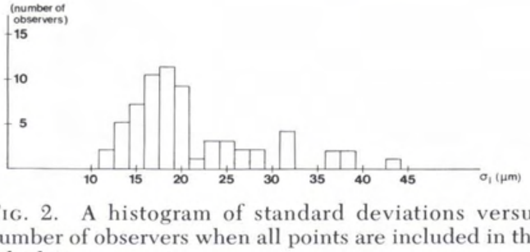

FIG. 2. A histogram of standard deviations versus number of observers when all points are included in the calculations.

FURTHER APPLICATION OF THE H-TEST

Experiences with the H-test resulted in two sharp, and in practice manageable, criteria being set for photogrammetrists who are in the service of the Survey Department:

- to be able to perform all usual measurements: $\sigma_1 \leq 20$ μm; and
- to be able to perform measurements limited to planimetry: $\sigma_1 \leq 30$ μm.

These criteria are based first on the requirements of the photogrammetric work of the Survey Department and second on the fact that most of the photogrammetrists are able to measure with an even better precision. Moreover, since the middle of 1979 the H-test has been used to follow the progress of operators in training. This has resulted in the following (Table 2):

- Having experience only with some simple exercises, almost every candidate scores a standard deviation $\sigma_1 > 30$ μm. Only twice did a nearly inexperienced candidate score a standard deviation $\sigma_1 < 20$ μm.
- The rate of progress differs from candidate to candidate. Some candidates had already met the criterion $\sigma_1 < 20$ μm after one training month, but most needed two or three months. However, some candidates needed half a year or even longer to meet this standard. The training of candidates showing no progress at all will be broken off.

The number of candidates for whom training must be broken off is not inconsiderable. It should be borne in mind that these candidates satisfactorily passed two tests on stereoscopic acuity at the time of their recruitment. If progress was not forthcoming, an ophthalmologist was consulted after deliberation with the Governmental Medical Department. In not a single case was a relevant eye defect found. All this led to the following thesis:

As long as the separate causes which exercise a negative influence on the stereoscopic measuring capability are unknown, one is obliged to teach a candidate-photogrammetrist to measure stereoscopically before it is possible to determine whether he is able to learn it or not.

TABLE 2. EXAMPLE OF A TRAINEE'S PROGRESS

	σ_1	σ_2	$\bar{x}_1$	$\bar{x}_2$	ua_1	ua_2	ub_1	ub_2
totally inexperienced	45	40	+2	+3	33	25	11	3
after two months training	20	20	+8	+8	4	3	1	0

At this moment the possibility to compile a training program, specifically aimed at height measurements, is being investigated. This program would be initiated to test whether or not it is possible to determine very soon after employment if stereoscopic measuring capacity can be developed satisfactorily.

Preliminary analyses show that training in one model with points distributed like those in the H-test will give highly correlated results. Although the observer only measures the model five times a day, he very soon puts his floating mark on the same height, even if measurements are separated by days. Of course, the training program itself precludes cheating.

RELATION TO OTHER TESTS

The relation to three other tests for stereoscopic acuity have been investigated:

(a) The TNO test for stereoscopic vision (random dot texture stereograms),
(b) the SSPO test for peripheric zones of the retina, and
(c) the Howard–Dolman test ("three-rod test").

Tests (a) and (b) are included in the recruitment procedure. These tests enable the employer to establish absence or defects of the natural capability for stereoscopic vision. Apart from that, the results of tests (a), (b), and (c) appear to have no predicting power with respect to stereoscopic measuring capability, which should be developed afterwards.

LIGHT SENSITIVITY

Thus far investigations made within the Survey Department into the causes of a weak or even bad stereoscopic measuring capability revealed one cause, namely, some kind of dominance of one of an observer's eyes. With three photogrammetrists it was found that during measuring stereoscopically one of their eyes dominated obviously. In other words, one half-mark is observed more sharply or intensively than the other.

Two of these three observers appeared to have this kind of dominance, but only to a low degree. Significant influence on their measuring precision was found only in situations where measuring indeed is difficult, for instance at places with only few correlation possibilities. In the H-test they therefore scored $\sigma_1 = 30$ μm but $\sigma_2 = 20$ μm. One of the third observer's eyes dominated to such a large extent that the criteria were even unapproachable for him, that is, σ_1 and σ_2 both being 40 to 50 μm.

Using illuminated marks adjusted to different intensities of light, it appeared to be possible to compensate for the effect of this type of dominance. The observer himself adjusts the marks in such a way that both marks seem to be of equal brightness to him. Thus, apparently, this kind of dominance is caused by a different light-sensitivity of the eyes, at least insofar as point-shaped sources of light are concerned. The observers prefer the plate illumination to remain balanced.

The difference between the light sensitivities of the eyes being not too big, it appeared to be possible for both observers first mentioned to meet the criterion $\sigma_1 < 20$ μm when measuring on an instrument with illuminated marks adjusted to the observer's personal needs (Table 3). The third observer thus far did not succeed in fully meeting this criterion, although his progress was astonishing (Table 4). His measurements on the beach are still of a poor quality, as can be seen from the difference between σ_1 and σ_2. However, by restarting his training, it may be possible that he also will meet the criterion after some time. For him it appeared to be necessary to color the faint illuminated mark, in order to prevent it from vanishing in light areas of the photo.

CONCLUSIONS

(1) The H-test proved to be a reliable test of personal stereoscopic measuring precision, giving highly reproducible results. It is a helpful tool in training new employees. Furthermore, it can be an important factor in quality control.

TABLE 3. EFFECT OF COMPENSATING THE DOMINANCE

	σ_1	σ_2	$\bar{x}_1$	$\bar{x}_2$	ua_1	ua_2	ub_1	ub_2
without compensation	28	20	−15	−12	16	7	5	1
using unbalanced illuminated marks	17	15	−11	−11	3	2	0	0

TABLE 4. THE THIRD OBSERVER'S PROGRESS

	σ_1	σ_1	$\bar{x}_1$	$\bar{x}_2$	ua_1	ua_2	ub_1	ub_2
without compensation	51	47	+10	+10	52	41	23	14
using unbalanced (colored) illuminated marks	31	20	−2	+2	8	3	5	1

(2) A sufficient stereoscopic acuity does not guarantee a sufficient stereoscopic measuring capability.
(3) At the beginning of his training almost every candidate-photogrammetrist starts with a measuring precision $\sigma > 30$ μm (at photo-scale).
(4) As far as can be seen now, 70 to 90 percent of the candidates can reach a precision $10 \leq \sigma \leq 20$ μm by training and further experience. The time they need to reach this differs greatly. Most need two or three months, while some need half a year or even longer. It should be borne in mind that a candidate-photogrammetrist has already successfully passed tests of stereoscopic acuity.
(5) The H-test proved the existence of a type of personal systematic error (size sometimes 2σ to 3σ), which really means trouble, since it appears only in terrain height measurements, as long as no well-defined points are measured. When measuring well-defined points, for instance signalized points, the observer obviously is able to eliminate his systematic error.
(6) As long as the separate causes that exercise a negative influence on stereoscopic measuring capability are unknown, one is obliged to teach a candidate-photogrammetrist to measure stereoscopically before it is possible to determine whether he is able to learn it or not.
(7) One of the deficiencies appeared to be a different light sensitivity of the observers' eyes. If this difference is not too large, it appeared to be possible to compensate for it in instruments using illuminated marks.
(8) A number of questions remain, for instance questions concerning personal systematic errors, correlation to age, and other deficiencies causing a weak or even bad stereoscopic measuring capability. In conclusion (4) it is stated that 70 to 90 percent of the candidates can meet a standard of $10 \leq \sigma \leq 20$ μm. Is it 90 or 80 percent, or perhaps even 70 percent?

It would be very valuable if further research could be done in cooperation with a greater number of (candidate-) photogrammetrists.

ACKNOWLEDGMENT

I would like to mention the intensive cooperation of Mrs. J. G. van der Kraan and ing. A. D. Baars in the development and application of the H-test.

REFERENCES

Anson, A., 1959. Significant findings of a stereoscopic acuity study, *Photogrammetric Engineering*, Vol. 25, No. 4, pp. 607-611.

Bedwell, C. H., 1971. The eye, vision and photogrammetry, *The Photogrammetric Record*, Vol. 7, No. 38, pp. 135-156.

Blank, A. A., 1960. Space perception through stereoptical instruments, *International Archives of Photogrammetry*, Vol. XIII-4.

Dwyer, R. F., 1960. Visual factors in stereoscopic plotting, *Photogrammetric Engineering*, Vol. 26, No. 4, pp. 557-564.

Fagerholm, P. O., 1960. Testing stereo vision quality, *International Archives of Photogrammetry*, Vol. XIII-6.

Krakau, C. E. T., 1969-1970. A device for the automatic, continuous recording of stereoscopic visual acuity, *Photogrammetria*, Vol. 25, (1969-1970) No. 4, pp. 115-123.

Lorenz, G. G., 1960. Measurements of the effect of unbalanced illumination upon elevation readings with a moving floating mark in a stereoplotter model, *International Archives of Photogrammetry*, Vol. XIII-4.

Moessner, K. E., 1955. A simple test for stereoscopic perception, *Photogrammetric Engineering*, Vol. 21, No. 3, pp. 331-339.

Palmer, D. A., 1960. Variations in binocular acuity and their influence on contour measurements, *The Photogrammetric Record*, Vol. 3, No. 16, pp. 357-369.

Reuss, J., 1960. Die einstellgenauiigkeit als funktion der Helligkeitsunterschiede im messbild, *International Archives of Photogrammetry* Vol. XIII-4.

Zorn, H. C., 1965. An instrument for testing stereoscopic acuity, *Photogrammetria*, Vol. 20, (1965) No. 6, pp. 229-238.

———, 1974. Binocular vision, *Textbook of Photogrammetry*, I.T.C., Enschede.

APPENDIX

ANALYSES OF THE OBSERVATION SERIES

A series of observations from a person is further referred to as a series.

(1) The observations are transformed to the system of the control points, at the scale of the exposures, by an overdetermined, three-dimensional similarity transformation. After

that, the *XY*-coordinates are neglected in the analysis.

(2) For each cross-section point a mean height observation is calculated from the observations of the participating persons. Outliers are omitted. In this way, besides the series originating from observations, a series consisting of calculated mean height observations is obtained.

(3) For each series and for each point the difference between mean and actual height observation is calculated. Simultaneously, for each series the numbers of outliers *ua* (>40 μm) and *ub* (>80 μm) are counted.

(4) Next, the mean error $\bar{x}$ and standard deviation σ are calculated using the well-known formulas

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

- if ub <5, outliers >80 μm are excluded from the calculations;
- if ub $\geq$5, outliers >160 μm are left out.

(5) *ua*, *ub*, $\bar{x}$, and σ are calculated twice, once including all points (resulting in ua_1, ub_1, $\bar{x}_1$, and σ_1), the second time excluding the 13 points on the beach (resulting in ua_2, ub_2, $\bar{x}_2$, and σ_2).

REMARKS

(1) In the first instance (see remark 2), for reliable results one should choose not too small clusters of series, e.g., at least 10 series, while the more experienced observers of a group should have the most influence. All series of a group should have been measured in an unchanged model, as the disturbing influence of differences in the relative orientation will be significant. A group consisting of series with $\sigma_1 > 30$ μm should be re-analyzed, after eliminating these series.

(2) Afterwards, for practical reasons, it might be advisable to select two extremely precise measuring operators ($\sigma_1 \leq 12$ μm). To test a third candidate-observer, it will be sufficient to create a group consisting of only these three. On an analytical plotter one is able to perform a relative orientation with high precision. Therefore, it is expected that in this case it will be possible to compare the measurements of a candidate with a formerly determined series of calculated mean height observations.

(3) The analysis method given above results in slightly biased estimations σ, due to neglecting the stochastic variability of the calculated mean height observations. The maximum effect is approximately 1 μm with the best operators (σ = 10 μm). In reality they are roughly 10 percent more precise.

(4) The test has been performed on several types of instruments (Wild A8, Galileo G6, and Kern PG2, all having handwheels). No significant influence of the type of instruments was found, nor was there any influence caused by differences of optical enlargement; the H-test was mainly performed with a sixfold optical enlargement and occasionally with a fourfold enlargement.

(Received 12 May 1981; accepted 1 August 1981; revised 16 September 1981)