34389

NAME OF AUTHOR/*NOM DE L'AUTEUR* KENNEY, Constance Franklin.

TITLE OF THESIS/*TITRE DE LA THÈSE* Estimating the Reliability of the Short Forms of the WISC-R

UNIVERSITY/*UNIVERSITÉ* of Alberta

DEGREE FOR WHICH THESIS WAS PRESENTED/ *GRADE POUR LEQUEL CETTE THÈSE FUT PRÉSENTÉE* M. Educ.

YEAR THIS DEGREE CONFERRED/*ANNÉE D'OBTENTION DE CE GRADE* 1977

NAME OF SUPERVISOR/*NOM DU DIRECTEUR DE THÈSE* Dr. Henry Janzen

DATED/*DATÉ* Sept. 21/77 SIGNED/*SIGNÉ* Constance Kenney.

PERMANENT ADDRESS/*RÉSIDENCE FIXE* 13822 Ravine Drive
Edmonton Alberta
T5N 3M1

THE UNIVERSITY OF ALBERTA

ESTIMATING THE RELIABILITY OF THE SHORT FORMS OF THE WISC-R

by

© CONSTANCE FRANKLIN KENNEY

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE
OF MASTER OF EDUCATION

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

EDMONTON, ALBERTA

FALL, 1977

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled Estimating the Reliability of the Short Forms of the WISC-R submitted by Constance Franklin Kenney in partial fulfilment of the requirements for the degree of Master of Education.

....
Supervisor

....

....

Date June 27, 1977

ABSTRACT

This study was undertaken to fulfill a need for more information concerning the reliabilities of the short forms of the WISC-R when used in a psycho-educational clinic. Before this assessment could be done a preliminary study, requiring the analysis of the standardization data of the WISC-R, was required to determine which short forms were most valid. The resulting 40 best short forms were tested for reliability on the clinic sample.

The concept of intelligence was discussed as well as the reliability and validity of the WISC-R. An outline of some of the research using a variety of the short forms was given. The relative merits and limitations of the "split-half" and "subtest combination" short forms were evaluated. It was concluded that the subtest combination short forms only had merit statistically.

The literature reviewed explained the spuriously high component in correlations involving two sets of scores obtained in just one administration of a test, and the process that was necessary to resolve the problem, namely, a corrected or revised formula for use with part-whole relationships.

Data from the standardization sample which were all available in the manual, were used to obtain a list of 40 best short forms after which their reliabilities with the research sample were analysed. In addition, comparisons between the short forms of the WISC and WISC-R were made and their similarities and differences discussed.

The results of the study were clear and unequivocal in their

support for the use of selected short forms of the WISC-R with individuals who might be referred to a psycho-educational clinic or school counsellor. Some suggestions on how to select the appropriate short forms were made.

The study contributed to the accumulating body of research supporting the reputation of the WISC-R as an even more robust test than its predecessor.

ACKNOWLEDGEMENTS

My sincere gratitude is offered to the following persons:

Dr. Henry Janzen, who gave so freely of his ideas and time, for his interest, encouragement and support.

Dr. Len Stewin, for his kind consideration and constructive criticism.

Professor Jeanette Funke, for her friendship, interest and enthusiasm.

Dr. Steve Hunka and Dr. Tom McGuire, for their expert guidance and assistance in the analysis of my data.

Mrs. Margaret Voice for her exceptional skills and concern in typing this thesis.

And very special thanks to my husband, Ted Kenney, for his continuous support, patience and understanding.

And finally, to my mother, Mollee Franklin, who very early in my life gave me the belief in myself which enabled me to prevail.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION TO THE STUDY

The Problem

Every year thousands of school children and adolescents are administered psychological tests in order to help teachers, parents and clinicians make important decisions regarding the development and progress of each individual child. Such tests are most frequently administered to children who have demonstrated that they are having some difficulty in adjusting to the school situation. Either they are not progressing at the expected rate in the basic school subjects or they are experiencing emotional and/or behavioral problems which appear to be interfering with their learning progress.

Foremost in any selection of tests is an intelligence test. If at all feasible an individually administered test such as the Stanford-Binet or one of the Wechsler Scales is used in preference to a group administered test such as the Lorge-Thorndike or Otis-Lennon. Although more expensive and time-consuming to administer, the individual tests offer a much more dependable measure of the child's intellectual ability. In addition, it offers an opportunity for the examiner to evaluate behavior in the test situation which might relate to performance in the classroom, such as the child's ability to understand what is expected of him, his attitude, his method of attacking the question and in general, any behaviors which might be hindering or interfering with his best performance.

Although the most reliable and valid of all tests, an intelligence test alone is rarely sufficient to diagnose a child's learning difficulties. In addition it is usually necessary to administer other kinds of tests in order to evaluate other aspects of his behavior. For example, examiners are interested in knowing: At what level is this child reading? What are his spelling and arithmetic skills? Does he speak well and get along well with his peers? Does he have perceptual-motor difficulties? Are aural and visual acuity adequate? For this reason, most examiners would agree that when a child is referred for assessment several tests rather than a single intelligence test is the preferred procedure.

While a carefully selected and properly administered battery of tests can offer a wealth of information, in most situations it is more an ideal than a reality. Such programs are found to be prohibitively expensive and very time consuming. The psychologist's or counsellor's dilemma has always been, and continues to be, how to get the most useful information concerning a child while making the best use of the time and resources available to him.

Background to the Problem

In response to the question of how to make the best use of time and resources available the first line of attack has been the development of better tests and a more thorough understanding and utilization of the test results. In the area of intelligence testing psychologists have explored two avenues of research in an effort to obtain the most information in as short a time as possible. One direction has led to the development and use of various abbreviated or modified short forms of the widely used Wechsler Scales (Doppelt, 1956; Silverstein, 1970; Tellegen & Briggs, 1967; Yudin, 1966), while the other direction has focused on a scatter or profile analysis of the individual subtest functions within the full-length test (Frost & Frost, 1962; Rhodes, 1969; Schofield, 1952).

The rationale of the short forms is that because of their high correlation with the full-length test they provide a reliable measure of global intelligence, that is, they are a good measure of an individual's overall capacity for understanding and coping with his world, while meeting an increasing demand for brief psychological tests (Clements, 1965; Patterson, 1953; Silverstein, 1970; Wechsler, 1958). On the other hand, the profile or scatter analysis approach to intelligence testing derives its rationale from ego psychology (Fromm et al, 1957; Rapaport et al, 1968). Variation among subtests is considered a means for assessing such ego functions as memory, concentration, attention, judgment, planning and concept formation (Rapaport et al, 1968). An evaluation of different subtests provides indices for development in specific areas and allows inferences to be made about personality, thought processes, defences, et cetera (Frost & Frost, 1962; Blatt & Allison, 1968).

In summary, the goal in using a short form is to get a quick and reliable estimate of global intelligence whereas the goal in using scatter analysis is to search for as much specific information as possible on each of the separate abilities or subtests of the test in order to gain as broad an understanding as possible from the administration of one high quality test. It almost goes without saying that the latter approach, in addition to requiring more administration time also requires a high degree of experience and interpretive ability on the part of the examiner. The primary reason for using a short form is that, theoretically, it allows more time for other supplemental kinds of testing which are necessary for a complete assessment. This paper is concerned with use of the various short forms of the WISC-R as a sound and sensible method of intelligence testing.

Short Forms

Historically, the Wechsler short forms have been generally of two kinds, namely, the "subtest combination" short form and the "split-half" short form.

In the subtest combination short form only certain selected subtests are administered. From the scores on these subtests, a full scale score and IQ is estimated. For example, the Vocabulary and Block Design subtests together are considered to be a suitable short form. Studies show that scores on these two subtests correlate very highly with the full scale score, although the standard error is somewhat higher (Kaufman, 1975; Silverstein, 1967a, 1970a). Combinations of two, three, four, five and even one single subtest have been considered as suitable short forms because of their high predictive validity. As a result, short forms have been recommended for use in lieu of the full-length test in certain circumstances.

Although a single subtest, such as Vocabulary can serve as an adequate short form (Kaufman, 1975) the standard error is quite high so that while it yields a reasonably accurate estimate of IQ, the confidence one can place on the estimate is less than with a longer test. With all short forms, the standard error of measurement is systematically reduced as the number of subtests in the short form is increased so that the confidence level is greatest for the longest short form. It can always be assumed that the predictive validity will increase, although sometimes only slightly, as the length of the short form is increased with the most accurate prediction being based on the results of the full-length test. These are some of the facts which have to be taken into consideration by the psychologist or clinician when deciding whether or not a short form could be used and if so, which one.

The "split-half" short form gets its name from the statistical procedure which reduces the length of the test by using only half the items. While the length of the test is reduced by half, a sampling from each subtest is retained. For example, by selecting every odd item in the Arithmetic subtest, every even item in the Vocabulary subtest, et cetera, a suitably sampled short form is developed. A few examples of this kind of short form (Yudin, 1967; Silverstein, 1968) are given in Appendix A. Because no subtest is omitted, a measure of all areas of intellectual functioning is obtained. For a short time after it first appeared the split-half procedure was eagerly received by clinicians because it lent itself so well to profile analysis. Later, it came under serious scrutiny for its statistical claims in the area of test reliability and following that its predictive validity. The development of these short forms will be discussed more fully in the review of the literature.

Rationale for Investigation

The thesis of this study is that the subtest combination short forms of the WISC and subsequently the WISC-R are a very practical and highly valid measure of intelligence. While these short forms have been given much attention and consideration in research and statistical studies, they have not been widely adopted for clinical purposes, mostly it would seem because of doubts associated with the ability of the short form to be a viable alternative to the full-length test.

Recent research (Kaufman, 1975; Sattler, 1974; Silverstein, 1967, 1971, 1974) has unhesitatingly supported the high validity of several of the short forms and recommended their use in testing as a general screening device and as a fast, accurate and economical estimate of intellectual ability. On the other hand, it has never been suggested and this study does not suggest, that a short form could replace the full-length test when a comprehensive assessment of intellectual functioning was required, nor would it be recommended for use when making significant classification decisions concerning mental retardation or special education (Finch, Ollendick & Ginn, 1973).

Statement of the Problem

The present study is designed to estimate the reliabilities of a select group of 40 subtest combination short forms of the WISC-R using 182 protocols obtained in a psycho-educational clinic. The purpose will be to determine whether or not any of these short forms are suitable for use in such a specialized clinic.

The best short forms will be selected by computing the validities for all possible subtest combinations of two, three, four and five, using the WISC-R standardization sample data presented in the WISC-R manual (1974) and following the procedure outlined by Silverstein (1970) in his evaluation of the best short forms for the WISC. The selected short forms will be the top ten of each combination. The result will be a list of 40 suitable short forms from which any examiner may select the one most appropriate for his purpose.

The concepts of reliability and validity are used throughout this study to describe characteristics of the short form and its relationship to the full-length test. To clarify the meaning of these terms in the context of the study it should be mentioned that when scores on a short form are correlated with those on the full-length test, the result is considered to be a measure of the validity of the short form, that is, a measure of the degree to which the short form is measuring the same thing as the full-length test. In the studies using the standardization data to determine the 40 best short forms, for example, Silverstein (1970), Sattler (1974), and Part 1 of the present study, the resulting coefficients are a measure of validity.

When the short forms are to be used in lieu of the full-length test, as is the case in Part 2 of this study, what is required is an understanding of the dependability of the short form. How reliable is the short form? Does it consistently give an accurate estimate of IQ? How does the reliability of the short form compare to that of the full-length test? The cross validation analysis is a measure of the reliability of the short forms.

Significance of the Study

The decision to examine the reliability of the short form of the WISC-R was made for several reasons. Because intelligence tests and particularly the WISC-R, are so widely used in schools and clinics, it was felt that the short form could not be ignored by serious users of this test, particularly after recent studies which show very high validities for certain of these forms (Silverstein, 1970; Kaufman, 1975). It is important to know which, if any, of these short forms can b dered a viable alternative for assessments done by Psycho-Educational Clinics.

If the reliability and validity of the short forms for this specialized sample are as high as some studies indicate they might be, then their advantages would be well demonstrated. Not only would these tests be worthwhile because of time saved in administering the tests but the administration skills necessary to administer the short form would be less demanding and therefore more easily acquired which, clearly, would be an important advantage. Teachers themselves might well be able to do some of this testing on their own.

In addition, if short forms are a serious alternative for this particular sample, emphasis could move toward more criterion-related testing in addition to the IQ measure and perhaps more follow-up testing and evaluation of recommendations.

Finally, it was felt that the study provided an opportunity to build a data base from existing files in the Education Clinic and to commence research with clinical data which might lead to other research questions.

CHAPTER II

REVIEW OF THE LITERATURE

Introduction

The Wechsler Intelligence Scale for Children, Revised (WISC-R) is the 1974 revision of the WISC which was first published in 1949 and which has proven by research and reputation to be one of the most rigorous and highly regarded intelligence tests in use today. Changes made in the WISC-R were primarily a rewording of some of the more outdated items, a modernization of subtest stimuli, some clarification for test administrators, and an adjustment in the age range to make it more compatible with the Wechsler test for younger children (WPPSI) and the one for adults (WAIS). However, the WISC-R is substantially the same as the WISC; that is, in all essential areas, the WISC-R is viewed as equivalent to the older test.

As more and more comparative research is being done on the two tests, it is gradually being conceded that in virtually all cases, the information acquired from research on the WISC can be generalized to the WISC-R with, in some cases, a tendency towards slightly more positive results on the WISC-R. This trend has led researchers to conclude that the WISC-R is an even finer test than its predecessor (Wechsler, 1974; Sattler, 1974; Kaufman, 1975; Silverstein, 1974). For this reason, in the following discussion it is quite reasonable to assume that statements and conclusions made concerning the WISC, also apply to the WISC-R.

The Concept of Intelligence as Expressed by the WISC-R

Wechsler (1974) describes intelligence as the overall capacity of a person to understand and cope with his world; he refers to it as a global entity meaning it is something multidetermined and multifaceted. The WISC-R supports this concept of intelligence by using twelve different subtests to measure as many different intellectual abilities with each contributing to the quotient considered to be a measure of general intelligence. No selected abilities are considered to be more crucial or important than the others.

The concept of mental age (MA) whereby a person was said to be functioning at a certain age level of mentality is not used with the WISC-R. Instead of an age-equivalent index, the deviation IQ, which is a measure of an individual's mental functioning in relation to others his own age, is used. The deviation IQ indicates the child's relative position in the age group to which he belongs. Therefore, a child's IQ will not vary from year to year unless his actual performance, compared to his peers, has changed.

Intelligence as measured by the WISC-R does not include all intellectual abilities but it does include those abilities which are considered valuable in our culture. For this reason, primary abilities such as conceptual learning, abstract reasoning and abstract or verbal problem-solving are emphasized in the WISC-R as they are in most intelligence tests (Wechsler, 1958). Intelligence as determined by the WISC-R is related to socially valued criteria; it is a measure of the degree to which an individual has those intellectual qualities which will get him through school and college and into the higher socio-economic occupations (Jensen, 1967). Biases inherent in an intelligence test favor convergent, analytical and scientific thinking while, at the same time almost ignoring the divergent, imaginative and artistic processes; it will always favor the scientists over the artists (Hudson, 1971).

Reliability and Validity of the WISC-R

The reliability coefficients for the WISC-R subtests and Verbal, Performance and Full Scale IQs are reported in the manual for 11 age groups (6 1/2 through to 16 1/2 years by one year intervals) in addition to an averaged index (Wechsler, 1974). The most reliable subtests were Vocabulary (.86), Information (.85), Block Design (.85) and Similarities (.81); the lowest reliabilities were for Object Assembly, Coding, Mazes and Picture Arrangement (.70-.73). Subtest correlations were split-half correlations, except for Coding and Digit Span which were obtained by a test-retest analysis. Reliability for the IQ scales were determined by a formula for a composite group. Reliability coefficients for Verbal IQ ranged from .91 to .96 (average .94), for Performance IQ from .89 to .91 (average .90) and for Full Scale IQ from .95 to .96 (average .96).

Stability coefficients for three age groups, 6 1/2-7 1/2, 10 1/2-11 1/2 and 14 1/2-15 1/2 (N = 303) were reported (Wechsler, 1974). They were determined by retesting the group after a one-month interval and are as follows: average coefficients for Verbal subtests ranged from .77 (Digit Span) to .88 (Information); average coefficients for Performance subtests ranged from .65 (Mazes) to .81 (Block Design); average coefficient for Verbal IQ was .93, for Performance IQ .90 and for Full Scale IQ .95.

Congruent validity for the WISC-R has been determined by correlating WISC-R scores with those on the WAIS and WPPSI for the age groups which overlap (age 6 1/2 and 16 1/2). Correlations with the WAIS for the three IQ Scales were Verbal .96, Performance .83 and Full Scale .95 with the WAIS yielding somewhat higher IQs than the WISC-R. Correlations with the WPPSI IQ Scales were Verbal .80, Performance .80 and Full Scale .82. The mean WPPSI IQs were slightly higher than the WISC-R IQs.

Research with Short Forms of the WISC and WISC-R

Most of the research with the WISC short forms has been in the form of validation study whereby the full-length scale was administered to a group of subjects after which certain subtest scores were combined and their total scores correlated to the full scale IQ. The correlation is a measure of the degree to which the short form is a satisfactory estimate of intellectual functioning. All the studies cited have obtained their data from one administration of the full-length scale.

The short form definitely has some limitations and Caldwell and Smith (1968) have suggested that possibly the short forms of the WISC may not be suitable for use with Negro children because of the fact that the intercorrelations of the subtests were found to be lower among Negroes than the standardization group. Hutt and Gibby (1965) stated that the use of the short form should never be considered in any evaluation of a retarded child. On the other hand, the short form was recommended for use with handicapped children (Nickols & Nickols, 1963). With a careful selection of a suitable short form they were able to get reliable results and ensure a positive experience for the child.

Using his abbreviated procedure for the WISC, Yudin (1966) obtained correlations with a group of emotionally disturbed children. Correlations for subtests ranged from .76 to .94 while those for Verbal, Performance and Full Scale scores ranged from .93 to .97. Reid, Moore and Alexander (1968) used Yudin's procedure with a group of brain-damaged and mentally retarded children and they reported the same satisfactory results. In a later study Gayton, Wilson and Bernstein (1970) did not get such unequivocal results using the Yudin procedure. Subjects were children in an outpatient psychiatric clinic. Correlations with the full scales were satisfactory for individual subtests but not for the Verbal-Performance discrepancy factor which they considered an important part of their evaluation.

From the results of a study with 145 emotionally disturbed children, Enburg, Rowley and Stone (1961) highly recommended the use of 30 WISC short forms based on correlations with the full scale IQ of .92 for three-subtest combinations, .94 for four-subtest combinations and .95 for five-subtest combinations. They concluded that the WISC subtest combination short forms were a suitable and reliable estimate of intellectual functioning for children with emotional problems but discouraged their use with problems of organic involvement and/or mental retardation.

In another study with disturbed children, Nickols and Nickols (1963) reported that the WISC short form consisting of Information, Arithmetic, Digit Span and Picture Completion (or Similarities if there was a visual handicap) had a good correlation (.92) with the full scale when used with disturbed children ranging in age from 6 to 16, and in IQ from 51 to 144. There was a slight tendency for the short form to over-estimate the IQ at the higher levels. Similar results had previously been reported for the short forms of the WAIS when used with an adult schizophrenic population (Nickols, 1962).

Clements (1965) examined 92 children between the ages of 9-0 and 12-11 who had been referred to a clinic for possible reading disabilities. Using a four-subtest combination short form of Similarities, Object Assembly, Arithmetic and Picture Arrangement, correlations of .947 with the full scale were obtained.

In general, the results of the studies using short forms of the WISC and WISC-R suggested that the abbreviated tests provided good estimates of the full scale IQ in many circumstances and with various specialized groups. However, the limitations of the research on the short form made it clear that identifying the kind of decision that has to be made was of primary importance (Levy, 1968).

"Subtest Combination" Short Form

McNemar (1950) set the pattern for studies dealing with the short forms of the WAIS and WISC. He was interested in determining the validities of the short form which consisted of various subtest combinations. He set the example by basing his own research on the original standardization sample of the Wechsler Bellevue Scale and by devising a formula for measuring the degree of correlation between the short form and the full-length test which did not require access to raw data. The formula which utilized data obtained directly from subtest intercorrelation tables presented in the Wechsler manual (1944) was as follows:

$$r_{xy} = \frac{k + \Sigma\Sigma r_{hj}}{\sqrt{k + 2\Sigma r_{gh}}\ \sqrt{n + 2\Sigma r_{ij}}}$$

Where r_{xy} = correlation between the subtest short form and the full-length scale (called the part-whole correlation or the validity coefficient for the designated short form).

k = number of subtests in the short form.

n = total number of subtests (10 for the WISC-R).

$\Sigma\Sigma r_{hj}$ = the sum of the correlations between each of the k subtests and all the other subtests.

Σr_{gh} = the sum of the intercorrelations of the k subtests.

Σr_{ij} = the sum of the correlations between each of the n subtests.

Note: $\sqrt{n + 2\Sigma r_{ij}}$ becomes a constant in all computations.

The results of McNemar's (1950) research was the systematic ranking of the ten best combinations of two, of three, of four and of five subtests from the Wechsler test. The correlation coefficients of the various short forms with the full-length test tended to be somewhat lower than those which had been published in earlier studies (Hunt, 1948; Patterson, 1948; Rabin, 1943) which was accounted for by the fact that the standardization sample was a much more homogeneous group compared to the samples used in other studies.

"Split-half" Short Form

Another kind of short form to gain acceptance by some researchers (Mogel & Satz, 1963; Pauker, 1963; Satz & Mogel, 1962; Yudin, 1966) was the "split-half" abbreviation of the WAIS and WISC, so called because only half of each subtest was administered. This short form was considered to be a superior form for clinical purposes because it provided a complete breadth of intellectual functioning by sampling each subtest. It was thought to have all the advantages of the full-length test and all the conveniences of a short form. By comparison, the subtest combination short forms were considered to be of limited use because of their emphasis on a single IQ index (Mogel & Satz, 1963).

The split-half procedure was first introduced by Wolfson and Bachelis (1960) who shortened the WAIS Verbal Scale by administering the odd items only on all of the verbal subtests except for the Digit Span subtest which was left unchanged. Correlations between the split-half scores on the Verbal Scale and the full Verbal Scale were reported to be .97.

The split-half form was further developed to include the Performance Scale as well as the Verbal Scale with very encouraging results. The number of items was reduced by as much as 54 percent compared to the full length test and part-whole correlations were reported to be .99 for Verbal IQ, .97 for Performance IQ and .99 for Full Scale IQ (Satz & Mogel, 1962). Following this success a procedure for accurately converting scores on the short form to IQ equivalents was established and the preparation of tables for the

WAIS and WISC (Satz & Mogel, 1962; Yudin, 1966; Silverstein, 1968a) made the use of the split-half short form both sound and practical.

Yudin (1966) proposed a split-half variation of the WISC which consisted of administering every third item on Information, Vocabulary and Picture Completion (scores were later multiplied by 3), every even item on Arithmetic and every odd item on Comprehension, Similarities, Picture Arrangement, Object Assembly and Block Design (scores multiplied by 2), and all items on Digit Span and Coding. Scores were corrected as indicated in the brackets after which the manual could be used to compute scaled scores and IQ's in the usual way. Correlations for this form with the full-length test were Verbal IQ, .96, Performance IQ, .95 and Full Scale IQ, .96. An outline of this short form is presented in Appendix A.

Yudin's (1966) abbreviated procedure for the WISC (and more currently the WISC-R) gained some acceptance particularly for use with specialized groups for whom a short form was more suitable. In a study with brain damaged and mentally retarded children the procedure was found to be very satisfactory (Reid, Moore & Alexander, 1968). In another study (Satz, Van de Riet & Mogel, 1967), the results led the authors to conclude that the Yudin procedure was not suitable for children with above normal IQs.

Although popular in some areas, the split-half procedure shortly began to receive some serious criticism for its statistical claims. A review of the research on the split-half abbreviations (Zytowski & Hudson, 1965) brought attention to the fact that the part-whole correlations in the studies they reviewed were almost all higher than the odd-even reliability estimates given in the WAIS and WISC manuals. They questioned how a short version of a test could be correlated more highly with the full-length test than the full-length test was correlated with itself. The validity coefficient between part and whole scales could not in reality be higher than the internal consistency coefficient reported by Wechsler. The spuriously high part-whole correlations claimed by the split-half procedure were attributed to a contamination of the results when an abbreviated scale was embedded in the full-length scale. For this reason the use of a profile interpretation obtained from the split-half procedure was strongly discouraged (Zytowski & Hudson, 1965).

Other researchers became highly critical of the split-half short form in the area of test reliability pointing out that short forms which reduced the length of a subtest instead of reducing the number of subtests were ignoring the depreciating effects on reliability (Tellegen & Briggs, 1967). Since reliability is a function of test length, when the number of items in a test is reduced the reliability of the test is decreased. Therefore, it was concluded that the short forms which used fewer items while retaining each subtest (Satz & Mogel, 1962; Yudin, 1966) were ill-suited for the very purpose for which they were designed, namely profile analysis (Tellegen & Briggs, 1967; Zytowski & Hudson, 1965). The same studies also underlined the fact that because the differences between subtest scores in general tend to be much less reliable than the tests themselves it was quite possible that while the total scores might be sufficiently reliable, the profile data would be decidedly unreliable.

It was finally concluded that the split-half abbreviated tests could be used, like the subtest combination tests, as a dependable measure of an IQ index only.

Correlations between the Short Form and the Full-Length Test

The contamination finding which is reflected in the high correlations claimed for both split-half and subtest combination short forms with the criterion test is dealt with by several researchers (Bashaw & Anderson, 1967; Levy, 1967; Silverstein, 1968; Tellegen & Briggs, 1967; Zytowski & Hudson, 1965). The importance of the part-whole correlation cannot be over-rated since it specifies the degree of equivalence between the short form and the full-length test (Tellegen & Briggs, 1967) and is a measure of the validity of the abbreviated test (McNemar, 1950).

When the part-score is obtained from the same test administration as the full-length test score, which is generally the case for most studies (one rare exception is Zytowski and Hudson, 1965), the part-whole correlation cannot be obtained in the traditional manner (Tellegen & Briggs, 1967). It was generally acknowledged that the dependence of the errors in the two sets of data greatly increases the correlation between them but for the wrong reason—incorporated in the correlation coefficient is the perfect correlation between the errors in each test giving it a spuriously inflated value (Levy, 1967). The next logical step in the research process was to define the way in which the part-whole correlations were spuriously high and to devise a formula for determining the validity of a short form which would be more realistic (Bashaw & Anderson, 1967; Levy, 1967; Tellegen & Briggs, 1967).

As an alternative to the part-whole correlation some researchers suggested and supported the suitability of the part-remainder correlation as a more meaningful validity measure for the short form (Zytowski & Hudson, 1965). The rationale for this consideration was based on the method for split-half correlations. When the part score, that is, the abbreviated form was correlated with the remainder of the test instead of with the whole test, the part-whole error variance was eliminated. The result was a lower and therefore more realistic measure of the part-whole relationship (Zytowski & Hudson, 1965). Other researchers disagreed; while this procedure eliminated error overlap, it also eliminated true-score overlap. For this reason the part-remainder correlation was considered an inappropriate index since it was not by definition a measure of the part-whole relationship (Tellegen & Briggs, 1967).

This evaluation was elaborated further by studies which suggested the probability that the part-whole correlation outlined by McNemar (1950) and the part-remainder correlation represented respectively the upper and lower boundaries of the "true coefficient" (Levy, 1967). Only in cases where the reliability of the part score was very high would the true coefficient approach the part-whole correlation coefficient and only when the reliability was very low would the true coefficient approach the part-remainder correlation coefficient. The true coefficient would lie somewhere between these two coefficients.

Modified Formula for Short-Form Validity

After further studies a formula was developed for obtaining a more accurate measure of the part-whole relationship (Levy, 1967). This index which corrected for the covariance due to correlated error, that is, it excluded part-score error variance, was as follows:

$$r_{xy}' = r_{xy} - (1 - r_{xx})\frac{\sigma_x}{\sigma_y}$$

Where r_{xy}' = corrected part-whole correlation coefficient.

r_{xy} = part-whole correlation coefficient (between short form and full-length test).

r_{xx} = reliability coefficient of each subtest.

$\frac{\sigma_x}{\sigma_y}$ = ratio of error variance to part-score variance.

Note: $(1 - r_{xx})\frac{\sigma_x}{\sigma_y}$ = error variance of the part-score.

This same formula which appeared to solve the problem of spuriously high part-whole correlations was arrived at independently by Bashaw and Anderson (1967) at about the same time.

Tellegen and Briggs (1967) without reference to Bashaw and Anderson (1967) or Levy (1967) approached the problem from a slightly different point of view by suggesting a modified index which replaced the perfect correlation between the part and itself which was assumed in McNemar's (1950) formula, by its reliability coefficient. The result was the following formula:

$$r_{xy}' = \frac{\Sigma r_{hh} + \Sigma\Sigma r_{hj}}{\sqrt{k + 2\Sigma r_{gh}}\ \sqrt{n + 2\Sigma r_{ij}}}$$

Where r_{xy}' = modified coefficient of correlation between composite part and composite whole. (Modified part-whole correlation.)

Σr_{hh} = sum of reliabilities of the subtests comprising the part.

$\Sigma\Sigma r_{hj}$ = sum of all the intercorrelations between each of the k subtests and all the other subtests.

Σr_{gh} = sum of correlations between any subtests g and h in the part.

Σr_{ij} = sum of correlations between any subtests l and m belonging to the whole.)

k = number of subtests included in the part.

n = number of subtests included in the whole.

Note: $n + 2\Sigma r_{ij}$ = becomes a constant in all calculations.

Although it is not immediately apparent to the non-statistician, the relationship of the above formula to that put forward by Bashaw and Anderson (1967) and Levy (1967) is that they are algebraically equivalent (Silverstein, 1971).

In conclusion to their study, Tellegen and Briggs (1967) hypothesized that if the modified index was used to reanalyse the standardization data, the results would yield coefficients lower than those reported by Doppelt (1956) for the WISC subtest short forms and would have the effect of altering the rank order and selection of the best short forms of each combination as determined in several earlier studies using the uncorrected part-whole formula (Clements, 1965; Enburg, Rowley & Stone, 1961; Howard, 1959; Jones, 1962; Maxwell, 1957; McNemar, 1950).

Following this hypothesis suggested by Tellegen and Briggs

(1967) and using the modified formula, the WISC standardization data, obtained from the Wechsler manual, were reanalysed by Silverstein (1970). The results of this study appear in Appendix B. Correlations with the full scale test of the ten best short forms for each combination of two, three, four and five subtests are given in descending order. The values given by the corrected formula averaged .032 lower than those obtained by the old formula (McNemar, 1950). In addition, it was found that 46 of the 120 best short forms selected differed from those selected by the McNemar formula (Silverstein, 1970). This supported the predictions made by Tellegen and Briggs (1967). The new reappraised values were considered to be a more realistic measure of the validity of the subtest short form (Tellegen & Briggs, 1967; Levy, 1967; Bashaw & Anderson, 1967) and the new list of best combinations determined by these validities was deemed to be the most accurate (Silverstein, 1970).

However there was a dissenting note to Silverstein's (1970) summation from McNemar (1974) who maintained that since the whole purpose was to select those short forms which correlated highest with the full scale test, then the original formula (McNemar, 1950) was the most suitable because it accomplished exactly that. Sattler (1974) in the revised edition to his book, appears to have been persuaded by this evaluation to some extent when he chose to revert back to the old uncorrected formula for analysing the standardization data for the new WISC-R. On the other hand, for the results on the WISC and WPPSI, he retained those obtained by Silverstein (1967a, 1967b, 1968, 1970) using the newer corrected formula.

The present writer felt that the argument made for a modified index was reasonable and well documented. The corrected formula should have been used for the new WISC-R data, if only for one important reason, that it makes it possible to compare the best short forms of the WISC and WISC-R. Although some indices obtained may differ only slightly from the old indices, the fact that a better theoretical method has been developed was thought to be a good reason to move on. For this reason, this study has included a reevaluation of the WISC-R data using the corrected formula which provided an updated and more reliable list of validity coefficients for the proposed WISC-R short forms.

In summary, there was much evidence to support the use of a corrected formula for computing part-whole correlations when both scores are obtained in a single administration. The results obtained with the two formulae were considered to be sufficiently different to warrant a reappraisal of the WISC-R standardization data using the corrected formula.

CHAPTER III

RESEARCH DESIGN AND PROCEDURE

The Problem

It was the purpose of this paper to analyse the standardization data of the WISC-R following the method described by Silverstein (1970). Following that analysis and using the resulting list of best Short Forms, a multiple correlational study was carried out, using clinical data obtained from 182 WISC-R protocols, to determine the reliability of the best subtest short forms for use with this particular population.

The Subjects

The subjects were 182 elementary and junior high students, aged 6 1/2-16 1/2, who were enrolled in schools in the Edmonton area and in some smaller rural communities outside the city. All the subjects were tested between September 1975 and July 1976 as part of a University testing program whose purpose was to offer its services to the community while training new psychologists. The protocols were administered by graduate students and examined by clinic supervisors for errors and ommissions. Only those subjects with scores on all ten subtests were included. This study was based on the assumption that the administration and scoring of the protocols was accurate and that the estimates for full scale IQ are reliable and valid.

A characteristic of the group was that all subjects were referred for testing invariably because of some difficulty with their school work or coping with the school situation. Because all subjects were experiencing a learning problem or coping with a social or emotional problem it was expected that norms for the research group would be lower than the norms for the standardization group.

The Procedure: Part 1

An important preliminary part of this study was to re-analyse the standardization data of the WISC-R using the revised formula (Tellegen & Briggs, 1967) in order to determine the validity coefficients for all possible short forms. It was felt that only then could the short forms of the WISC-R be compared in any meaningful way to those of the WISC which had been obtained statistically by this same method (Silverstein, 1970). The WISC short forms are listed in Appendix B. Also, the results of this analysis yielded the list of forty best short forms which were tested on the research data to obtain a measure of their reliability in the clinic sample.

The formula and method used for the foregoing analysis has been fully elaborated and discussed earlier in this paper. Briefly, the first part of this study was a replication of the analysis of the standardization data of the WISC carried out by Silverstein (1970) using instead the WISC-R data. All necessary data for use with this formula were available in the standardized tables published in the Wechsler Manual (1974). Using this procedure all possible combinations of two, three, four and five subtests were correlated with the full scale IQ and then rank ordered to obtain the ten best short forms in each group. (The total number of possible combinations of the ten subtests is 627; there are 45 different short forms when selecting any two subtests, 120 when selecting any three subtests, 210 when selecting any four subtests and 252 when selecting any five subtests (McNemar, 1950).)

Two subtests, Digit Span and Mazes were not considered for inclusion in any possible short forms. Both these subtests were excluded in the computation of IQs for the standardization data and for this reason were considered inappropriate for inclusion in a short form which attempts to predict that IQ.

The Procedure: Part 2

The second part of the study was a systematic analysis of the reliability of the 40 short forms to determine the extent to which these short forms were suitable for use with the specialized research group.

Data Analysis

Means, standard deviations and partial correlations were obtained. A step-wise multiple-regression procedure with double cross validation (Mulro 8) was used with the clinic sample to obtain predictive estimates for use with combinations of WISC-R subtests. Subjects were randomly assigned to one of two groups (A and B) for each short form. By this procedure a regression equation was developed for group A and then applied to group B to calculate validity (cross validation) after which the procedure was reversed giving a double cross validation analysis. Groups were randomly re-assigned for each short form analysed.

CHAPTER IV

RESULTS

The Research Sample

A total of 182 protocols of the WISC-R were used in the study. The subjects were 126 male and 56 female students ranging in age from 6 to 16 years. A summary of the percentage of males and females in each age group is presented in Table 1.

TABLE 1

CLINIC SAMPLE BY AGE AND SEX: PERCENT IN EACH AGE GROUP

	Percent in Each Age Group		
Age Group	Male	Female	Total
6-0 to 6-11	7.1	3.8	11.0
7-0 to 7-11	5.5	2.2	7.7
8-0 to 8-11	4.4	2.2	6.6
9-0 to 9-11	7.1	2.2	9.3
10-0 to 10-11	3.8	1.6	5.5
11-0 to 11-11	1.1	1.1	2.2
12-0 to 12-11	2.7	3.8	6.6
13-0 to 13-11	18.1	6.0	24.2
14-0 to 14-11	10.4	3.3	13.7
15-0 to 15-11	7.1	3.3	10.5
16-0 to 16-11	1.6	1.1	2.7
Total	69.2	30.8	100.0

Approximately 70 percent of the sample were boys and 30 percent were girls. This proportion for the clinic group is a reflection of the general finding that many more boys than girls have learning difficulties.

The largest represented age group was 13-0 to 13-11 which accounted for 24 percent of the group. This is the age at which it becomes increasingly difficult for a child to compensate for learning disabilities and/or deficits. Children aged 13-0 to 15-11 accounted for 68 percent of the sample with over two-thirds of those boys. The next highest represented age group was 6-0 to 6-11 accounting for 11 percent of the sample.

A breakdown of the percentage of subjects in each IQ classification compared to the standardized group is summarized in Table 2.

TABLE 2

CLINIC SAMPLE BY IQ CLASSIFICATION COMPARED TO STANDARDIZATION SAMPLE: PERCENT IN EACH CLASSIFICATION

	Standardization Group	Research Group
Mentally Deficient (-69)	2.2	7.1
Borderline (70-79)	6.0	15.9
Dull Normal (80-89)	16.5	24.7
Average (90-109)	49.4	45.0
Bright Normal (110-119)	16.5	5.4
Superior (120-129)	7.4	1.6
Very Superior (130+)	2.3	0.0
Total	100.3	99.7

Because of the specialized nature of the research group IQs tended to be generally lower than those in the standardization group, that is, there were a larger number of subjects with below average IQ and a smaller number with average or above average IQ in the research group. This was an expected result.

The means and standard deviations for the three IQ indices were all lower for the research group than the standardization sample. These results are shown in Table 3. Compared to the standardized mean of 100 with a SD of 15, the mean Verbal IQ was 87.86 (SD 15.85), Performance IQ was 92.44 (SD 15.77) and Full Scale IQ was 89.07 (SD 15.67).

TABLE 3

MEANS AND STANDARD DEVIATIONS FOR VERBAL IQ, PERFORMANCE IQ AND FULL SCALE IQ FOR THE CLINIC SAMPLE (N = 182)

	Mean	SD
Verbal IQ	87.86	15.85
Performance IQ	92.44	15.77
Full Scale IQ	89.07	15.67

Note: Mean and SD for the Standardized distributions for all three IQ indices are 100.0 and 15.0.

Scaled scores for every individual subtest were lower than the standardized mean of 10 while standard deviations were usually somewhat higher than the standardized 3. Similarly Sums of Scaled Scores for Verbal, Performance and Full Scale were lower, 40.38 (SD 12.92) compared to 50.25 (SD 12.14), 44.64 (SD 11.80) compared to 50.19 (SD 10.89) and 85.02 (SD 22.76) compared to 100.44 (SD 21.01). These results are summarized in Table 4.

Results for Part 1: 40 Best Short Forms

The standardization data for the WISC-R were analyzed according to the method suggested by Silverstein (1970) to determine

TABLE 4

WISC-R MEANS AND STANDARD DEVIATIONS OF SCALED SCORES FOR THE CLINIC SAMPLE (N = 182)

	Mean		SD	
Information	7.75	(10.00)*	3.03	(3.00)*
Similarities	8.26		3.18	
Arithmetic	7.87		3.01	
Vocabulary	8.08		3.10	
Comprehension	8.42		3.00	
Sum of Scaled Scores—Verbal	40.38	(50.25)	12.92	(12.14)
Picture Completion	9.31	(10.00)	3.05	(3.00)
Picture Arrangement	9.36		3.58	
Block Design	8.50		3.20	
Object Assembly	9.56		2.86	
Coding	7.88		3.17	
Sum of Scaled Scores—Performance	44.64	(50.19)	11.80	(10.89)
Sum of Scaled Scores—Full Scale	85.02	(100.44)	22.76	(21.01)

Note: *Figures in parentheses are the mean scaled scores of the subtests and sum of verbal, performance and full scale scores of the standardization sample.

which short forms correlated best with the full scale. The 10 best short forms in each combination of two, three, four and five subtests were determined and are listed with their validity coefficient in Table 5.

The 40 best short forms which were recommended by Silverstein (1970) for the WISC are marked with an asterisk; there are 22 short forms which appear on both lists. An informative comparison can now be made with Table 5 by referring to the complete list of 40 best short forms for the WISC in Appendix B.

More than half the WISC-R short forms are the same as those selected for the WISC which is an expected finding considering the high degree of similarity between the two tests and the continued high subtest reliabilities of the subtests Vocabulary, Information, Similarities and Block Design which appear so frequently in the short forms. The degree to which the two sets vary can be explained by the improved reliability of some of the subtests of the WISC-R, in particular, Comprehension and Picture Completion. Comprehension, for example, had a reliability range of .59 to .73 on the WISC compared to .69 to .87 on the WISC-R; Picture Completion had a reliability range from .59 to .69 on the WISC compared to .68 to .85 on the WISC-R. As a result these two subtests appeared with more frequency in the WISC-R short forms: Comprehension appeared 10 times compared to 5 on the WISC; Picture Completion appeared 11 times compared to twice on the WISC. These subtests replaced in frequency other subtests which appeared in fewer of the WISC-R short forms: Arithmetic appeared in only 4 compared to 15 short forms of the WISC; Picture Arrangement,

TABLE 5

VALIDITY COEFFICIENTS FOR THE 40 BEST WISC-R SHORT FORMS: A REPLICATION OF THE PROCEDURE RECOMMENDED BY SILVERSTEIN (1970)

Dyad		Triad		Tetrad		Pentad	
Short Form	r	Short Form	r	Short Form	r	Short Form	r
*V BD	.882	*I V BD	.903	I V PC BD	.916	I S V PC BD	.926
*I BD	.862	*S V BD	.902	*I S V BD	.914	I V C PC BD	.925
S BD	.856	*I S BD	.896	*S V PA BD	.912	*I S V PA BD	.925
C BD	.846	*I C BD	.896	I V C BD	.911	*S A V PA BD	.924
*V OA	.839	*V C BD	.894	I S C BD	.911	*I S V OA BD	.923
*S V	.838	V PC BD	.887	*I V PA BD	.911	I S C PC BD	.922
V PC	.836	*S C BD	.887	S V PC BD	.910	I V PC PA BD	.922
*I V	.834	*A V BD	.886	*S A V BD	.909	I S C PA BD	.922
I S	.832	V PA BD	.885	I C PC BD	.909	I V PC CO BD	.922
I PC	.825	*S V OA	.881	I V OA BD	.908	*S A V OA BD	.922

* Short Forms which also appear on the WISC list of best 40 (Silverstein, 1970).

Note 1: Abbreviations: I = Information; C = Comprehension; A = Arithmetic; S = Similarities; V = Vocabulary; PC = Picture Completion; PA = Picture Arrangement; BD = Block Design; OA = Object Assembly; CO = Coding.

Note 2: SE_{est} for Dyad range from 7.08-7.59; Triad 6.44-7.09; Tetrad 6.01-6.29; Pentad 5.66-5.81.

in 7 compared to 14; Object Assembly, in 5 compared to 9 short forms of the WISC. The Coding subtest appeared in only one short form for each test.

The subtests appearing most frequently in the short forms of the WISC-R were those with the highest reliability: Block Design (.85), Vocabulary (.86), Information (.85) and Similarities (.81); respectively they were represented in 33, 28, 22 and 19 of the 40 short forms. All except Vocabulary appeared more frequently (3 to 5 times) in the WISC-R tests.

A further comparison was made between the results in Table 5 and those obtained by the McNemar (1950) method, which was a comparison of the results obtained by two different procedures. The latter were reported in Sattler (1974) and are presented in Appendix C. While varying noticeably in validity coefficient values, the two lists were very similar in selection of short forms. Of the 40 short forms, 29 are the same for both lists; of those which changed, 6 were pentads, 4 were tetrads and 1 was a triad; all dyads remained unchanged.

The two lists of short forms are similar to the degree that the results are largely dependent on the reliabilities of the individual subtests. Both methods used the reliability coefficients reported in the manual (Wechsler, 1974). The extent of their variability can be accounted for by the different procedures used. In general, the corrected formula yielded lower validity coefficients for the short forms, an outcome which was expected and preferred as being a more realistic value (Tellegen & Briggs, 1967; Silverstein, 1970).

The standard errors of estimate (SE_{est}) for the WISC-R were calculated as 7.08 IQ points for the best dyad, 6.44 for the best triad, 6.01 for the best tetrad and 5.66 for the best pentad. It should be noted that these values are somewhat higher than those reported by the other method (refer to Appendix C). Because SE_{est} is a function of the validity coefficients their value increased when the coefficients were decreased. As a result, in addition to a more realistic validity coefficient, the correct formula yields what is probably a more realistic standard error of measurement.

The list of correlation coefficients presented in Table 5 is intended to replace the list shown in Sattler (1974) which was considered inappropriate for use in comparing the WISC and WISC-R. By referring to Table 5 and Appendix B which lists the 40 best WISC short forms, it was observed that indeed the WISC-R short forms, largely because of improved subtest reliability as mentioned earlier, were a more valid measure of IQ than those of the WISC. The validity coefficients in general, tended to be moderately, but consistently, higher compared to the WISC. For example, for the best dyad (V BD) the validity coefficient was .882 compared to .856, for the best triad (I V BD) it was .903 compared to .887, for the best tetrad (I V PC BD) it was .916 compared to .904 and for the best pentad (I S V PC BD) it was .926 compared to .915. Validities were improved to the extent that for each group the new range of validity coefficients was equivalent to the range of the next largest group on the WISC, for example, the validity for the dyads was equal to that of the triads on the old test. This finding supports the general conclusions concerning the WISC-R, that it is an even more rigorous test than its predecessor

(Sattler, 1974; Wechsler, 1974).

Results for Part 2: Cross Validation Analysis

The data from 182 protocols were used to evaluate the reliability of the 40 best WISC-R Short Forms. This analysis was obtained by a multiple correlation procedure with a double cross validation procedure. For each short form analysis subjects were randomly reassigned to one of two groups and regression equations calculated for each group. The results were expressed by the squared multiple correlation coefficient (R^2) which gave a measure of the proportion of the variability in the Full Scale IQ which could be accounted for by the regression equation of each particular short form. The results of this analysis are summarized in Table 6.

Three R^2 calculations were made for each short form: The first (R^2) was a measure of the relationship between the short form and the full-length test for the research sample; the second (R'^2) was an unbiased estimate of R^2 for the population calculated according to Olkin and Pratt (1958) and Tatsuoka (1969). The third was the cross-validated R^2 whereby scores calculated on Group 1 from the regression equations for Group 2, and on Group 2 from regression equations for Group 1, were correlated with the Full Scale IQ. This final statistic was the strongest one that could be made when both predictor and criterion variables were obtained at one time. The results in general were very strong and appear to overwhelmingly support the efficacy of many of the WISC-R short forms. Theoretically, R^2 would be greater than R'^2 or the cross-validated R^2; R'^2 would be smaller than R^2 because it was an estimate of the population; cross-validated R^2 would

TABLE 6

CROSS VALIDATED MULTIPLE CORRELATION COEFFICIENTS (R^2) FOR 40 WISC-R SHORT FORMS OBTAINED FROM RESEARCH SAMPLE
(N = 182)

	Group 1			Group 2		
Short Form	R^2	R'^2	R^{2*}	R^2	R'^2	R^{2*}
V BD	.85	.85	.85	.80	.80	.80
I BD	.85	.85	.84	.82	.82	.81
S BD	.75	.75	.73	.75	.75	.71
C BD	.80	.80	.78	.74	.74	.72
V OA	.81	.81	.81	.81	.81	.80
S V	.75	.75	.75	.75	.75	.75
V PC	.80	.80	.80	.81	.81	.81
I V	.78	.78	.78	.78	.78	.78
I S	.79	.79	.79	.76	.76	.76
I PC	.82	.82	.82	.77	.77	.77
I V BD	.89	.85	.88	.88	.83	.87
S V BD	.86	.83	.85	.86	.83	.85
I S BD	.89	.86	.89	.84	.80	.84
I C BD	.88	.82	.88	.87	.84	.87
V C BD	.84	.80	.83	.86	.86	.86
V PC BD	.90	.86	.89	.82	.78	.81
S C BD	.87	.80	.87	.76	.67	.75
A V BD	.91	.86	.91	.82	.78	.82
V PA BD	.82	.76	.82	.93	.89	.92
S V OA	.76	.71	.76	.90	.87	.90
I V PC BD	.90	.90	.89	.92	.92	.91
I S V BD	.90	.90	.87	.90	.90	.88
S V PA BD	.88	.88	.88	.93	.93	.93
I V C BD	.89	.89	.89	.89	.89	.89
I S C BD	.89	.89	.88	.89	.89	.89
I V PA BD	.92	.92	.91	.92	.92	.89
S V PC BD	.91	.91	.90	.88	.88	.87
S A V BD	.92	.92	.91	.88	.88	.86
I C PC BD	.93	.93	.92	.89	.89	.88
I V BD OA	.92	.92	.91	.92	.92	.91
I S V PC BD	.93	.90	.92	.92	.92	.92
I V C PC BD	.92	.90	.92	.93	.92	.93
I S V PA BD	.93	.92	.92	.94	.93	.93
S A V PA BD	.94	.92	.93	.95	.93	.94
I S V BD OA	.92	.91	.90	.94	.93	.93
I S C PC BD	.94	.91	.94	.92	.90	.91
I V PC PA BD	.93	.92	.92	.94	.92	.93
I S C PA BD	.93	.91	.93	.93	.92	.93
I V PC BD CO	.93	.91	.93	.94	.92	.94
S A V BD OA	.92	.91	.92	.93	.91	.93

Note:

R^2 = the squared multiple correlation coefficient of the subtest with the Full Scale IQ.

R'^2 = an unbiased estimate of the squared multiple correlation coefficient for the population. It is calculated according to Olkin and Pratt (1958) and Tatsuoka (1969) by a formula which corrects for shrinkage.

R^{2*} = the cross validated R^2, calculated for Group 1 from the regression equation constructed for Group 2, and for Group 2 from the regression equation constructed for Group 1. It is a measure of predictive efficiency.

be expected to be lower than either R^2 or R'^2. In fact, cross validated R^2 was equal to and often exceeded the other coefficients which is to say the actual data yielded higher coefficients than would be theoretically expected.

It should be mentioned again that the values presented in Table 6 are the squared multiple correlations and express the variance which can be accounted for by each short form in predicting full scale IQ. With this consideration in mind some noteworthy observations can be made. Dyads have a wide variability; at their worst they accounted for only 71 to 73 percent of the variance in predicting IQ, at their best they accounted for 81 to 85 percent. The triads showed much the same kind of variability as the dyads but at their best they could make a prediction which accounted for 91 to 92 percent of the variance. When the short forms were increased from three to four subtests, the variability was reduced considerably. The tetrads were able to predict full scale IQ which accounted for 86 to 93 percent consistently. The five-subtest short form compared favorably with the full scale test by accounting for 90 to 94 percent of the variance when predicting IQ.

In general, the reliability of the short forms increased as the length of the short form increased. It was observed that the range of correlation coefficients narrowed as the short form lengthened indicating less variability and better reliability for the longer test. For example, when the results for Group 2 were examined, it was found that multiple R^2 correlations for the two-subtest short forms ranged from .71 to .81, for the three-subtest short forms from .75 to .92, for the four-subtest short forms from .86 to .93 and for the five-subtest short forms from .91 to .94.

The ability of the short form to predict the full scale IQ successfully was increased as the number of subtests in the short form was increased. However, while the differences in reliability between the dyad and triad, and between the triad and tetrad were quite noticeable, the differences between the tetrad and pentad were not as great. For example, the best multiple R^2 correlations for a dyad were .85 and .81 (highest for each group) compared to .91 and .92 for a triad, .92 and .93 for a tetrad and .94 and .94 for a pentad. It was concluded that the four-subtest short forms were the most elegant short forms giving a good reliable estimate of IQ for the very reasonable amount of time equired.

The evaluation of the efficacy of the short forms in predicting Verbal IQ and Performance IQ has been summarized in Table 7.

In general, the short forms were better predictors of Verbal IQ (V IQ) than Performance IQ (P IQ). For example, the multiple R^2 correlation coefficients for the dyads ranged from .70 to .91 for V IQ compared to .43 to .70 for P IQ; for the triads, from .82 to .91 for V IQ compared to .58 to .81 for P IQ; for the tetrads, .88 to .96 for V IQ compared to .68 to .85 for P IQ and in the pentads, .90 to .95 for V IQ compared to .73 to .91 for P IQ. The variability for each group of short forms was greater when predicting P IQ. For example, in the most reliable group, the pentads, the multiple R^2 ranged from .73 to .91 for P IQ compared to .90 to .95 for V IQ.

The short forms also were able to predict Verbal IQ somewhat better than Full Scale IQ (FS IQ) (see Table 6; compare Table 5). For

TABLE 7

CROSS VALIDATED MULTIPLE CORRELATION COEFFICIENT (R^2) FOR VERBAL IQ AND PERFORMANCE IQ FOR CLINIC SAMPLE
(N = 182)

	Verbal IQ		Performance IQ	
Short Form	Group 2 per Group 1 Equation	Group 1 per Group 2 Equation	Group 2 per Group 1 Equation	Group 1 per Group 2 Equation
1	.81	.85	.66	.73
2	.77	.80	.69	.73
3	.70	.72	.69	.64
4	.79	.74	.67	.66
5	.87	.78	.55	.70
6	.87	.92	.32	.45
7	.79	.84	.66	.70
8	.87	.91	.43	.43
9	.90	.85	.42	.43
10	.84	.70	.69	.70
11	.88	.92	.71	.65
12	.88	.91	.71	.63
13	.89	.86	.63	.79
14	.90	.91	.63	.79
15	.89	.89	.75	.60
16	.85	.83	.87	.72
17	.86	.88	.76	.58
18	.90	.89	.71	.68
19	.86	.82	.86	.81
20	.89	.91	.68	.61
21	.91	.89	.80	.82
22	.93	.96	.73	.69
23	.88	.91	.81	.85
24	.94	.95	.68	.76
25	.96	.94	.63	.76
26	.90	.90	.81	.82
27	.91	.88	.81	.79
28	.95	.95	.74	.68
29	.91	.88	.85	.78
30	.90	.90	.78	.85
31	.94	.94	.83	.77
32	.92	.94	.82	.79
33	.94	.95	.82	.86
34	.96	.93	.82	.86
35	.95	.94	.82	.82
36	.95	.94	.80	.81
37	.89	.91	.89	.91
38	.93	.95	.79	.86
39	.90	.90	.88	.88
40	.96	.94	.83	.73

example, the range of correlations for the dyads was .71 to .81 for the FS IQ compared to .70 to .91 for V IQ; for triads, .75 to .92 for FS IQ compared to .82 to .91 for V IQ; for tetrads, .86 to .93 for FS IQ compared to .88 to .96 for V IQ; for pentads, .91 to .94 for FS IQ compared to .90 to .95 for V IQ.

For this study multiple regression weights and constants were determined for each short form. These are given in Table 8. Since the results of Group 1 and 2 were so similar throughout the analysis, it was decided that only the data from Group 1 would be presented. The regression equation for computing the estimated criterion variable (Y') from the predictor variables (b_1, b_2, . . . b_5) is shown at the bottom of the table.

The multiple regression weights and constants shown in the table are those derived from the sample data and used in the double cross validation analysis. Theoretically the derived regression equations could be used to predict full scale IQs in another sample, particularly in a sample which was similar to the research group. The limitations of the predictions made on this basis are the same as those for any regression equation and that is that the weights and constants are derived from a specific sample and not from the population. However, many of the short forms as determined by this study, have a sufficiently high correlation (see Table 6) to warrant a high degree of confidence in their ability to predict the full scale IQ.

TABLE 8

MULTIPLE REGRESSION EQUATION WITH ADDITIVE CONSTANT (a) AND REGRESSION WEIGHTS FOR USE IN PREDICTING FULL SCALE IQ

	a	b 1	b 2	b 3	b 4	b5
V BD	44.76	3.08	2.22			
I BD	44.75	2.97	2.49			
S BD	48.26	2.07	2.78			
C BD	42.58	2.44	3.02			
V OA	39.04	3.77	1.98			
S V	47.57	1.95	3.10			
V PC	42.23	3.48	2.05			
I V	49.54	2.68	2.25			
I S	48.24	3.16	1.89			
I PC	41.15	3.56	2.14			
I V BD	42.89	2.14	1.60	1.93		
S V BD	42.89	1.29	2.25	2.02		
I S BD	43.13	2.49	1.47	1.72		
I C BD	41.89	2.43	1.48	1.91		
V C BD	42.20	2.26	1.36	1.99		
V PC BD	39.50	3.01	1.37	1.45		
S C BD	38.18	1.47	2.35	2.28		
A V BD	41.53	1.73	2.50	1.57		
V PA BD	42.36	2.53	1.21	1.80		
S V OA	41.16	1.25	2.55	1.76		
I V PC BD	42.60	2.16	1.29	.95	1.26	
I S V BD	42.63	2.25	.92	.89	1.69	
S V PA BD	39.51	1.10	2.13	1.23	1.41	
I V C BD	4[illegible].29	1.88	1.28	.84	1.68	
I S C BD	42.67	2.26	.93	.96	1.61	
I V PA BD	39.51	1.55	1.44	.96	1.98	
S V PC BD	37.72	1.12	1.96	1.30	1.60	
S A V BD	39.85	1.18	1.54	1.62	1.60	
I C PC BD	34.46	2.14	1.37	1.20	1.79	
I V BD OA	37.85	1.87	1.56	1.45	1.24	
I S V PC BD	38.72	1.25	.99	1.53	1.08	1.25
I V C PC BD	37.36	1.34	1.15	1.16	1.28	1.30
I S V PA BD	41.56	1.05	.82	1.36	1.23	1.17
S A V PA BD	40.24	.90	1.09	1.57	1.30	.95
I S V BD OA	35.90	1.55	.72	1.49	1.14	1.40
I S C PC BD	39.40	1.64	.85	1.11	1.03	1.35
I V PC PA BD	41.37	1.54	1.56	.74	.55	1.39
I S C PA BD	37.88	1.64	.95	1.11	1.06	1.42
I V PC BD CO	37.01	1.53	1.51	1.04	1.13	1.08
S A V BD OA	37.65	.86	1.01	1.82	1.10	1.24

Note: $Y' = a + b_1X_1 + b_2X_2 \ldots b_5X_5 + e_i$

CHAPTER V

SUMMARY, DISCUSSION, IMPLICATIONS AND CONCLUSIONS

The results reported in Chapter IV give encouraging support to the thesis of this study which was that the WISC-R short forms are a sound and sensible alternative to the full-length test in many situations and can be considered by the serious and busy test administrator as part of a battery or as a single quick measure of intelligence.

The study explored the spuriously high component in correlations involving two sets of scores obtained in just one administration of a test and the process that was necessary to resolve the problem satisfactorily. The new corrected formula for obtaining part-whole correlations was suggested by Tellegen and Briggs (1967) and considered by Silverstein (1970) to be a sound solution.

The replication of Silverstein's (1970) procedure on the WISC-R standardization data produced a list of 40 best short forms (Table 5) which could be compared and evaluated in a meaningful way to the list of 40 best short forms for the WISC (Appendix B). A comparison of the two lists indicated important similarities and differences.

It was observed that 22 of the short forms appeared on both lists, a finding which was expected since the two full-length tests were so similar. The changes in the remaining 18 short forms of the WISC-R were accounted for by the increased appearance of two subtests, Comprehension and Picture Completion which, due to improved subtest reliability, tended to preempt the subtests Arithmetic, Picture Arrangement and Object Assembly which appeared more frequently on the old test.

The validity coefficients of the WISC-R short forms were found to be consistently higher than those for the WISC leading to the conclusion that the revised test was more robust than the WISC as suggested by Wechsler (1974), Sattler (1974) and Kaufman (1975).

The results of the cross validation analysis shown in Tables 6-8 indicated that the short forms, in general, were a very reliable prediction of the full scale IQ and could be used with confidence to estimate the intellectual level of a child.

It was noted that the short forms were better predictors of Verbal IQ than Performance IQ, in fact, they were better predictors of Verbal IQ than Full Scale IQ. This finding is explained by noting that verbal subtests, or those correlating highly with the verbal scale, were heavily represented in the short forms. Vocabulary, Information and Similarities were each represented in more than half the short forms while Block Design, which correlates as well with the verbal scale as it does with the performance scale, was represented in 33 of the 40 short forms. The high reliability of these subtests, along with their frequency in the short forms, ensures that the short forms will be excellent measures of Verbal IQ.

The squared multiple correlations for all short forms were higher than theoretically expected. It was found that the two-subtest short form could make a prediction of IQ which accounted for 71 to 85 percent of the variance. As the length of the short form was increased the ability to predict the full scale IQ was improved, so that any three-subtest short form could predict an IQ within a 75 to 92 percent range of variance, a four-subtest short form within an 86 to 93 percent range of variance and any five-subtest short form within the 90 to 94 percent range of variance. These results gave strong support for their use in groups such as the research sample where children were experiencing some school difficulties.

The tetrads or four-subtest short forms were considered to be the most efficient and most useful because of their high predictive ability in relation to the amount of administration time required.

From the list of 40 best short forms of the WISC-R an examiner could select those which suited his needs and preferences. It was suggested that some care in selecting an appropriate short form would ensure the probability of both a good reliable result and a positive experience for the child.

Short forms were not recommended for use with retarded children or where classification decisions were to be made, such as opportunity class placement. Short forms however, should be very useful with children who have been referred to a clinic or counsellor for emotional disturbance or behavioral problems. They were also recommended for use with handicapped children who might find a longer test tiring, and for those who had a possible reading disability (Clements, 1965).

Because some children find it difficult to keep still for very long or for some reason cannot spend a long time on one task, any short form would likely be a better choice over the full-length test.

Individual subtest characteristics and functions were important considerations in selecting a short form. Vocabulary, for example, is a long and demanding task for which many children lack the skill, in which case a short form with Information or Similarities which are shorter tests and seem easier, would be the better choice. Block Design was included in 33 of the 40 short forms and luckily is a test which most children enjoy; errors in this test are often not apparent to the child, so there is no discouragement. For a blind child the only short forms would be those without any of the Performance tests which leaves only three dyads to choose from, Similarities—Vocabulary, Information—Vocabulary and Information—Similarities. For a child who is deaf but able to read, there would be no problem in choosing short forms; if the child could not read, the short forms would not be appropriate. With a behavior problem child, that is, one who was restless or inattentive or not easily engaged, a full-length test would be difficult; a short form which is quick and interesting would be more successful, for example, Information—Comprehension—Picture Completion—Block Design.

Limitations of the Study

The major limitation of this study was that again the issue of the reliability of short forms was decided on data obtained from a single administration of the full-length test. Although the statistics used were the most rigorous available for making corrections under these conditions, the question must be explored further in a study which would use short form data which had been obtained from a separate administration.

Future research could address itself to this problem by testing a group with a short form and some time later with the full-length test. Since both tests would not have to be administered by the same person, the study could be made in cooperation with the Individual Testing psychology course at the University. Short forms could be administered to those who would be or had been tested by students in the program.

Implications of the Study

The purpose of this study was to, firstly, analyse the standardization data of the WISC-R in order to determine the 40 best short forms, and secondly, to determine the reliabilities of these short forms with a clinic sample. The results of the study were positive and unequivocal. In terms of this study the implications are as follows:

1. The list of 40 best short forms of the WISC-R which was determined by this study (Table 5) is intended to be used by all those who do testing to select appropriate short forms according to needs and preferences. In addition, the list is intended to replace a similar list published in Sattler (1974).
2. This study implies that short forms of the WISC-R can be used with confidence as reliable and valid measures of intellectual level in a variety of situations. The short forms work very well for the purpose for which they were intended, that is, a good estimate of intellectual functioning.

3. This study adds to the body of research which supports the use of the WISC-R short forms in general practice.
4. Similarly, it contributes to the research which supports the reputation of the WISC-R as an even more robust test than its predecessor.

Conclusion

The list of suggested short forms derived from this study and tested on the research group appear to have excellent utility for such a specialized group. Whereas previous studies in general supported the use of the short form, there was some question as to whether or not they would be suitable for use with children referred to a psycho-educational clinic.

An evaluation of the data indicated that some short forms had better utility than others. If the criterion was only to save time then any of the two-subtest short forms was clearly the best; if high reliability was the criterion then the five-subtest short forms were the best. However, the three-subtest and four-subtest short forms while only slightly more time consuming were appreciably more reliable than the two-subtest short form and for all practical purposes equally as reliable as the five-subtest short form. For this reason they were considered the most preferred and strongly recommended. Adding a fifth subtest did not increase reliability sufficiently to warrant the required extra administration time, although it could be easily included when circumstances demanded it.

Within the set of triads and tetrads other criteria could certainly be used by the administrator in selecting an appropriate test. The fact that certain subtests are longer or less interesting for the child while others are quicker and more fun, and the fact that some require special test materials while others are entirely verbal would be determining factors. It was felt that a short form could be selected on the basis of the individual, the environment and the tester according to suitability and preference.

When using the WISC-R Short Forms some caution must be exercised in the interpretation of results. The standard error for the short form is definitely higher than the full-length test. The standard error of the estimate (SE_{est}) for the WISC-R short forms varies from 7.08 IQ points for the best dyad to 5.66 for the best pentads (refer to note in Table 5). The average standard error of measurement (SE_m) for the full scale IQ is 3.19, for Verbal IQ 3.60. Therefore any individual IQ can be over- or under-estimated to a large degree. No single administration of a short form can be interpreted as anything more than a rough estimate. However, this fact was not considered to be a great disadvantage in as much as it was thought that it should reduce the tendency of those interested in labeling a child with a number and thinking of that number as invariable. The explicit assumption of a short form result is that it is an imperfect estimate of general intelligence indicating roughly a child's intellectual development as compared to other children his age.

The WISC-R short forms yield a quick and very reliable estimate of an individual's intellectual level. They will also detect any gross impairments which would indicate the necessity for administering the full-length test or to continue with further testing. They are a very economical way of obtaining a highly valid measure of intelligence.

There is much evidence to support the use of the short forms with any public school population. However, they are not recommended for diagnosing learning disabilities or for making educational decisions about mental retardates or special education candidates.

The advantages of the short form are that they are practical, save time, serve as good screening devices, allow for more frequent testing and can be administered by less than fully qualified psychologists. A classroom teacher, for example, can use a WISC-R Short Form to test or support a hypothesis concerning a child, almost immediately without too much formality and have the information very soon afterwards.

The various WISC-R short forms are to some degree less valid than the full-length form but not so much so that they cannot be seriously considered for use in many testing situations. Every user of the short form then must consider the nature of the decision to be made and evaluate the relationship between validity lost and time saved.

BIBLIOGRAPHY

BIBLIOGRAPHY

Bashaw, W. L. & Anderson, H. E. Jr. A correction for replicated error in correlation coefficients. *Psychometrika*, 1967, *32*, 435-441.

Blatt, S. J. & Allison, J. The intelligence test in personality assessment. In A. I. Rabin (Ed.), *Projective techniques in personality assessment*. New York: Springer, 1968, 421-460.

Caldwell, M. B. & Smith, T. A. Intellectual structure of southern Negro children. *Psychological Reports*, 1968, *23*, 63-71.

Clements, C. R. An abbreviated form of the Wechsler Intelligence Scale for Children. *Journal of Consulting Psychology*, 1965, *29*, 92.

Doppelt, J. E. Estimating the full scale score on the WAIS from scores on four subtests. *Journal of Consulting Psychology*, 1956 *20*, 63-66.

Draper, M. R. & Smith, H. *Applied regression analysis*. New York: John Wiley & Sons Inc., 1966.

Enburg, R., Rawley, V. N. & Stone, B. Short forms of the WISC for use with emotionally disturbed children. *Journal of Clinical Psychology*, 1961, *17*, 144-149.

Finch, A. J., Ollendick, T. H. & Ginn, F. W. WTSC short forms with mentally retarded children. *American Journal of Mental Deficiency*, 1973, *78*(2), 144-149.

Fromm, E., Harman, L. D. & Marschak, M. Children's intelligence tests as a measure of dynamic personality functioning. *American Journal of Orthopsychiatry*, 1957, *27*, 134-144.

Frost, B. P. & Frost, R. The pattern of WISC scores in a group of juvenile psychopaths. *Journal of Clinical Psychology*, 1962, *18*, 354-355.

Gayton, W. F., Wilson, W. T. & Bernstein, S. An evaluation of an abbreviated form of the WISC. *Journal of Clinical Psychology*, 1970, *26*, 466-468.

Glasser, A. J. & Zimmerman, I. L. *Clinical interpretation of the Wechsler Intelligence Scale for Children*. New York: Grune & Stratton, 1967.

Hunt, W. A. The validity of some abbreviated intelligence scales. *Journal of Consulting Psychology*, 1948, *12*, 48-52.

Hutson, B. A. Psychological testing; misdiagnosis and half-diagnosis. *Psychology in the Schools*, 1974, *11*, 388-391.

Hutt, M. L. & Gibby, R. G. *The mentally retarded child* (2nd ed.). Boston: Allyn and Bacon, 1965.

Jensen, A. R. How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, 1969, *39*, 1-123.

Joesting, J. & Joesting, R. Quick test validation scores of adults in a welfare setting. *Psychological Reports*, 1972, *30*, 537-538.

Kaufman, A. S. Factor analysis of the WISC-R at 11 age levels between 6 1/2 and 16 1/2 years. *Journal of Consulting and Clinical Psychology*, 1975, *43*(2), 135-147.

Kelly, F. J., Beggs, D. & McNeil, K. A. *Research design in the behavioral sciences: Multiple regression approach.* Southern Illinois University Press, 1969.

Levy, P. The correction for spurious correlation in the evaluation of short-form tests. *Journal of Clinical Psychology*, 1967, *23*, 84-86.

Levy, P. Short-form tests: A methodological review. *Psychological Bulletin*, 1968, *69*, 410-416.

McNemar, Q. On abbreviated Wechsler-Bellevue Scales. *Journal of Consulting Psychology*, 1950, *14*, 79-81.

McNemar, Q. *Psychological statistics* (3d ed.). New York: Wiley, 1962.

McNemar, Q. Correction to a correction. *Journal of Consulting and Clinical Psychology*, 1974, *42*, 145-146.

Maxwell, E. Validities of abbreviated WAIS scales. *Journal of Consulting Psychology*, 1957, *21*, 121-126.

Mogel, S. & Satz, P. Abbreviation of the WAIS for clinical use: An attempt at validation. *Journal of Clinical Psychology*, 1948, *12*, 147-152.

Nickols, J. Brief forms of the Wechsler Intelligence Scales for research. *Journal of Clinical Psychology*, 1962, *18*, 167.

Nickols, J. & Nickols, M. Brief forms of the WISC for research. *Journal of Clinical Psychology*, 1963, *19*, 425.

Olkin, I. & Pratt, J. W. Unbiased estimate of certain correlation coefficients. *Annals of Mathematics Statistics*, 1958, *29*, 201-211.

Patterson, C. H. A further study of two short forms of the Wechsler-Bellevue Scale. *Journal of Consulting Psychology*, 1948, *12*, 147-152.

Patterson, C. H. *The Wechsler-Bellvue scales: A guide for counsellors.* Springfield, Illinois: Charles C. Thomas, 1953.

Pauker, J. D. A split-half abbreviation of the WAIS. *Journal of Clinical Psychology*, 1963, *19*, 98-100.

Rapaport, D., Gill, M. M. & Schafer, R. *Diagnostic psychological testing* (Rev. ed.). New York: International Universities Press, 1968.

Reid, W. B., Moore, D. & Alexander, D. Abbreviated form of the WISC for use with brain-damaged and mentally retarded children. *Journal of Consulting and Clinical Psychology*, 1968, *32*, 236.

Rhodes, F. *Manual for the Rhodes WISC scatter profile.* San Diego, California: Educational and Industrial Testing Service, 1969.

Sattler, J. M. *Assessment of children's intelligence* (Rev. ed.). Philadephia: W. B. Saunders, 1974.

Satz, P. & Mogel, S. An abbreviation of the WAIS for clinical use. *Journal of Clinical Psychology*, 1962, *18*, 77-79.

Satz, P., Van de Riet, H. & Mogel, S. An abbreviation of the WISC for clinical use. *Journal of Consulting Psychology*, 1967, *31*, 108.

Schofield, W. Critique of scatter and profile analysis of psychometric data. *Journal of Clinical Psychology*, 1952, *8*, 16-22.

Silverstein, A. B. "Validity" of short-form intelligence tests. *Journal of Consulting Psychology*, 1965, *29*, 392-393.

Silverstein, A. B. A short form of the WISC and WAIS for screening purposes. *Psychological Reports*, 1967, *20*, 682. (a)

Silverstein, A. B. A short form of the Wechsler scales for screening purposes. *Psychological Reports*, 1967, *21*, 842. (b)

Silverstein, A. B. Validity of a new approach to the design of WAIS, WISC, WPPSI short forms. *Journal of Consulting and Clinical Psychology*, 1968, *32*, 478-479.

Silverstein, A. B. Reappraisal of the validity of the WAIS, WISC and WPPSI short forms. *Journal of Consulting and Clinical Psychology*, 1970, *34*, 12-14. (a)

Silverstein, A. B. Reappraisal of the validity of a short form of Wechsler's scales. *Psychological Reports*, 1970, *26*, 559-561. (b)

Silverstein, A. B. A corrected formula for assessing the validity of WAIS, WISC and WPPSI short forms. *Journal of Clinical Psychology*, 1971, *27*, 212-213.

Silverstein, A. B. A short short form of the WISC-R for screening purposes. *Psychological Reports*, 1974, *35*, 817-818.

Silverstein, A. B. Comment: A reply of McNemar. *Journal of Consulting and Clinical Psychology*, 1975, *43*, 423-425.

Tatsuoka, M. M. *Selected Studies in Advanced Statistics. Number 5.* Champaign, Illinois: The Institute for Personality and Ability Testing, 1969.

Tellegen, A. & Briggs, P. F. Old wine in new skins: Grouping Wechsler subtests into new scales. *Journal of Consulting Psychology*, 1967, *31*, 499-506.

Tipton, R. M. & Stroud, L. H. Abbreviated forms of the WAIS. *American Journal of Mental Deficiency*, 1973, *78*(2), 150-152.

Wechsler, D. *The Measurement of Adult Intelligence* (3d ed.). Baltimore: Williams & Wilkins, 1944.

Wechsler, D. *Manual for the Wechsler Intelligence Scale for Children*. New York: Psychological Corporation, 1949.

Wechsler, D. *Manual for the Wechsler Adult Intelligence Scale*. New York: Psychological Corporation, 1955.

Wechsler, D. *The Measurement of Adult Intelligence* (4th ed.). Baltimore: Williams & Wilkins, 1958.

Wechsler, D. *Manual for the Wechsler Intelligence Scale for Children* (rev. ed.). New York: Psychological Corporation, 1974.

Yudin, L. W. An abbreviated form of the WISC for use with emotionally disturbed children. *Journal of Consulting Psychology*, 1966, *30*, 272-275.

Zytowski, D. G. & Hudson, J. The validity of split-half abbreviations of the WAIS. *Journal of Clinical Psychology*, 1965, *21*, 292-294.

APPENDICES

APPENDIX A

ABBREVIATED PROCEDURES OF THE WECHSLER SCALES USING THE "SPLIT-HALF" METHOD

APPENDIX A

ABBREVIATED PROCEDURES OF THE WECHSLER SCALES USING THE "SPLIT-HALF" METHOD

Appendix A contains three examples of the "split-half" method:

(a) An Abbreviated Procedure for the WAIS (Satz & Mogel, 1962)

(b) An Abbreviated Procedure for the WISC (Yudin, 1966)

(c) A Procedure for Reducing the Number of Items in Subtests of the WISC (Silverstein, 1968)

AN ABBREVIATED PROCEDURE FOR THE WAIS
(SATZ & MOGEL, 1962)

Subtest	Items Used	Multiply Score By	Correction
Information	Every 3rd	3	Subtract 1 from scaled score
Comprehension	Odd only	2	
Arithmetic	Odd only	2	
Similarities	Odd only	2	
Digit Span	Unchanged	1	
Vocabulary	Every 3rd	3	
Digit Symbol	Unchanged	1	
Picture Completion	Every 3rd	3	
Block Design	Odd only	2	Subtract 1 from scaled score
Picture Arrangement	Odd only	2	
Object Assembly	Odd only	2	

AN ABBREVIATED PROCEDURE FOR THE WISC (YUDIN, 1966)

Subtest	Recommended Items	Multiply Score By	Correction Factors
Information	Every 3rd	3	Subtract 1 from scaled score
Comprehension	Odd only	2	None
Arithmetic	Even only	2	Add 1 to scaled score
Similarities	Odd only	2	None
Vocabulary	Every 3rd	3	Subtract 4 from scaled score
Digit Span	All items	1	None
Picture Completion	Every 3rd	3	Subtract 2 from raw score
Picture Arrangement	Odd only	2	Subtract 3 from raw score
Block Design	Odd only	2	Subtract 6 from raw score
Object Assembly	Odd only	2	None
Coding	All items	1	None

A PROCEDURE FOR REDUCING THE NUMBER OF ITEMS IN SUBTESTS OF THE WISC (SILVERSTEIN, 1968)

Subtest	Recommended Items	Multiply Score By
Information	Every 3rd	3
Comprehension	Odd only	2
Arithmetic	Even only	2
Similarities	Odd only	2
Vocabulary	Every 3rd	3
Digit Span	Not Administered	-
Picture Completion	Every 3rd	3
Picture Arrangement	Odd only	2
Block Design	Odd only	2
Object Assembly	Odd only	2
Coding	All items	1

APPENDIX B

VALIDITY COEFFICIENTS FOR THE 40 BEST WISC SHORT FORMS
(SILVERSTEIN, 1970)

APPENDIX B

VALIDITY COEFFICIENTS FOR THE 40 BEST WISC SHORT FORMS (SILVERSTEIN, 1970)

Dyad		Triad		Tetrad		Pentad	
Short Form	r	Short Form	r	Short Form	r	Short Form	r
V BD	.856	A V BD	.887	I V PA BD	.904	I A V PA BD	.915
I BD	.836	I V BD	.886	A V PA BD	.903	A S V PA BD	.915
I V	.825	S V BD	.885	A S V BD	.900	I S V PA BD	.913
A V	.822	I V OA	.873	S V PA BD	.900	C A V PA BD	.912
V PA	.822	V PA BD	.873	C A V BD	.897	A S V BD OA	.908
S V	.819	C V BD	.869	I A V BD	.897	I A V BD OA	.908
S BD	.817	I V PA	.867	I S V BD	.896	I S V BD OA	.908
I PA	.816	I C BD	.867	S V BD CO	.892	A S V PC BD	.908
I OA	.811	I S BD	.867	I V PA OA	.892	I C V PA BD	.908
V OA	.811	A V OA	.866	A V PC BD	.891	I A V PA OA	.908

Note 1: Abbreviations: I = Information; C = Comprehension; A = Arithmetic; S = Similarities; V = Vocabulary; PC = Picture Completion; PA = Picture Arrangement; BD = Block Design; OA = Object Assembly; CO = Coding.

Note 2: The formula used to compute these part-whole correlations was suggested by Silverstein (1970). The purpose of this modified formula is to correct for the spuriously high coefficients previously found in part-whole correlations.

APPENDIX C

VALIDITY COEFFICIENTS FOR THE 40 BEST WISC-R SHORT FORMS (SATTLER, 1974)

APPENDIX C

VALIDITY COEFFICIENTS FOR THE 40 BEST WISC-R SHORT FORMS (SATTLER, 1974)

Dyad		Triad		Tetrad		Pentad	
Short Form	r	Short Form	r	Short Form	r	Short Form	r
V BD	.906	S V BD	.931	I V C BD	.947	S A V PA OA	.963
I BD	.888	I V BD	.929	S V PA BD	.947	S A V PA BD	.962
S BD	.885	I C BD	.928	I C PC BD	.945	S A V BD OA	.960
C BD	.878	I S BD	.925	S A V OA	.944	I C PC BD CO	.960
V OA	.878	V C BD	.924	I V PA BD	.944	I V PC BD CO	.960
V PC	.868	S C BD	.921	I S C BD	.944	S A C PA OA	.960
S V	.864	S V OA	.919	I C PA BD	.944	I S C PA BD	.960
I S	.860	V PA BD	.919	I S PA BD	.943	I V C PC BD	.959
I PC	.858	A V OA	.919	S V PC BD	.943	A V C BD OA	.959
I V	.857	V PC BD	.919	S V BD OA	.943	A V C PA BD	.958

Note 1: Abbreviations: I = Information; S = Similarities; A = Arithmetic; V = Vocabulary; C = Comprehension; PC = Picture Completion; PA = Picture Arrangement; BD = Block Design; OA = Object Assembly; CO = Coding.

Note 2: The formula used to compute these part-whole correlations was suggested by McNemar (1974).

Fig. 30. Binary gradient picture of Edmonton I.

A two dimensional cross correlation matrix was computed by shifting the windows horizontally and vertically for a displacement of 8 units. The maximum correlation value in the matrix gave the location of best matching integer coordinates. The correlation procedure was repeated for getting fractional alignments of one half pixel.

For control points from either uniform regions or regions where changes occurred in the second picture, correlation values were less than 0.5. Thus 3 control points were dropped. Both measures of calculating the figure of merit were used. The second approach of connected edge elements is more reliable, because it takes care of noisy regions and isolated edge pixels. The 6 control point areas are outlined in white squares in Fig's 28 and 29. The central pixel within the square represents the control point chosen.

In the second test case, two pictures of the fertilizer region were chosen. The pictures are 600*500 and are shown in Fig's 31 (Fertilizer 2, red band) and 32 (Fertilizer 3, red band). Both pictures do not have well distributed feature points, and have more than minor variations in rotation and translation. The binary gradient picture of the Fertilizer 2 picture was obtained with threshold set to 5. The binary picture was divided into sub-pictures of maximum size 256*256, resulting in 2 sub-pictures of 256*256, 2 sub-pictures of 256*244, 1 sub-picture of 88*256 and a sub-picture of 88*244.

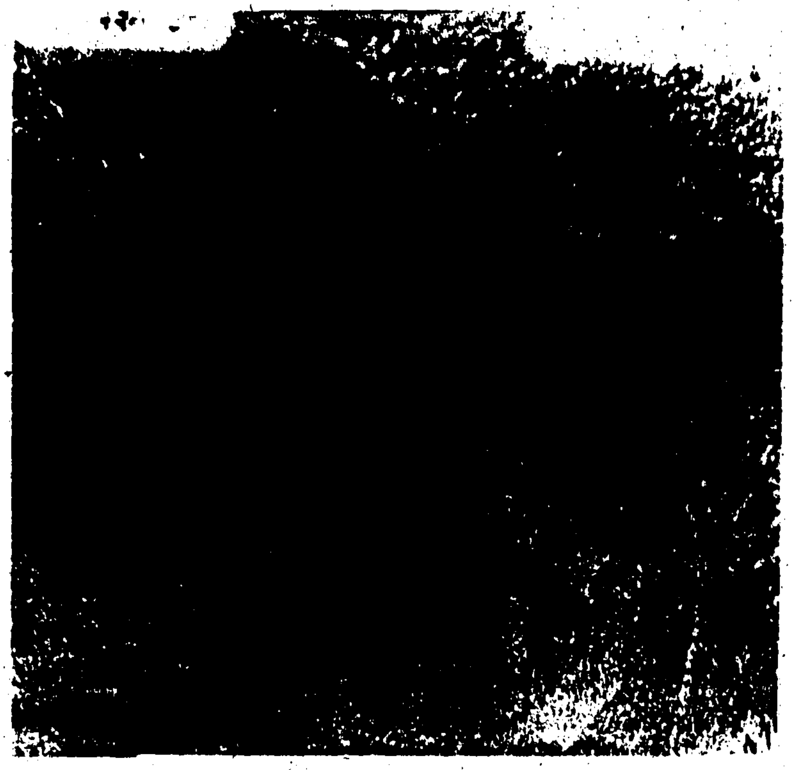

Fig. 31. Fertilizer 2 with control points.

Fig. 32. Fertilizer 3 with control points.

The connected edge element technique was used to compute the lists of potential control points. A set of 9 final control points were chosen from the control points obtained over the whole picture.

The procedure followed in selecting the 9 control points and in checking their validity is same as in the first test case. The correct control points obtained are shown in Fig's 31 and 32. It is worth noting that except for one control point, all the rest were chosen at the edges of the objects of the picture. This particular control point was chosen to satisfy the criterion that at least one control point should come from the region in which no control point was selected. This problem will arise for uniform regions, since they will produce weak edges and consequently the connected edge element measure will have a low value.

3.4 Iterative Feedback Approach

As mentioned in section 3.2, if the control points are not selected properly, mis-registration occurs. For the improperly selected control points, the error was found to be greater than 5 pixels. In order to handle this problem, an iterative feedback technique for registration was developed.

The proposed approach is a combination of the control point and correlation approach. As in the control point technique, some points are chosen and coefficients of a transformation F are calculated. Next, a small window corresponding to a control point is extracted and its corresponding transformed window (larger in size) is also obtained. These two windows are then cross correlated for a shift of ±2 units in both directions. The small shift range is useful for convergence in the next iterative loop, as the shift values are used in feedback correction. The two dimensional displacement vector is computed for all the control points, and computation of the correlation matrix is repeated for obtaining fractional alignments in terms of one half pixel. Next, a feedback correction of the displacements is done for all the control points of the original picture.

The whole process of computing the coefficients of the transformation, extracting windows and computing the two dimensional displacement vector is repeated until either convergence is attained or a maximum number of iterations has been reached. Convergence is defined to imply that the displacement vector is zero for all control points. For a non-zero displacement vector, the best set of control points is chosen from the minimum displacement vector. The procedure tries to adjust the control points of the original picture to give the best possible coordinates. The transformation is then generated using the minimum displacement vector. As a final check, the original picture and the transformed picture are cross correlated at both specific and random test points. The proposed method is given in algorithm D.

<u>Algorithm</u> D:

<u>Step</u> 1: Set max=1, maxiter: maximum number of iterations, p: shift parameter. Read control point pairs from the two pictures.

<u>Step</u> 2: Compute the coefficients of the transformation by least squares analysis for the control points.

<u>Step</u> 3: Extract a small window S of size w*w around one control point from the original picture.

<u>Step</u> 4: Calculate the larger transformed window WT of size w+p*w+p, corresponding to S, from the picture to be transformed.

<u>Step</u> 5: Cross correlate window S for a shift of p units with the transformed window WT, giving a two dimensional matrix, from which the relative displacement with respect to the control point of the original picture is calculated.

<u>Step</u> 6: Repeat steps 3 to 4 for all control points.

<u>Step</u> 7: max = max + 1, if max ≥ maxiter, go to step 8. If the relative displacement vector is zero in both directions, go to step 8. Add the relative displacement vector to the control points of the original picture in both the directions (x,y), and save the vector. Repeat the procedure beginning with step 2.

<u>Step</u> 8: Choose the control point pair, according to the minimum relative displacement vector obtained from

several iterations.

Step 9: Repeat step 2 with new control point pairs and transform the whole picture.

To illustrate the iteration procedure, suppose the control point coordinates of the original and second picture are (40,10) and (100,25) respectively. After extracting the window and transformed window and computing the cross correlation, let the relative displacement vector be 1 unit in the x direction and -2 units in the y direction. After applying the feedback correction, the new control point pair becomes (41,8) and (100,25).

It should be pointed out that the transformation function generates row and column coordinates as non-integer numbers. In the nearest neighbor approach, the numbers are rounded, i.e., 0.5 is added and integer portions are taken as coordinates. These fractions can also be used as weights in the two dimensional interpolation. If w1 and w2 represent fractions of coordinates generated, and a(x,y) is the gray level of the pixel at row x, column y, then the gray level of the output pixel is given by:

$$(w [illegible] y) + (1-w1)*a(x+1,y) + w2*a(x,y+1) + (1-w2)*a(x+1,y+1))/2.$$

In order to test the performance of the algorithm, four different sets of pictures were chosen for

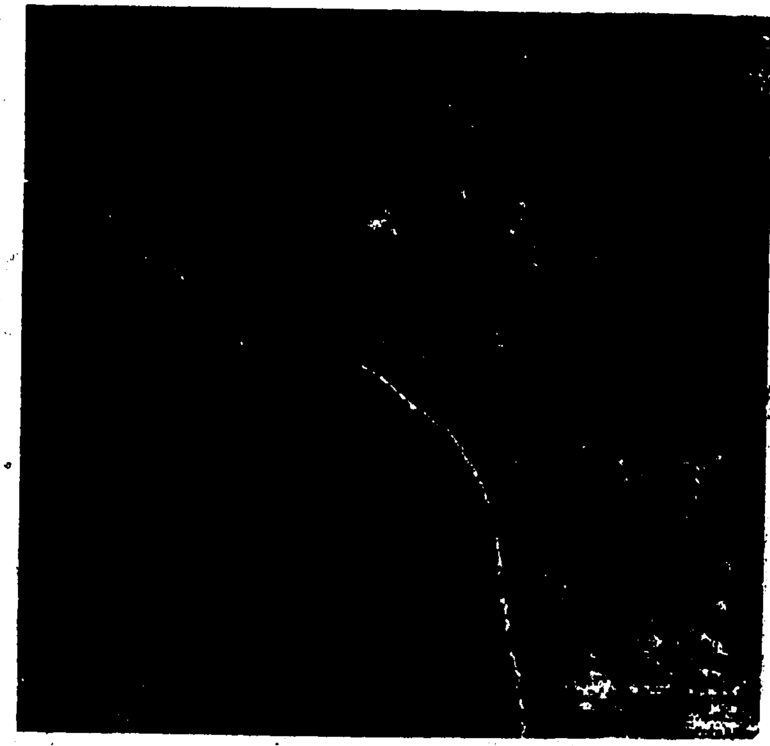

Fig. 33. Lake I.

Fig. 34. Lake II.

registration. It should be noted that the size of the registered pictures varied from 256*256 to 501*501. In the first case, two sub-pictures from LANDSAT images 1043-18341 (Sept. 1972) and 1600-18242 (March 1974) were extracted. The size of the pictures is 256*256 and 384*384 and they are shown in Fig's 33 (Lake I, band 5) and 34 (Lake II, band 5). The Lake I picture was obtained from image 1043-18341, with coordinates (2030, 2285, 2691, 2946). The Lake II picture was extracted from image 1600-18242, with coordinates (15,398,1536,1919). The window size around the control points was chosen as 9*9 with the maximum number of iterations set to 5. The registered picture of size 256*256, as shown in Fig. 35, was generated by a nearest neighbor approximation and by applying a first order transformation, whose coefficients are:

$a = 0.99 \quad b = -0.019 \quad c = 1.87,$

$d = 0.036 \quad e = 1.01 \quad f = 52.97.$

The corresponding windows of the two registered pictures were cross correlated. The final registration accuracy for 20 reference points was about 1 to 2 pixels. In the window matching, normalized correlation failed at those points where objects had undergone marked changes, even though objects in the two pictures were essentially the same, namely, a lake from fall and winter seasons. To correct this problem, instead of using normalized correlation, it is proposed that the edges of the two windows be obtained by a gradient operator. The two binary gradient

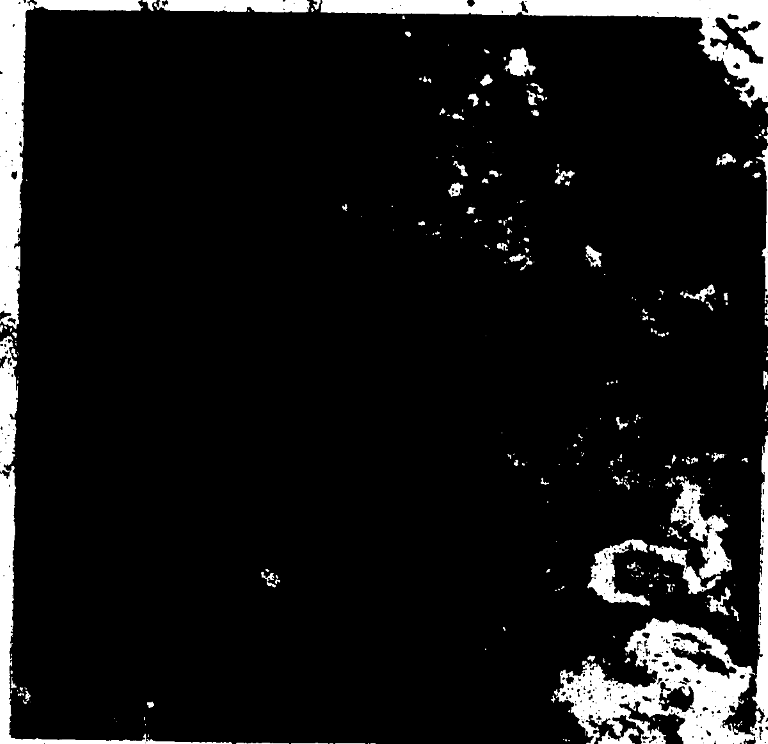

Fig. 35. Lake II registered with Lake I using nearest neighbor approximation.

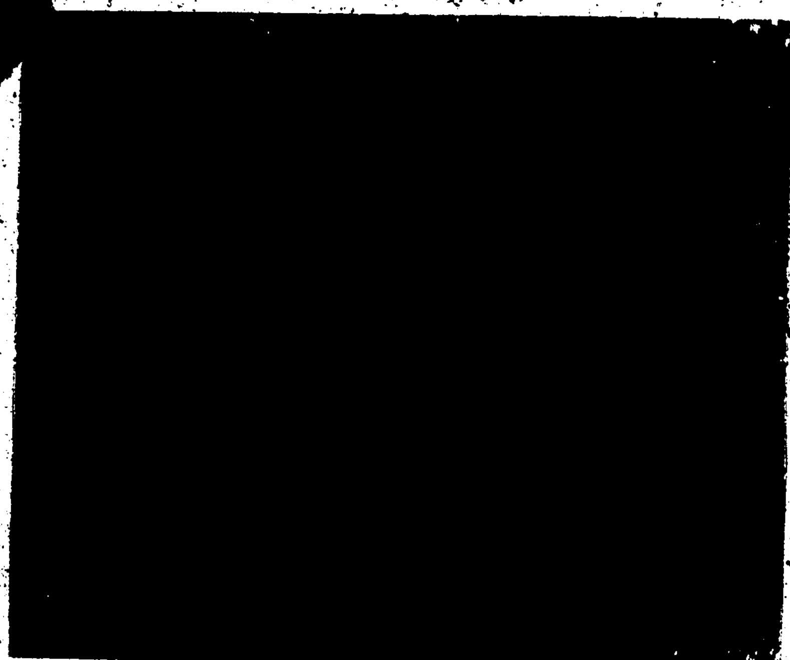

Fig. 36. Fertilizer 3 registered with Fertilizer 2 using nearest neighbor approximation.

pictures can then be registered. Comparing with Jayroe et al's [27] method, it follows that there is no need to generate full binary gradient pictures, but only 2*N*I windows, where N is the number of control points, I is the number of iterations (maximum of 5). Thus this scheme of registration is much faster than Jayroe et al's.

In the second test case, the initial control points were obtained by the automatic selection technique. The pictures are given in Fig's 31 and 32. The window size around the control points was chosen to be 15*15, with the maximum number of iterations set to 5. The best set of control points was chosen for generating the registered sub-picture of size 420*350. The registered sub-picture, as shown in Fig. 36, has a registration accuracy of 1 pixel for the second order transformation. The accuracy was calculated by using the two dimensional cross correlation function at 20 specific and random test points. The coefficients of the second order transformation are:

a = -60.71 b = 0.89 c = 0.11 d = -0.7E-6 e = 0.6E-4,
f = -0.11E-3 g = 201.35 h = -0.11 i = 0.88 j = 0.4E-6,
k = -0.7E-4 l = 0.2E-4.

The registered picture was also generated by using a two dimensional polynomial function among the input pixels as shown in Fig. 37. Bilinear interpolation smooths pictures and takes care of discontinuities. The nearest neighbor and bilinear techniques were also applied to the green and blue bands of the color photograph, of the fertilizer region pictures. The registration accuracy was within 1 pixel for all bands. It should be noted that the two dimensional polynomial function was also used in generating the windows for the iterative feedback process. The set of final control points is more accurate, and therefore, the transformation coefficients are not the same as obtained for the nearest neighbor approximation. The coefficients of the second order transformation are:

a = -61.5 b = 0.90 c = 0.10 d = -0.7E-6 e = 0.8E-4,
f = -0.12E-3 g = 201.7 h = -0.12 i = 0.89 j = 0.4E-6,
k = -0.11E-3 m = 0.4E-4.

The registered picture was also generated by using the cubic convolution function [47]. The gray level of the output pixel was computed by using the 16 gray levels of the input picture. The weights used were derived from the truncated sinx/x function. The coefficients of the second order transformation are:

a = -60.2 b = 0.9 c = 0.10 d = -0.14E-5 e = 0.7E-4,
f = -0.1E-3 g = 203.8 h = -0.13 i = 0.89 j = 0.9E-6,
k = -0.13E-3 m = 0.72E-4.

The registered picture is given in Fig. 38 and has a registration accuracy of about 1 pixel. The accuracy was checked by cross correlating corresponding windows as in the previous interpolation techniques.

The third test case is concerned with the registration of Edmonton I and Edmonton II pictures as shown in Fig's 28 and 29. Using the automatic control point approach, control

Fig. 37. Fertilizer 3 registered with Fertilizer 2 using bilinear approximation.

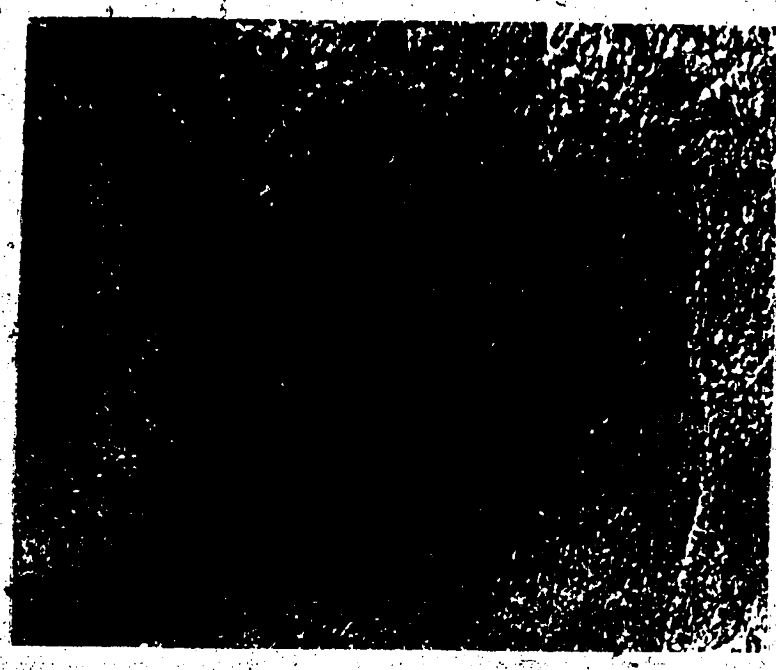

Fig. 38. Fertilizer 3 registered with Fertilizer 2 using cubic convolution.

Fig. 39. Edmonton I registered with Edmonton II using nearest neighbor approximation.

points were obtained for the least squares analysis. The coefficients of the second order transformation are:

a = 61.20 b = 1.0 c = -0.012 d = 0.76E-7 e = 0.1E-5,
f = 0.3E-4 g = 64.73 h = 0.038 i = 1.03 j = -0.46E-7,
k = -0.15E-4 m = -0.14E-3.

The registered picture, given in Fig. 39, has a registration accuracy of about 1 pixel and was checked both visually on a line printer map and by correlation at 20 random points. As band to band variation is of interest in change detection, the other three bands of the two pictures were also registered. It should be noted that the same transformation function was used for obtaining registered pictures of all bands.

The final test cas[illegible] with the registration of forestry cut areas of northern Alberta. Digital data from LANDSAT images 1043-18341 (Sept. 1972) and 1439-18325 (Oct. 1973) were extracted. The pictures are 512*512, but only 256*490 windows are shown in Fig's 40 (Cut I, band 5) and 41 (Cut II, band 5). Cut I picture was extracted from the 1043-18341 image, with coordinates (256,767,1731,2242). Cut II picture was obtained from the 1439-18325 image, with coordinates (200,711,1880,2391). The control points were selected automatically and corrected by the iterative feedback approach. The coefficients of the second order transformation are:

a = 134.8 b = 1.0 c = 0.021 d = -0.32E-6 e = -0.25E-4,
f = -0.65E-4 g = 11.31 h = 0.17E-1 i = 1.01 j = -0.15E-6,

Fig. 40. Forestry Cut I.

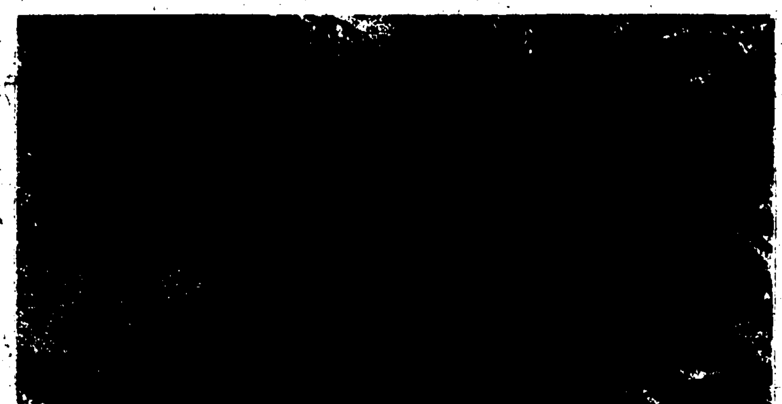

Fig. 41. Forestry Cut II.

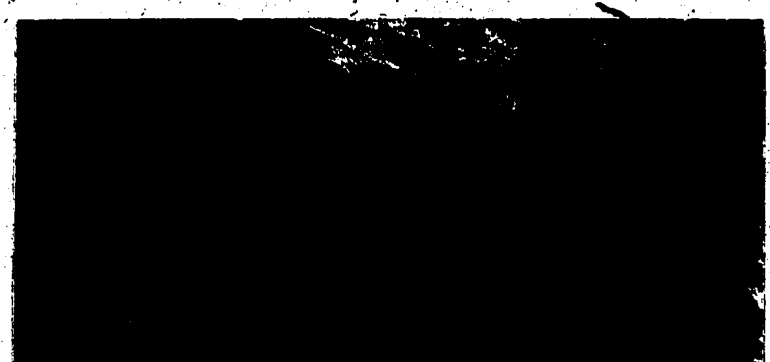

Fig. 42. Forestry Cut I registered with Forestry Cut II using nearest neighbor approximation.

$k = 0.97E-5$ $m = -0.1E-3$.

The registered picture shown in Fig. 42 was generated by applying the transformation to the picture given in Fig. 40. The registration accuracy was about 1 pixel and was checked visually and by doing correlation at 20 selected and random points. The automatic control point procedure gave a list of 196 potential control points. 90 control points were not considered since they were missing in the second picture. The root mean square error for the 70 control points was 0.8 pixels. The 20 control points with low correlation values were not taken into consideration.

3.5 Registration with Large Scale Differences

This section deals with the problem of registration of pictures which have been obtained by different methods, e.g., aircraft and LANDSAT. Previous registration techniques have limited the scale factors to be about 1:1.5. In surveying registration techniques, it was noticed that alignment of pictures with scale differences as large as 1:5, had not been reported. A technique is presented to solve this problem and is tested on digital pictures with scale differences of about 1:6 (horizontally) and 1:4.2 (vertically). Since the first LANDSAT satellite was launched in 1972, the proposed technique is very useful for detecting changes before 1972 (assuming aircraft data exists).

In the standard *nearest neighbor approach*, the gray level corresponding to the rounded address from the large picture is used for the registered picture. If this approach is used for registration, then the registered picture will be degraded. For example, if the scale factor is 1:b, then this technique chooses one pixel from the b*b neighborhood. If b is greater than 2 and if the neighborhood contains edge transitions, a pixel corresponding to either high or low contrast can be chosen. This causes the straight edges in the registered picture to look irregular, and non-noisy regions to appear noisy.

In the <u>averaging approach</u>, by knowing the scale factors, the larger picture can be averaged to reduce the scale and then the two pictures can be registered. If this is done, registration may not be accurate. The reason for this is that averaging changes the relative coordinates and size of the objects in the picture. Also, because of averaging, the registered picture will be smoothed. The larger picture can also be <u>sampled</u> to reduce its size, but the reduced picture will still be degraded.

In order to take care of these problems, an <u>improved technique</u> is presented as follows: Suppose I is the row number of the registered picture to be generated, then the row and column addresses for rows I-1, I, and I+1 are generated as shown in Fig. 43. Let the column under consideration be J, i.e., the output pixel is located at row I and column J. The registered picture is generated row by row. The advantage is in the saving of computer storage, since the registered picture need not be kept in core.

```
           J-1     J     J+1

   I-1      X-----X-----X



    I       X-----X-----X



   I+1      X-----X-----X
```

Fig. 43. Rectangular region.

In the proposed technique, the equations of four lines which are midway between (I-1,I), (I,I+1), (J-1,J) and (J,J+1) are calculated. These lines define a rectangular region, the coordinates of which are shown as four dots in Fig. 43. The output pixel is assigned the average gray level of the region defined by the four dots, whose shape also varies with I and J. Since the boundary of this region is not aligned with the grid, two approaches for defining the area within the region exist, which are:

1. <u>Zeroth order approach</u>. In this approach, the addresses of the pixels along the boundary are rounded, i.e., a pixel is included or not depending on whether or not its coordinates are inside or outside the boundary. This approach is simple to use but may not be very accurate.
2. <u>Refined approach</u>. In this approach, the pixel's contribution to the region, as defined by the boundary line, is calculated, i.e., the gray level of the

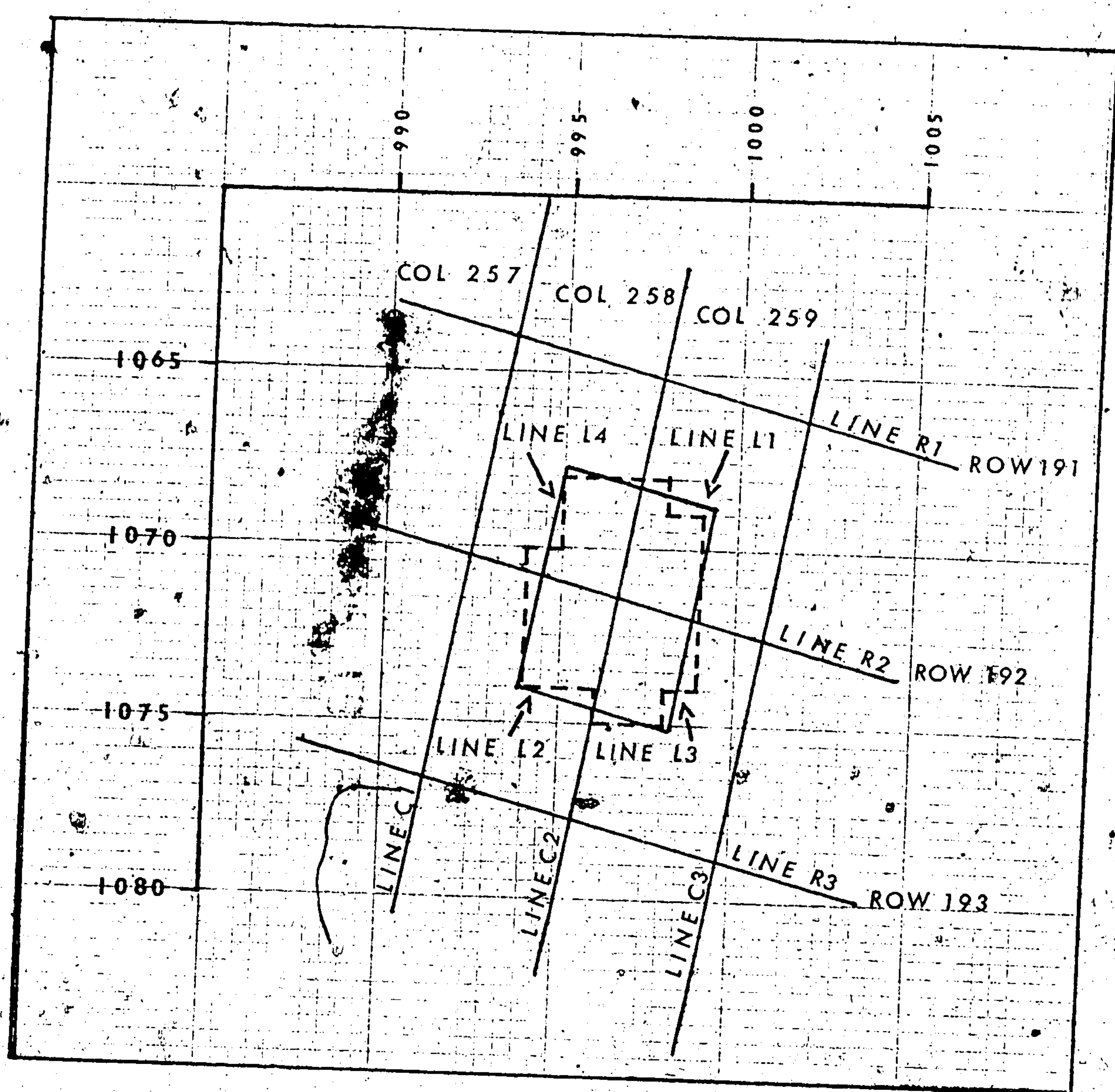

Fig. 44. Zeroth order approximation to the rectangular region.

boundary pixel is proportional to the area occupied by the pixel within the boundary line. This approach is very accurate; however, it is computationally expensive.

In order to compare the nearest neighbor and the proposed approximation, processing for one output pixel will be discussed in detail. For the output pixel corresponding to [illegible]ow 192 and column 258, the address generated is (1071.4,996.6), see Fig. 44. The gray level corresponding to the address (1071,997) of the picture will be used for the nearest neighbor approximation. It is evident from Fig. 44 that this approximation is not only very poor but also does not use any information from the pixel's neighborhood. Also, due to large scale differences, the distance between lines R1, R2, R3 and C1, C2, C3 connecting the input pixels is large. In fact, [illegible]e difference is about 5.99 pixels horizontally and about 4.26 pixels vertically.

In the proposed approximation, the equations of lines L1, L2, L3, and L4 which are midway between lines (R1,R2), (R2,R3), (C2,C3) and [illegible] respectively are calculated. These four lines defi[illegible]ngular region, which is the scaled down version of [illegible]on defined by lines R1, R3, C1, and C3. The area within this region is of interest for computing the average gray level for the input pixel with coordinates (1071.4,996.6). Since the boundary of this region is not aligned with the grid, the _zeroth order approach_ along the boundary is taken. The dotted line in Fig. 44 represents the approximation to the rectangular region. The average of the gray levels of the region defined by the dotted lines is assigned as the gray level for the input pixel with coordinates (1071.4,996.6). It should again be noted that the shape of this approximation is not fixed but varies from pixel to pixel.

The test problem is concerned with registration of an aerial picture of Edmonton (1971) with a LANDSAT picture (1975) of size 600*600, extracted from 1160-17553 image. The LANDSAT picture as shown in Fig. 45 is of size 281*400 (window of Fig. 29) and the aerial picture's size is 2048*2048. A 400*256 window of the aerial picture is shown in Fig. 47. In order to compare the proposed technique with the averaging approach, the aerial picture was averaged as shown in Fig. 46, i.e., from a 4*4 neighborhood was averaged to produce the size reduction. Control points in both pictures were selected manually and for the nearest neighbor and the proposed technique, control points were obtained from their original pictures, i.e., from the aerial (2048*2048) and the Edmonton II (600*600). The registered pictures as shown in Fig's 48-50, were generated by a second order transformation, whose coefficients for the averaging approach are:

$a = -94.12$ $b = 1.46$ $c = 0.38$ $d = 0.4E-6$ $e = -0.35E-3$,
$f = 0.14E-3$ $g = 12.45$ $h = -0.167$ $i = 1.092$ $j = -0.6E-6$,
$k = -0.39E-3$ $m = 0.9E-4$.

The transformation coefficients for the nearest neighbor and proposed approximation are:

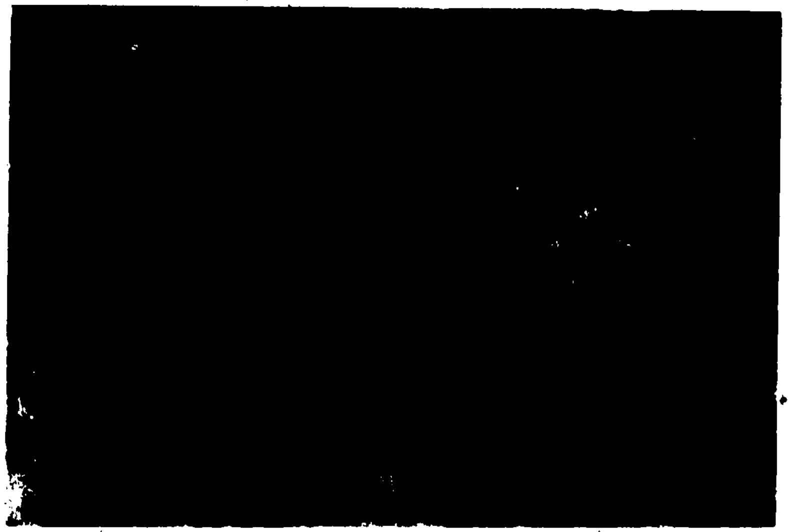

Fig. 45. Window of Edmonton II.

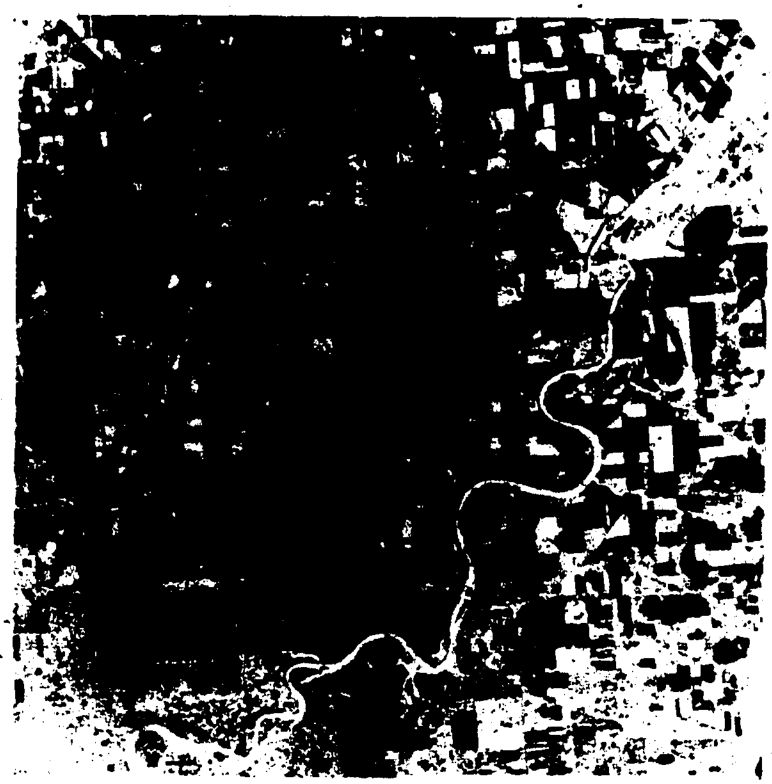

Fig. 46. 4*4 average of aerial picture.

Fig. 47. Window of aerial picture.

Fig. 48. Average aerial picture registered with Edmonton II using nearest neighbor approximation.

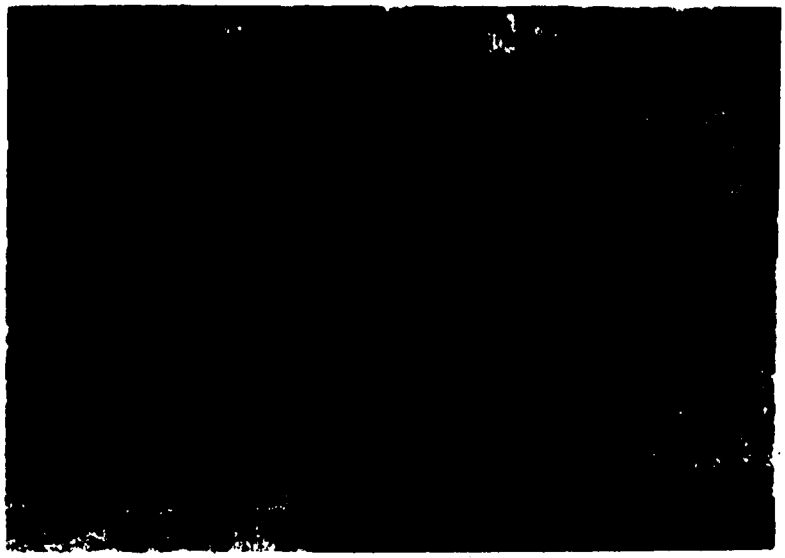

Fig. 49. Aerial picture registered with Edmonton II using nearest neighbor approximation.

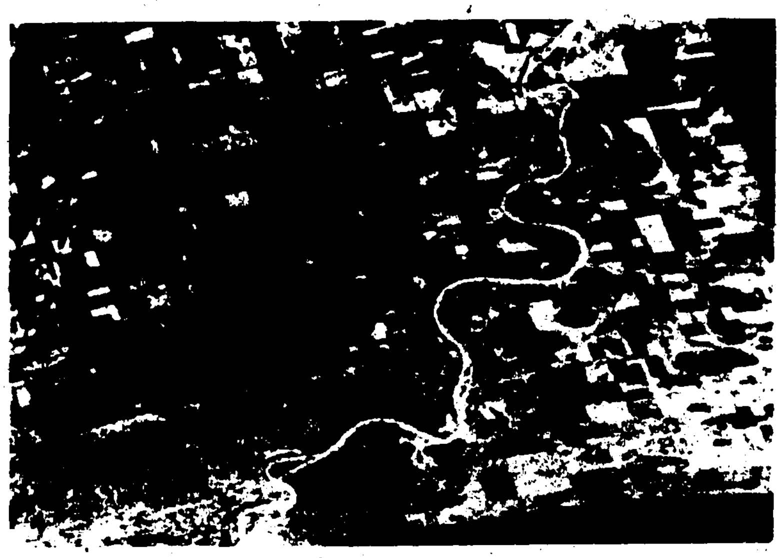

Fig. 50. Aerial picture registered with Edmonton II using proposed approximation.

a = -376.51 b = 5.84 c = 1.52 d = 0.16E-5 e = -0.14E-2,
f = 0.57E-3 g = 49.80 h = -0.66 i = 4.36 j = -0.2E-5,
k = 0.38E-3 m = -0.15E-2.

The nearest neighbor picture was generated by using the gray level of the pixel, which is closest to the calculated row and column addresses of the original aerial picture. In the proposed technique, the zeroth order approach was used in calculating the rectangular region's average gray level, see Fig. 44 for a particular output pixel. Clearly, this picture is the best visually by comparing with the other two registered pictures. In order to check registration accuracy, the correlation approach cannot be followed, because of severe changes. Visually, the registration accuracy was found to be reasonable, i.e., within 1 to 2 pixels. The computational processing time for registration was high for both the nearest neighbor and the proposed technique. The equations of lines need not be generated for every output pixel, i.e., for each output row, equations of lines L1 and L2 will not change, and only line L3 needs to be recomputed.

3.6 Discussion of Results

This chapter deals with registration of pictures. In the automatic control point selection, some basic features of a control point and problems encountered in selecting them are given. Control points are selected on the edges of the objects in the pictures. The technique is independent of gray levels in the pictures and does not use any pre-specified objects or templates.

In the iterative feedback approach, the control points are corrected to produce the best set of points. The registration accuracy was between 1 to 2 pixels for the first order transformation and 1 pixel for the second order transformation. The accuracy was computed by cross correlating corresponding windows at 20 reference points. The root mean square error was 0.8 pixels and the size of registered pictures varied from 256*256 to 501*501.

In the registration of pictures with large scale differences, a new technique is presented. The row and column coordinates of the input picture for the three output rows, i.e., previous, current, and the next are calculated. These coordinates define a region, the shape of which shows the scale differences. In the nearest neigbor approach for the output pixel, only 1 pixel is used from this region. If the size of the region is large, most of the important information about the region is lost. In the new technique, 4 boundary lines define a smaller version of the region. Since the output picture's grid is not aligned with the input picture's grid, a zeroth order approach along the boundary is taken. The average gray level of the new region is computed and used for the output pixel. The technique is compared with the nearest neighbor and the averaging approaches. The picture generated by the proposed technique is the best visually, as objects look smooth and non-noisy.

Chapter IV

TEMPLATE MATCHING

Template matching is a basic technique of pattern recognition, which has been used extensively in character recognition. Templates from a data base are compared with the incoming patterns and a correlation coefficient is calculated for each template. The template with the greatest resemblance to the incoming pattern is the one having the maximum correlation coefficient. In digital picture matching, for finding an object or its multiple copies, a template or window is displaced over all possible points of the picture. In this chapter, methods for the movement of the template across the picture and for comparing with the sub-pictures will be examined. Various search techniques for moving a template include, sampling at certain points [21,22], and planning [30].

If correlation is negative, full computation is still done for all the corresponding rows of windows and sub-pictures. Thus Barnea and Silverman's [6] observation that computations need not be done with high accuracy at every reference point is very important. In their approach, random samples from the sub-picture and a window are taken and a difference sum is continuously monitored. If that sum exceeds some pre-specified threshold, the reference point is dropped from further investigation.

One of the main problems in using a difference measure is that the mean of the window and means of the reference sub-pictures have to be computed. Also, the difference measure does not take care of the constant offset problem. Moreover, the threshold has to be specified. Barnea and Silverman do not suggest any method for choosing the pre-specified threshold; however, recently Onoe and Saito [41] have proposed a technique for automatic threshold setting.

Tasto and Block [60] describe a template matching method for detecting objects of known shapes, size and orientation. The template and the picture are reduced to binary images by gradient methods, to produce a contour map and a line drawing. The template is moved over the picture and the city block distance

$$d\{(i,j);(h,k)\} = abs(i-h)+abs(j-k),$$

is computed. The authors also give a reformulation of computational equations for efficiency. Rosenfeld and Pfaltz [52] suggest using the following measure:

$$d\{(i,j);(h,k)\} = \max[(2*abs(i-h)+abs(j-k))/3, \max\{(abs(i-h),abs(j-k))\}].$$

Nagel and Rosenfeld [39] sort the gray levels of a window in increasing order and note their coordinates. The same procedure is also used for the sub-pictures. The authors report that the absolute difference measure sum grows much faster for a non-match point. This technique is shown by the authors to be faster than the row by row correlation. The basic concept is the same as that of Barnea and Silverman.

4.1 The Proposed method

It is assumed that the object to be located is present in the picture. To take care of constant gain and offset, normalized correlation will be used as a similarity measure. The correlation between a window and a sub-picture is calculated by using all the rows; however, intermediate results can give an insight into the trend of the sign or value of the correlation coefficient. If the value is either negative or small, then the reference point can be dropped. The concept of a partial correlation coefficient is defined as:

Definition 5: A partial correlation coefficient is an intermediate correlation coefficient, when successive rows of a window and sub-picture are correlated. If the row size of a window is N, then a maximum of N partial correlation coefficients exist.

Let the window size be 4*4, then while matching, 4 partial correlation coefficients are computed for each of the 4 rows. In other words, the first coefficient is computed by matching the first row of a window with the first row of the sub-picture. The second coefficient is computed by matching the corresponding second rows and updating the first coefficient. It should be noted that the value of the Nth partial correlation coefficient is same as the correlation coefficient.

The proposed method will take the following points into consideration:

1. For matching a window with sub-pictures, not all the rows need to be correlated for a non-match point. This point is implemented by monitoring the partial correlation coefficient at three check points.
2. The unit shift displacement of a window can be speeded up in both directions, i.e., a number of rows and columns of reference points can be skipped. For one approach, alternate rows and columns are evaluated to check whether the correlation is above some threshold.
3. While correlating rows, the order in which rows are correlated can be changed.

Let k1,k2,k3 be pre-specified constants. The range of these constants is between -1 and +1. For an M*M window, the partial correlation coefficient is computed and the value is monitored at three check points, namely M/4, M/2 and 3M/4. If the value of the coefficient is ≤ k1, when M/4 rows have been correlated, then computation is stopped for that particular reference point. If not, computation is carried on until M/2 rows have been processed. If the value of the coefficient is ≤ k2, the reference point is dropped, otherwise computation is continued until 3M/4 rows have been processed. At the third check point, 3M/4, if the partial correlation coefficient is ≤ k3, the reference point is skipped, otherwise computation is continued to M rows.

For <u>skipping column reference points</u>, the correlation is computed at the first reference point. If the correlation value is ≤ k3, the next column point is skipped. Processing resumes at the third column point. For example, for columns 1,2,3,4,5,...,M, points 3,5,..., are evaluated, if points 2,4,..., were ≤ k3. Thus the first point is computed, and if the value is ≤ k3, the second point is skipped. The third point is then computed and so on.

For <u>skipping complete rows of reference points</u>, the first row is evaluated. If correlation values for all the row elements are ≤ k3, the evaluation of the next row is not done. Processing resumes with the third row. The motivation for skipping rows and columns of reference points and of monitoring correlation values at check points, is to <u>decrease the computational time</u>.

It was assumed that in matching, the first row of the two pictures was correlated, then the second row was correlated and so on. The <u>ordering of rows</u> can be changed by using a random number generator to generate a non-repeating integer sequence of numbers between 1 to M. Then, rows can be correlated according to the ordering of the integer numbers.

A general FORTRAN program was written to implement the proposed method, Barnea and Silverman's method, and conventional correlation. A further option could choose the prespecified or random ordering of rows for matching. The fast method with random row ordering will be called __Fast Random__. Barnea and Silverman's method was implemented with Onoe and Saito's [41] automatic threshold setting technique. The absolute difference measure and normalized absolute difference measure were used as similarity measures for the Barnea and Silverman's method. For the normalized absolute difference measure, means of the template and sub-pictures were subtracted before correlating them. The normalized absolute difference measure with random row ordering will be called as __NADM Random__. The values of k1,k2,k3 for the test problems were taken to be 0.0, 0.2, and 0.3 respectively.

In the first test problem, LANDSAT pictures from two different years were registered, using the iterative feedback correction technique, see Section 3.4. The two registered pictures are given in Fig. 33 (original) and Fig. 35 (registered) respectively. A 32*32 window was extracted from the coordinates (112,64) of the original picture. The problem was set to locate this window in the second picture. The range of search coordinates was limited to (100,30), (100,79), (149,30) and (149,79). The total number of reference points is 50*50=2500. For a 32*32 window, the total number of rows to be correlated is 32*2500=80,000. Details of using six matching techniques are given in Table XIII. It should be noted that CPU time is the total computational time to read in the pictures, complete the processing, search for minimum or maximum, and print the results.

Table XIII. Computational Times for Correlation Methods I.

Method	Rows correlated	Saving %	CPU secs	Ratio	Match point
1. Fast	8455	89.4	5.942	1.00	(112,64)
2. Fast Random	10086	87.4	11.267	1.89	(112,64)
3. Barnea with abs. diff. measure	59218	27.5	19.628	3.30	(101,30)
4. Barnea with normalized abs. diff. measure	56229	31.2	26.318	4.42	(138,56)
5. NADM Random	57939	29.1	39.222	6.60	(138,56)
6. Conventional	80000	0.0	27.158	4.57	(112,64)

In the second test problem, two pictures from a LANDSAT image were extracted. The 91*96 pictures are from band 6 and band 7 of image 1600-18242. The pictures were extracted with coordinates (181,271,1744,1839). The pictures correspond to the lake area, as shown for band 5 in Fig. 33. A 32*32 window corresponding to the lake in the band 6 picture was used as a template. The coordinates of the top left corner are (23,31). The range of search coordinates was limited to (20,20), (20,59), (59,20) and (59,59). The total number of reference points is 40*40=1600 and the total number of rows to be correlated is 32*1600=51,200. All six matching techniques were applied and the results are given in Table XIV. The proposed method was also used to locate control points for 3 test pictures.

Table XIV. Computational Times for Correlation Methods II.

Method	Rows correlated	Saving %	CPU secs	Ratio	Match point
1. Fast	10730	79.0	3.862	1.00	(23,31)
2. Fast Random	12009	76.5	8.867	2.29	(23,31)
3. Barnea with abs. diff. measure	28773	45.3	8.639	2.23	(23,31)
4. Barnea with normalized abs. diff. measure	17686	66.9	10.605	2.74	(23,31)
5. NADM Random	16668	68.9	19.088	4.94	(23,31)
6. Conventional	51200	0.0	16.123	4.17	(23,31)

For the first test problem as summarized in Table XIII, it follows that the proposed method is the best among all the other methods. Barnea and Silverman's method failed to locate the proper match point. The two different measures gave different incorrect matching points. By applying random row ordering, the proposed method and Barnea and Silverman's method took more computational time. Again, Barnea and Silverman's method did not find the correct match point. As expected, conventional correlation by taking more time found the correct match point. The surprising thing is that Barnea and Silverman's normalized measure took about same time as conventional correlation for this example. One of the reasons is that sub-picture means have to be subtracted at every reference point.

In the second test problem, as summarized in Table XIV, all six methods obtained the correct matching point. The proposed method is again the best. For random ordering, computational time was high with no apparent gain. In order to summarize, the following conclusions are drawn:

1. The proposed method is the best among the six tested. This method is applicable to any type of pictures, as constant gain and offset problems are taken care of.
2. Random row ordering should not be done. It consumed more computational time, with no apparent gain.
3. Barnea and Silverman's method should be used very carefully and only for those pictures where constant gain and offset problems do not arise.
4. The proposed method is 2 times faster than Barnea and Silverman's method and about 4 times faster than conventional correlation for the test problems.

4.2 Fast Correlation Coefficient

The purpose of this section is to compute the upper bound of the correlation coefficient R, in terms of multiplications and additions and to propose faster methods for computing R for two dimensional data, i.e., picture and window matching. The notion of running sums will be used to save computations. In picture matching, for finding the best template match, a similarity measure like normalized cross correlation is often used, which is given by:

```
                 Sum {(x-xmean)*(y-ymean)}
R  = -----------------------------------------------,
      Sqrt{(Sum(x-xmean)**2)*(Sum(y-ymean)**2)}
```

where x,y are the n-tuple data vectors, and xmean,ymean are the respective means. For computational efficiency, this can be written as:

```
                    n*Sum x*y -Sum x*Sum y
R = -----------------------------------------------------.
     Sqrt[{n*Sum x**2-(Sum x)**2}*{n*Sum y**2-(Sum y)**2}]
```

Consider a large picture X of size L*L and a small window Y of size M*M as shown in Fig. 51. The total number of reference points where this window can be displaced is $(L-M+1)^2$. The number of multiplications and additions for a conventional method for computing Sum x and Sum x^2 terms are $(L-M+1)^2*M^2$ and $(L-M+1)^2*(M^2-1)$ respectively. It is shown that by using algorithm F, the computations can be reduced to $L*(2L-M)$ and $(2L-M)^2-1$ multiplications and additions respectively. Also, algorithm F's functions are monotonic decreasing with respect to M, as M increases. As M-->L, the functions are L^2 and L^2-1, where as M-->0, the functions are $2L^2$ and $4L^2-1$.

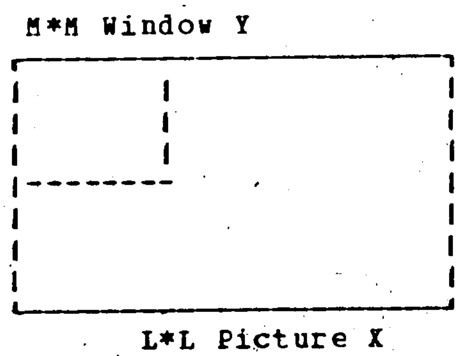

Fig. 51. Large and small pictures.

It should be noted that since the terms Sum y, Sum y^2 in R remain the same, they have to be computed only once. Also, the upper bound will only be considered for terms like Sum x, Sum y, Sum x*y, Sum x^2 and Sum y^2. The total number of operations for Sum x^2, Sum y^2, and Sum x*y terms are M^2 multiplications and M^2-1 additions. Similarly for Sum x and Sum y terms, the total number of operations are 0 multiplications and M^2-1 additions. The total number of <u>multiplications</u> for the conventional method are:

$$M^2+(L-M+1)^2*(M^2+M^2).$$

The first factor consists of M^2 multiplications for the term Sum y^2, and the second factor is for M^2 multiplications for the terms Sum x*y, Sum x^2 for all $(L-M+1)^2$

reference points. The total number of <u>additions</u> for the conventional method are:

$$2*(M^2-1)+3*(M^2-1)*(L-M+1)^2.$$

Since Sum y, Sum y^2 terms are to be computed once only and Sum x*y <u>may have</u> to be computed for all points, the only place one can speed up the process is in the computation of Sum x and Sum x^2 terms. The Sum x*y term can be efficiently computed by convolution, see [4]. Therefore algorithms have been designed to compute Sum x and Sum x^2 terms efficiently.

4.2.1 <u>Algorithm E for Sum x and Sum x^2</u>

Suppose a window W is being matched with the large picture, as shown in Fig. 52. Let Sum a1 be the <u>horizontal sum</u> of the first four elements of the first row and Sum $a1^2$ be the sum of squares of first four elements of the first row. Then Sum a1, Sum a2, Sum a3, Sum a4, Sum $a1^2$, Sum $a2^2$, Sum $a3^2$, Sum $a4^2$, are computed, where the range of the row sum is from 1 to 4.

```
                              4*4 Window W
a1  . . . . . . .             . . . .
a2  . . . . . . .             . . . .
a3  . . . . . . .             . . . .
a4  . . . . . . .             . . . .
a5  . . . . . . .
```

Fig. 52. Computation of running sums.

Next, for computing correlation for a displacement of

1 in the column direction, the terms Sum a5, Sum $a5^2$, are added to Sum a2, Sum $a2^2$, Sum a3, Sum $a3^2$, Sum a4, Sum $a4^2$ respectively. Thus for every column displacement, only one row has to be computed, i.e, M multiplications and 4(M-1) additions, instead of M^2 multiplications and $2(M^2-1)$ additions. When the template is moved to the right, for displacement in the x-direction, this procedure can be repeated. The algorithm can be summarized, as follows:

<u>Step</u> 1: Compute a1,a2,a3,...,aM for an M*M window.

<u>Step</u> 2: Compute aM+1,aM+2,...,aL which gives the ith column of the correlation matrix.

<u>Step</u> [illegible]mpute a1,a2,a3,...,aM for the (i+1)th column.

<u>Step</u> 4. Repeat steps 2-3 for all possible displacements.

The total number of <u>multiplications</u> for Algorithm E are:

$$M^2+(L-M+1)^2*M^2+Multa,$$

where

$$Multa = (L-M+1)*(M^2+M*(L-M)).$$

The first and second factors are contributions from the terms Sum y^2 and Sum x*y respectively. The first term in Multa is the contribution from the initial processing, the second factor has M multiplications per L-M rows per (L-M+1) points. The term Multa can be written as:

$$L*M*(L-M+1).$$

The total number of <u>additions</u> are:

$$2(M^2-1)+(L-M+1)^2*(M^2-1)+addx+addx^2,$$

where

$$addx = (L-M+1)*((M-1)*M+M-1)+2(L-M+1)*(M-1)*(L-M).$$

The first term is the contribution from Sum y and Sum y^2 and the second term is due to Sum x*y. The expression for addx can be simplified to:

$$(L-M+1)*(2M*L-M^2-2L+2M-1).$$

Similarly $addx^2$ term has also the same value.

4.2.2 Algorithm F for Sum x and Sum x^2

The motivation of this algorithm is to compute running sums in two dimensions, with no extra storage. The operations are done 'in place', with the exception of two vector arrays of maximum row size. This algorithm is an improved version of algorithm E.

Firstly, row sums are computed for the range of 1 to M. This sum is again summed, i.e., now a column sum (two dimensional sum) is obtained. Next, while the window moves down for a column displacement, only one row is summed horizontally, the previous row sum (row refers to the window which it has left upon this displacement) is subtracted from the two dimensional sum, and the new row summed is added to the two dimensional sum. In this manner, column processing is done. When window moves right, the procedure of adding one element and subtracting the previous element is followed. The original vector is modified now and processing starts again as in the column displacement. The algorithm F can be summarized as follows:

Step 1: Compute a1,a2,a3,...,aM for an M*M window, where a1 is the row sum of first row, a2 is the row sum of second row, and so on. Similarly, compute the sum of squares terms, i.e., $a1^2$, $a2^2$,...,aM^2.

Step 2: Summation of a1,a2,...,aM gives the Sum x term for the first correlation point. Save this sum as Tem. Similarly, add $a1^2$, $a2^2$,...,aM^2 and save this sum as Tem2.

Step 3: For computing the second correlation point (column), compute aM+1, add aM+1 and -a1 to the previous sum Tem. Similarly, repeat for the sum of squares term. Repeating step 3 in this way, completes the processing of first column of the correlation matrix.

Step 4: When the window is moved right by a unit displacement, modify a1,a2,a3,...,aL by subtracting the previous element and by adding the next element needed. For example, for a 3*3 case,

$$a1 = a1-a11+a14, \quad a1^2 = a1^2-a11^2+a14^2,$$

where a11 and a14 are the first and fourth elements of first row. Repeat steps 3 and 4 for different columns.

Step 5: Repeat step 4 for all possible displacements.

The total number of multiplications for Algorithm F are:

$$M^2+(L-M+1)^2*M^2+Multb,$$

where

$$Multb = L*M+2(L-M)*L.$$

The first and second factors are contributions from the terms Sum y^2 and Sum $x*y$. The first factor in Multb is from M multiplications for L rows and the second factor is 2 multiplications per L rows per (L-M+1-1) columns. The total number of <u>additions</u> are:

$$2(M^2-1)+(L-M+1)^2*(M^2-1)+addx+addx^2,$$

where

$$addx = L*(M-1)+2(L-M)*L+(M-1)*(L-M+1)+2(L-M)*(L-M+1).$$

The first factor is the contribution from Sum y and Sum y^2 terms and the second factor is due to Sum $x*y$ term. The term for $addx^2$ is the same as addx. The first factor in addx is (M-1) additions per L rows (once only, column sum), the second factor is for 2 additions per column for (L-M+1-1) columns. The third factor is the (M-1) additions (first time) per column for (L-M+1) total columns, and the last factor is for 2 additions per row (L-M) for (L-M+1) total rows. The addx factor can be simplified to:

$$(2L-M)^2-1.$$

The number of <u>multiplications</u> for conventional, algorithm E, and algorithm F can be summarized as follows:

$M^2+(L-M+1)^2*M^2+(L-M+1)^2*M^2$ for conventional,

$M^2+(L-M+1)^2*M^2+(L-M+1)*L*M$ for algorithm E,

$M^2+(L-M+1)^2*M^2+L*(2L-M)$ for algorithm F.

Similarly, the number of <u>additions</u> are:

$2(M^2-1)+3(L-M+1)^2*(M^2-1)$ for conventional,

$2(M^2-1)+(L-M+1)^2*(M^2-1)+2(L-M+1)*(2L*M-M^2-2L+2M-1)$ for algorithm E,

$2(M^2-1)+(L-M+1)^2*(M^2-1)+2(2L-M)^2-2$ for algorithm F.

The number of multiplications and additions are given in the first and second rows of Table XV. The ratio of the operations taken by the conventional method as compared to the algorithm F is also given. The ratio indicates that algorithm F is about 1.9 to 2 times faster and about 2.6 to 3 times faster than the conventional method in the number of multiplications and additions, respectively. It should be noted that L is usually very large with respect to M.

Table XV. Number of Multiplications and Additions for Matching Methods.

L	M	CONVENTIONAL	ALG. E	ALG. F	RATIO
256	8	0.7936E+07	0.4478E+07	0.4097E+07	1.937
		0.1172E+08	0.5667E+07	0.4414E+07	2.655
256	16	0.2974E+08	0.1586E+08	0.1500E+08	1.983
		0.4443E+08	0.1840E+08	0.1530E+08	2.903
256	64	0.3051E+09	0.1557E+09	0.1527E+09	1.998
		0.4576E+09	0.1635E+09	0.1529E+09	2.992
256	128	0.5453E+09	0.2769E+09	0.2728E+09	1.999
		0.8179E+09	0.2853E+09	0.2730E+09	2.997
256	192	0.3115E+09	0.1590E+09	0.1559E+09	1.999
		0.4673E+09	0.1638E+09	0.1560E+09	2.995
256	256	0.1966E+06	0.1966E+06	0.1966E+06	1.000
		0.3277E+06	0.3277E+06	0.3277E+06	1.000
10000	10	0.1996E+11	0.1098E+11	0.1018E+11	1.961
		0.2965E+11	0.1348E+11	0.1068E+11	2.776
10000	100	0.1961E+13	0.9902E+12	0.9805E+12	2.000
		0.2941E+13	0.1019E+13	0.9810E+12	2.998
10000	1000	0.1620E+15	0.8111E+14	0.8102E+14	2.000
		0.2431E+15	0.8136E+14	0.8102E+14	3.000
10000	10000	0.3000E+09	0.3000E+09	0.3000E+09	1.000
		0.5000E+09	0.5000E+09	0.5000E+09	1.000

4.3 Discussion of Results

In this chapter, template matching methods are discussed. The need for using a normalized correlation coefficient as a similarity measure is pointed out, and a fast template matching method is developed. For one test problem, Barnea and Silverman's method did not locate the correct match point. The proposed method is about 2 times faster than Barnea and Silverman's method and about 4 times faster than conventional correlation for the test problems. The upper bound for the correlation coefficient is computed in terms of the number of multiplications and additions. Two algorithms for fast correlation are given. The number of multiplications for picture matching can be reduced from

$$M^2+2(L-M+1)^2*M^2 \text{ to}$$

$$M^2+(L-M+1)^2*M^2+L*(2L-M).$$

The number of additions can be reduced from

$$2(M^2-1)+3(L-M+1)^2*(M^2-1) \text{ to}$$

$$2(M^2-1)+(L-M+1)^2*(M^2-1)+2(2L-M)^2-2.$$

Algorithm F is about 1.9 to 2 times faster than the conventional method in the number of multiplications, and is about 2.6 to 3 times faster in the number of additions.

Chapter V

CHANGE DETECTION

Changes in a scene can result from both relative movement of objects and modifications in stationary objects. To detect a motion change, it is usually assumed that objects change in displacement only. If this restriction is removed, and changes in displacement, rotation and scaling are permitted, detection of motion changes by a machine becomes cumbersome. Changes in stationary objects occur due to shape, intensity, texture, etc., or some combination of them. Each factor introduces a type of change, which a general change detection system may not be able to detect. This chapter will deal with changes due to intensity, shape and location of objects in various scenes. An application in machine vision is that of a walking robot searching for different things. Various other applications of change detection include crop growth, detection of forest fires and natural diasters, environmental changes, military reconnaissance, land use management, urban growth, etc.

In the following paragraphs, various techniques of change detection will be surveyed, and the changes will be classified into different types. The technique developed for detecting changes is described and the results obtained on applying them on five test sets of pictures are discussed in the section 5.3.

5.1 Survey

Kawamura [29] describes an automatic change detection system for city planners. The two pictures are registered by least squares analysis, and then divided into small regions called cells. Measurements like correlation, entropy and probability are calculated for each cell pair. Using pattern classification techniques, the cells are classified into change category and no-change category. This system was the first automatic system for change detection.

Quam [45] has designed a system for detecting differences between pictures of the planet Mars. The two pictures are registered by a variant of the control point method, see Section 3.2. For corresponding sub-pictures, a two dimensional cross correlation matrix is computed for the integer shift parameters. By using interpolation between the correlation values, shift parameters in terms of fractional accuracy are obtained. The registered pictures are then subtracted pixel by pixel and displayed.

Lillistrand [35] describes a system, where portions of a scene are processed one at a time. Each sub-picture is cross correlated to obtain the shift parameters and then the sub-pictures are geometrically corrected to remove distortions. After applying the transformation, the registered sub-picture is generated and changes between the sub-pictures are detected. The system has the further capability of suppressing shadows or other features which are not of interest.

Ulstad [61] reports an algorithm for detecting small differences between two pictures. It consists of applying nonlinear spatial registration, then matching central moments of local areas for normalization of data, and finally, computing the pixel by pixel difference picture. The point of matching the local mean and variance is emphasized for accurate computation of the difference picture.

Price and Reddy [44] approach the change detection problem differently. The two pictures are registered by the control point technique and are then segmented into regions by thresholding the histograms. The statistical features of the regions are computed and then compared for any type of change to be detected. This is one of the first symbolic approaches to the change detection problem.

For land use changes, much work has been done at the U.S. Geological Survey [38,66]; however, the system is not automatic. Changes are identified by image interpreters who mount the two pictures on a zoom transfer scope. Land which is in the process of development or under construction can be recognized, by comparison with previous photographs. An automatic digital system would be very useful for this application.

Welch and Pannell [71] have used digital pictures for land use changes in China. Pattern classification algorithms are used for classifying individual pixels into various classes. The classification maps are compared for urban growth and land use in various sectors. The authors discuss the significance of changes for land use.

Ellefsen [16] begins by geometrically registering and correcting the two LANDSAT pictures. The four bands of both pictures are used to generate a new picture, such that every pixel has 8 gray levels assigned to it. The pixels which are similar to each other in the new picture, are grouped. A spectral map of various groups is obtained for interpretation. The process is iterative in the sense that, if any group cannot be identified properly, re-grouping of pixels is done again.

De Gloria et al [14] also use LANDSAT pictures to detect spring and summer changes. They do not mention any registration steps for picture alignment. The changes are obtained by manual as well as by digital means. The input pictures are classified into various categories such as water, snow, meadow, trees, etc., by applying pattern classification algorithms. The number of pixels belonging to each category is counted for each picture. The change in the number of pixels from one date to another is then computed. A similar approach has been followed by Christenson and Lachowski [11].

5.2 Type of Changes

A change can take place either in a short time or in a long time. A rapid change takes less time to occur, e.g., forest fires, oil leakage in sea, landslides, etc. A slow change takes long time to develop, e.g., land use pattern of developing cities, cutting and planting of forests, plant growth, etc. The type of changes can be classified as: small, large, false, pre-specified, motion, intensity, and texture. Each type of change is heuristically summarized as:

1. small: A small change is hard to see visually and is minute in size. An algorithm should be able to detect this change, as it may be important for military applications.
2. large: A large change is easy to locate, both visually and algorithmically.
3. false: A false change is important to detect, since it may occur due to many reasons, e.g., incorrect results produced by an algorithm. A change may exist, but an algorithm interprets it as a no-change and vice versa. It can also result from difference pictures which have been subtracted from unnormalized pictures.
4. pre-specified: In a pre-specified change, it is already known that a change has occurred at some location or in some sub-region, e.g., monitoring of the missile sites.
5. motion: In a motion change, some objects may have moved with respect to the other objects in the same scene,

e.g., monitoring the number of cars at a traffic circle by a robot or a TV camera.

6. intensity: An intensity change is due to variation in gray levels, because of different environmental conditions.
7. texture: The texture of objects can change with respect to time, e.g., plants have different texture at different stages of growth.

5.3 Method of Solution

In this chapter, the change detection problem will be studied with respect to the computation of difference pictures. If pictures have been obtained under identical conditions, then difference picture analysis will show changes clearly; however, under different conditions, normalization of pictures is essential. It should be noted that even though the pictures may be normalized, a registration error of more than 2 pixels will produce an incorrect difference picture, i.e., changes may not be reliable. The difference picture consists of the absolute difference of the gray levels of the two pictures. In order to know the magnitude of the gray level, which can be classified as a change, the threshold T should be computed. The selection of this parameter is not easy and is dependent on the nature of changes to be detected. If the range of the difference picture is (a,b), then gray levels from a to

T are assigned a gray level of 0 and from T+1 to b are assigned a gray level 1. The resulting picture is a binary change picture, where a gray level of 0 represents a change pixel and a gray level of 1 represents a no-change pixel.

If the binary picture contains noise or areas where registration accuracy is in question, an N*N averaging can be done for smoothing purposes. Isolated change pixels (0's) can be removed by an N*N cleaning operation, i.e., in a neighborhood size of N*N, the number of pixels with gray level 0 is found. If this number is less than (N*N)/2, then the gray level of the pixel is changed from 0 to 1.

For registered pictures, the absolute difference between the gray levels of the corresponding pixels will vary. Also, for a particular gray level, the difference does not remain the same, as the spatial coordinates of the pixels change. For example, the first picture may have many pixels with a gray level of 5. To the corresponding pixels of the second picture, the gray levels may vary from 1 to 30. A one dimensional histogram for gray level 5 can then be computed, which would give the frequency of numbers in the range (1,30).

The notion of a one dimensional histogram can now be easily extended to <u>two dimensional histograms</u>, by considering other gray levels in the first picture. The two dimensional histogram can be represented by a matrix, where the rows of the matrix are the frequencies of gray levels in the second picture, corresponding to the first picture's gray level. The two dimensional histogram can be defined as:

<u>Definition</u> 6: Let a1,a2,...,aN and b1,b2,...,bN be the gray levels of the first and second picture. Let Hi1,Hi2,...,HiN be the frequency of the mapping of the corresponding pixels for gray level ai to the gray levels b1,b2,...,bN. The matrix is given by:

	b1	b2	b3	...	bN
a1	H11	H12	H13	...	H1N
a2	H21	H22	H23	...	H2N
..	...	...	...	...	...
aN	HN1	HN2	HN3	...	HNN

Given two registered pictures and a binary change picture, the notion of the two dimensional histogram can be used to generate the change and no-change histograms, by considering only those pixels which have changed or not changed. Since the mapping for a particular gray level of the first picture to the set of gray levels of the second picture is one to many, the <u>weighted</u> <u>average</u> of the gray levels of the set is computed. This average is the mapping of the particular gray level in the first picture to the second picture. The mean deviation is then the absolute difference of the two gray levels. The procedure for computing the two dimensional histogram is summarized in algorithm G.

Algorithm G:

Step 1: Given original, registered and binary change pictures, generate change and no-change two dimensional histograms according to the following criterion. Scan the three pictures pixel by pixel and repeat the following for all the pixels. For the original picture's gray level, check the gray level in the registered picture and increment the corresponding column by one. If the pixel has changed according to the binary picture, the column of the change histogram is updated, otherwise increment the no-change histogram's column.

Step 2: Compute the sum and weighted average of the numbers in a row and assign this average to the average vector. Repeat step 2 for all the rows and both the change and no-change histograms.

Step 3: Compute the deviation vectors for both the histograms between the row index and the average vector, by considering only those gray levels which are present in both the pictures. Finally, compute the means of the deviation vectors, which gives the mean deviation for change and mean deviation for no-change pixels.

The computation of the change and no-change histograms can be illustrated with the help of an example. Consider the sub-picture matrices A, B and the binary change picture

C (for a threshold of 4) as given in Tables XVI-XIX.

Table XVI. Original Matrix A.

5	5	6	6
10	10	10	10
5	5	6	6
6	6	6	6

Table XVII. Registered Matrix B.

8	7	7	7
8	8	20	20
8	8	8	7
7	7	7	7

Table XVIII. Binary Change Matrix C.

1	1	1	1
1	1	0	0
1	1	1	1
1	1	1	1

Table XIX. Two Dimensional Histograms.

No-Change

	7	8	20	ave.	dev.
5	1	3	0	8	3
6	7	1	0	7	1
10	0	2	0	8	2

Change

	7	8	20	ave.	dev.
5	0	0	0	-	-
6	0	0	0	-	-
10	0	0	2	20	10

In order to detect various kinds of change, five sets of pictures were registered, see Chapter III. The types of changes consisted of: large, slow, small, and small with texture. In another set of pictures, the registered pictures were subtracted pixel by pixel for registration accuracy. It should be pointed out that for subtracting the two pictures, the following operation will be used:

$$p = 255 - abs(f - g),$$

where f,g are the gray levels of the two pictures and p is the gray level of the difference picture. The operation simply reverses the gray scale for displaying changes as black on the white background. The range of the Gaussian normalization for all the pictures was (0,255).

In the first test case, the two registered pictures as shown in Fig's 33 and 35, were Gaussian normalized. The resulting pictures, as shown in Fig's 53 and 54, were then subtracted pixel by pixel to generate a difference picture. The histogram of the difference picture is given in Fig. 55. From the histogram of the difference picture, the threshold was found to be 230. The difference picture was then thresholded at 230 to give a binary change picture, as shown in Fig. 56. Visually, the changes in the lake are clearly evident, which are due to snow on the lake. Empirically, the proposed algorithm was applied on the two registered pictures and the binary picture to yield a set of statistical parameters. The mean change deviation and mean no-change deviation from the first picture to the registered picture is given in Table XX (picture name Lake).

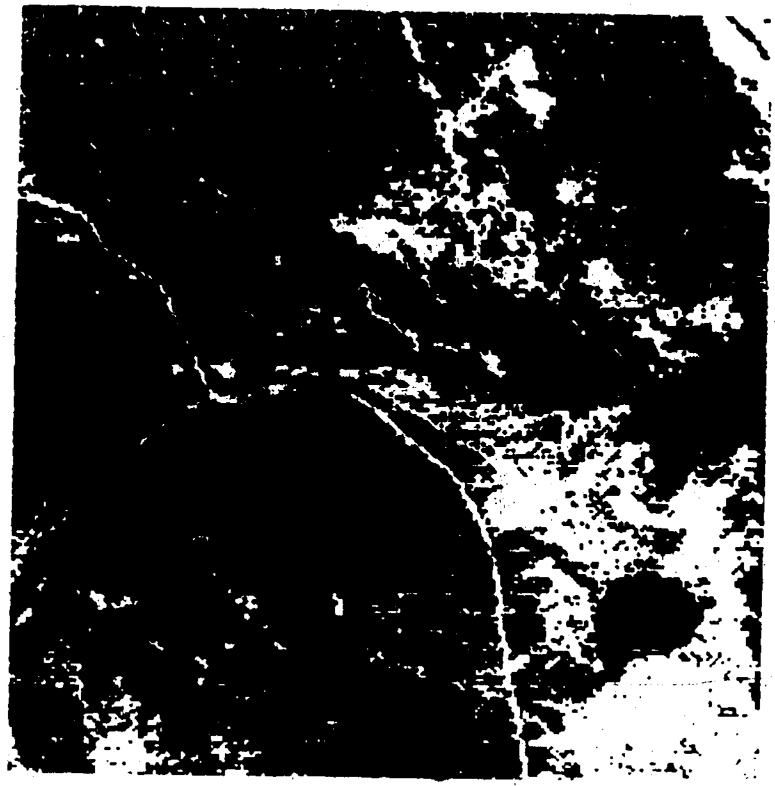

Fig. 53. Lake I with Gaussian histogram.

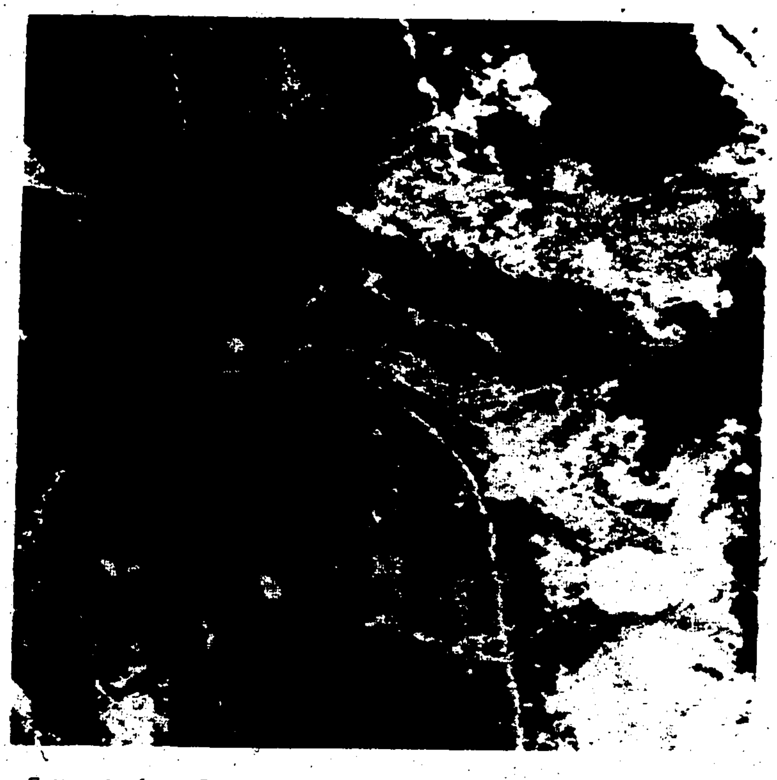

Fig. 54. Lake II (Gaussian) registered with Lake I (Gaussian) using nearest neighbor approximation.

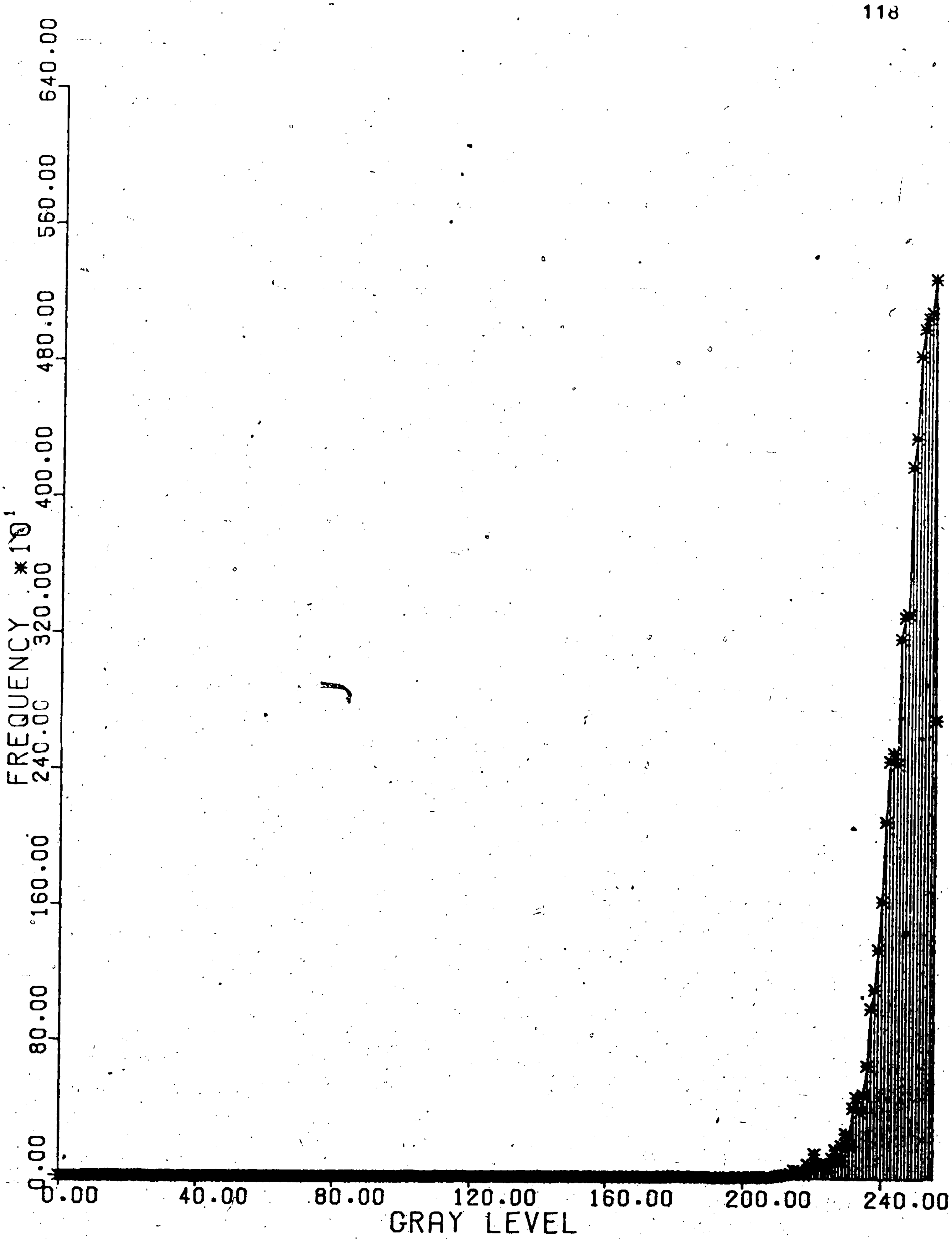

Fig. 55. Histogram of difference picture of Lake area.

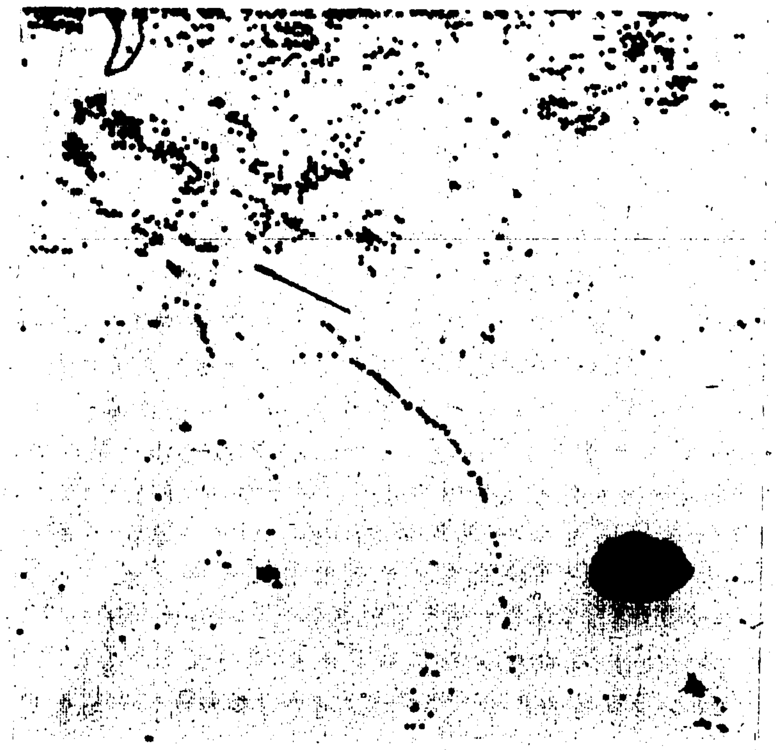

Fig. 56. Binary change picture of Lake area.

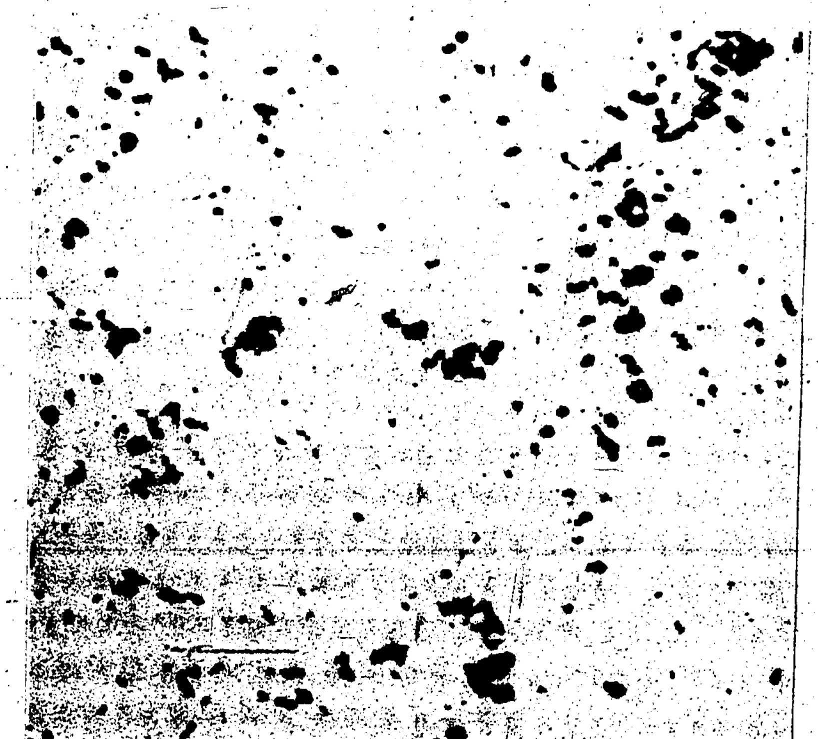

Fig. 57. Binary change picture of Edmonton using bands 4 and 5.

The number of pixels which changed is also given.

Table XX. Means of Change and No-Change Gray Levels.

Name	No-change Ave	Change Ave	Pixels Not Changed	Pixels Changed	Pixels Total
1. Lake	10.51	28.45	63567	1969	65536
2. Edmonton					
Band4	10.21	20.56	240064	10937	251001
Band5	19.87	33.59	240064	10937	251001
Band6	25.18	56.02	238072	12929	251001
Band7	22.58	68.38	238072	12929	251001
3. Aerial	6.17	21.39	81239	31161	112400
4. Cut	6.94	20.8	122231	3209	125440
5. Fertilizer					
N. N.	5.13	20.92	129548	17452	147000
Bilinear	5.06	21.64	130736	16264	147000
Cubic	4.75	23.60	131582	15417	147000

In the second test case, all four bands of the two pictures were registered. The band 5 pictures are shown in Fig's 29 and 39. The four bands of the first registered picture were then subtracted pixel by pixel, from the second registered picture to yield four difference pictures. The histograms of the difference pictures gave a threshold of T=15 for bands 4 and 5 and a threshold of T=53 for bands 6 and 7. The difference pictures were then thresholded to give binary pictures.

Next, another binary picture was generated pixel by pixel by a logical 'or' of bands 4 and 5, e.g., if either of the pixels of band 4 or band 5 has a gray level 1, the new picture will also have a gray level 1. In the similar manner, another binary picture was generated by the logical 'or' of pixels of bands 6 and 7. This processing was done, because bands 4 and 5 and bands 6 and 7 are very close together, both visually as well as statistically. Although, some features are differentiated by each band separately, in this study this pcint will not be considered.

The two new binary pictures were then cleaned by using a 7*7 operator. The resulting pictures are shown in Fig's 57 and 58. In order to point out the land use changes, the pictures were overlayed with the registered picture and are shown in Fig's 59 and 60. Visually, the urban growth in the Edmonton area can be seen easily. The new developing areas of Castle Downs, Mill Woods, St. Albert, Carleton Estate, Albert Park, Thorncliffe, Patricia Hts., are pointed out in Fig. 59. Rural changes such as cutting of crops, etc., are evident in Fig. 60. It should be noted that bands 4 and 5 emphasize land use changes and bands 6 and 7 emphasize forestry and rural type of changes. Empirically, the mean change deviation and mean no-change deviation in gray levels for all bands is given in Table XI (picture name Edmonton). It should be noted that the cleaned binary picture (of bands 4 and 5) was used as a mask for both bands 4 and 5.

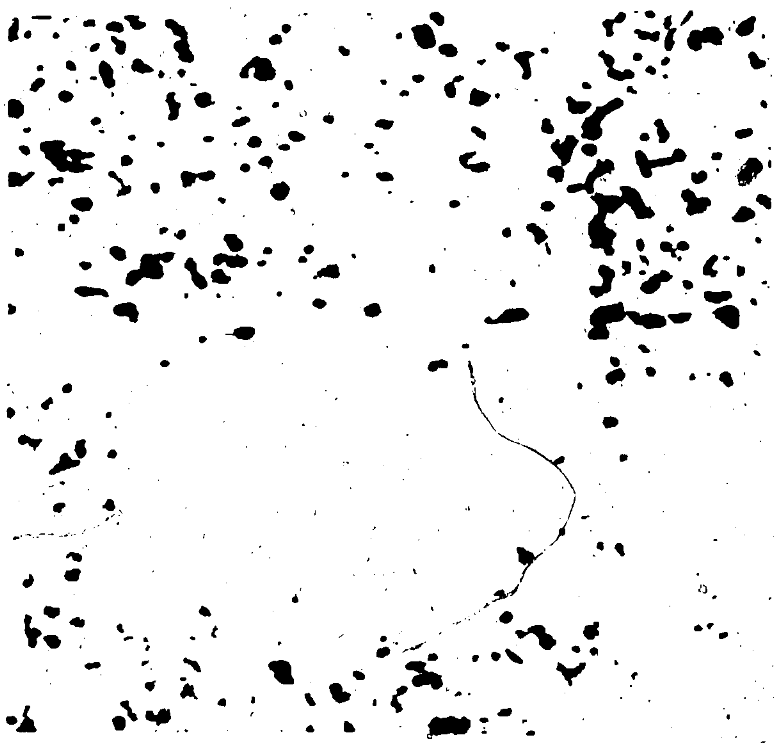

Fig. 58. Binary change picture of Edmonton using bands 6 and 7.

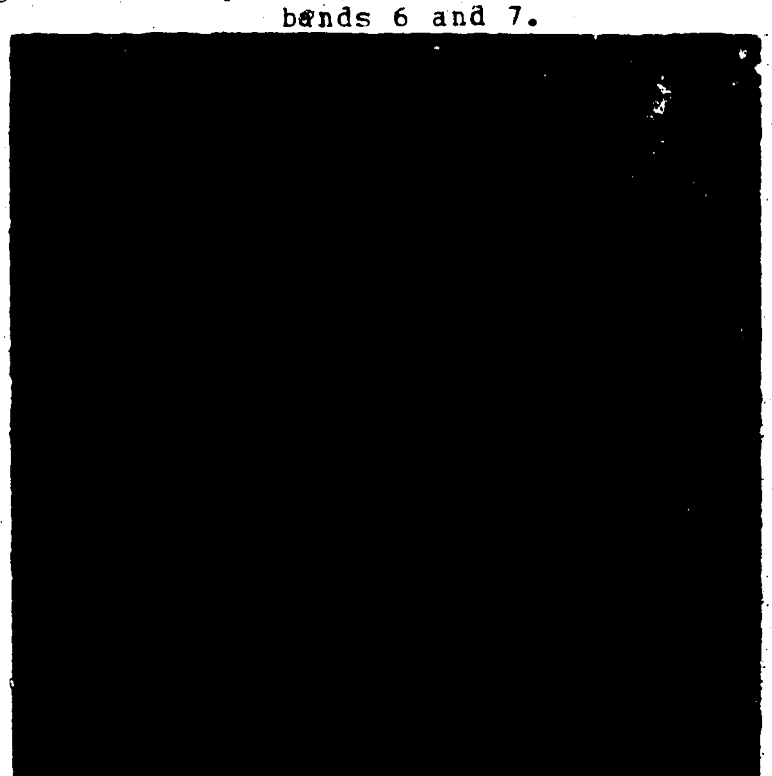

Fig 59. Overlaid change picture of Edmonton II using bands 4 and 5.

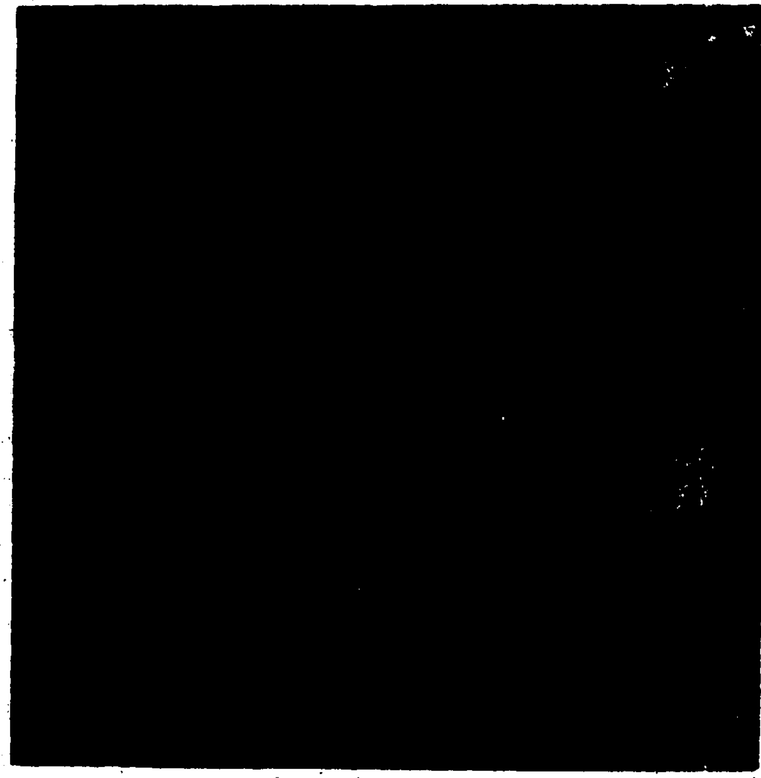

Fig. 60. Overlaid change picture of Edmonton II using bands 6 and 7.

Fig. 61. Window of Edmonton II with Gaussian histogram.

Similarly, the other cleaned binary picture was used for both bands 6 and 7.

In the third test case, the registered pictures as shown in Fig's 45 and 50, were Gaussian normalized. The purpose of this test was to check the registra' .on accuracy of pictures. The resulting pictures as shown in Fig's 61 and 62, were then subtracted pixel by pixel. From the histogram of the difference picture, the threshold was chosen to be 224. The binary change picture was obtained by thresholding the difference picture at 224. The binary change picture as shown in Fig. 63, show large changes in four years. The registration accuracy was found to be reasonable, i.e., within 1 to 2 pixels. The mean change deviation and mean no-change deviation in gray levels is given in Table XX (picture name Aerial).

In the fourth test case, the registered pictures as shown in Fig's 42 and 41, were Gaussian normalized. The resulting pictures as shown in Fig's 64 and 65, were then subtracted pixel by pixel to generate a difference picture. From the histogram of the difference picture, the threshold was chosen to be 234. The difference picture was then thresholded at 234 to give a binary change picture. The binary picture was cleaned using a 3*3 operator to yield the final binary change picture, as shown in Fig. 66. Visually, the small changes such as cutting of forests were detected easily. It should be pointed out that cleaning does

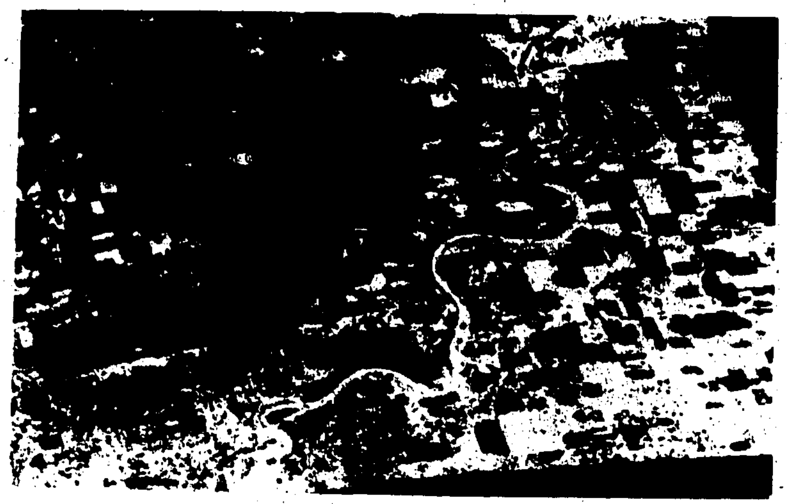

Fig. 62. Aerial picture (Gaussian) registered with Edmonton II (Gaussian) using proposed approximation.

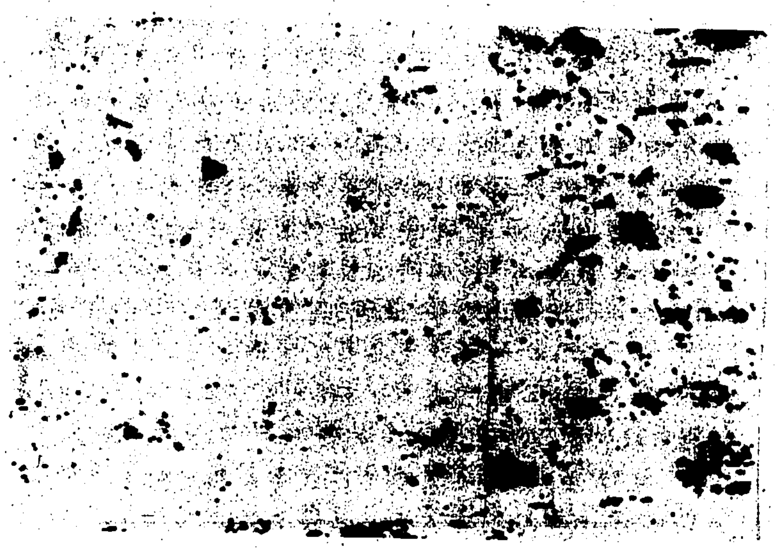

Fig. 63. Binary change picture of aerial and Edmonton II.

Fig. 64. Forestry Cut I (Gaussian) registered with Forestry Cut II (Gaussian) using nearest neighbor approximation.

Fig. 65. Forestry Cut II with Gaussian histogram.

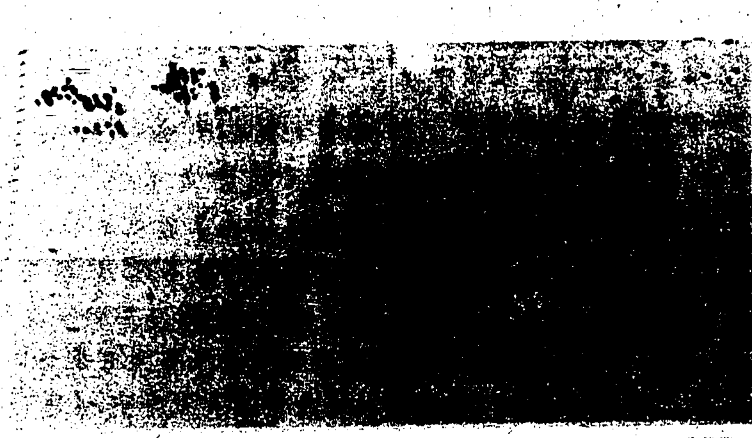

Fig. 66. Binary change picture of cut area.

in fact alter the shape of the detected changes. The mean change deviation and mean no-change deviation in gray levels is given in Table XX (picture name Cut).

In the final test case, the Fertilizer 2 picture and the registered Fertilizer 3 pictures, as shown in Fig's 31, 36-38, were Gaussian normalized. The resulting pictures are shown in Fig's 67-70. The normalized Fertilizer 2 picture was then subtracted from the three normalized registered pictures to give three difference pictures. From the histograms of the difference pictures, the threshold was chosen to be 237. The difference pictures were then thresholded at 237 to give three binary pictures. These pictures were cleaned using a 3*3 operator to give the final change pictures, as shown in Fig's 71-73.

Visually, the small changes detected are good. Since the pictures have texture, the difference pictures and subsequently the binary pictures, show spikes of noise. It should be noted that texture was not taken into account for detecting changes. Thus this test case raises the question of using texture information in detecting changes for texture oriented pictures. Also, the change picture obtained through the cubic convolution interpolation is much better than that obtained by the nearest neighbor approximation. The mean change deviation and mean no-change deviation in gray levels for both the nearest neighbor and bilinear approximation is given in Table XX (picture name Fertilizer).

Fig. 67. Fertilizer 2 with Gaussian histogram.

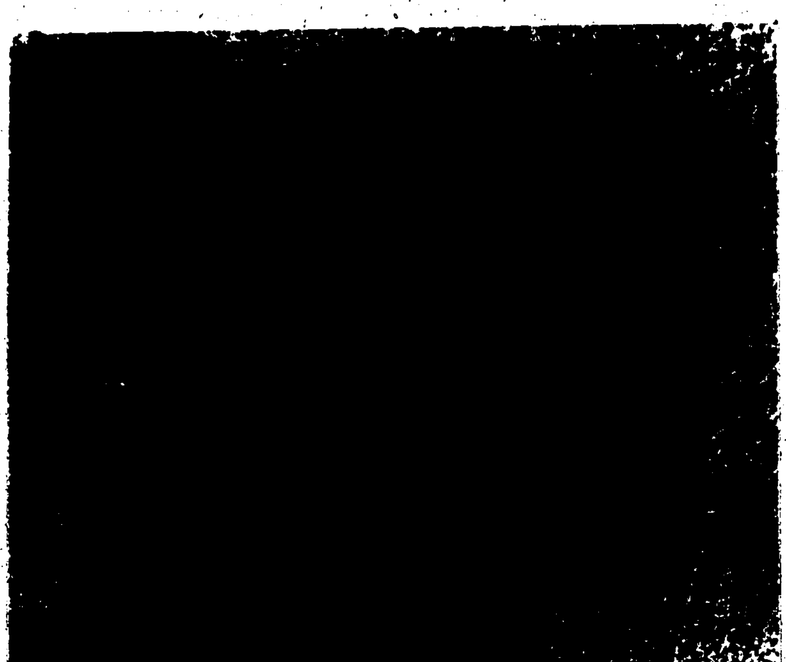

Fig. 68. Fertilizer 3 (Gaussian) registered with Fertilizer 2 (Gaussian) using nearest neighbor approximation.

Fig. 69. Fertilizer 3 (Gaussian) registered with Fertilizer 2 (Gaussian) using bilinear approximation.

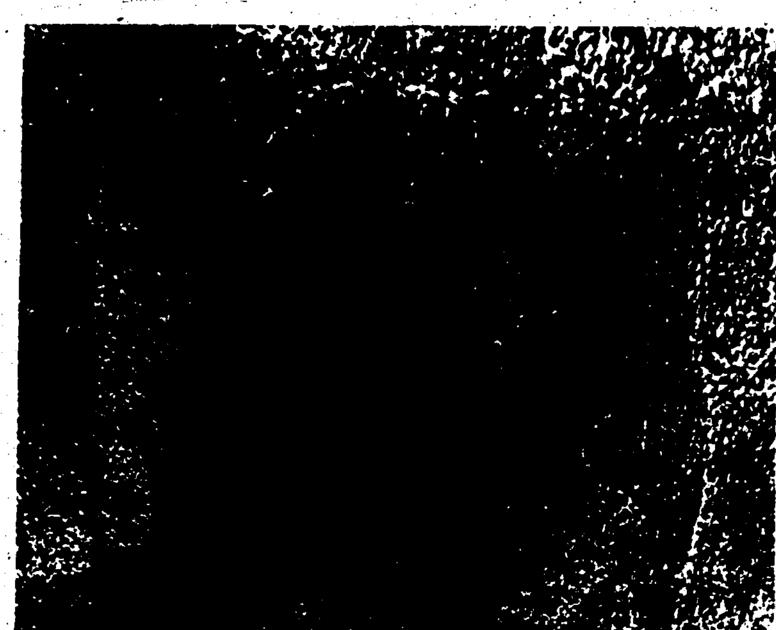

Fig. 70. Fertilizer 3 (Gaussian) registered with Fertilizer 2 (Gaussian) using cubic convolution.

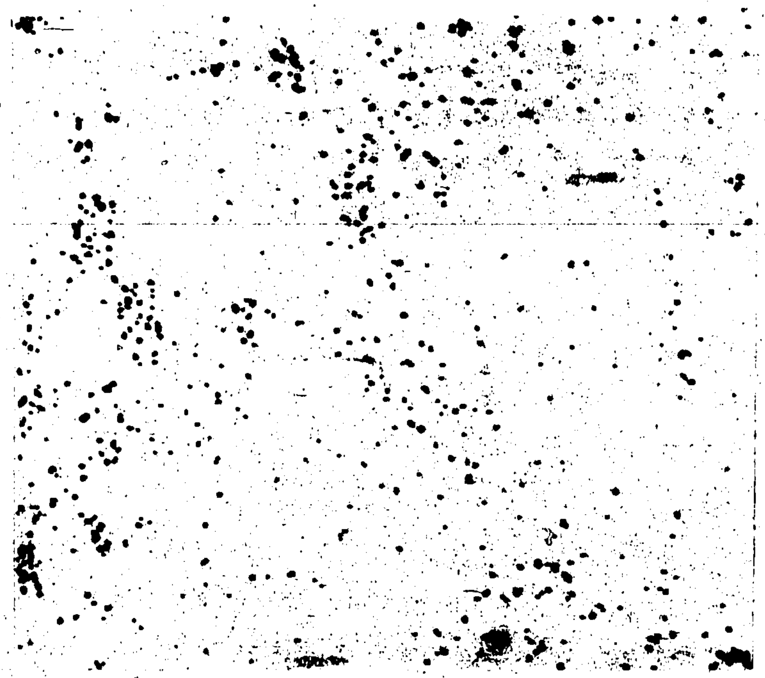

Fig. 71. Binary change picture of Fertilizer area using nearest neighbor approximation.

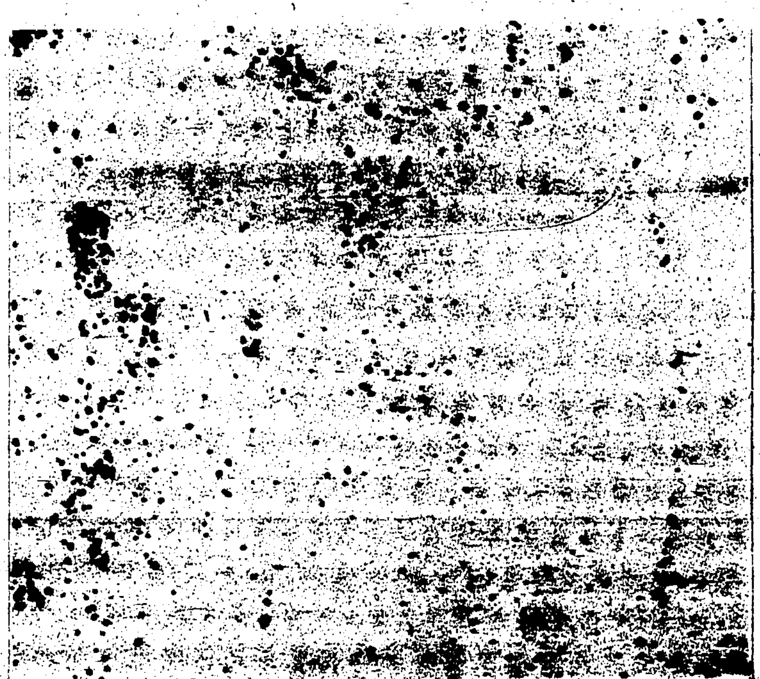

Fig. 72. Binary change picture of Fertilizer area using bilinear approximation.

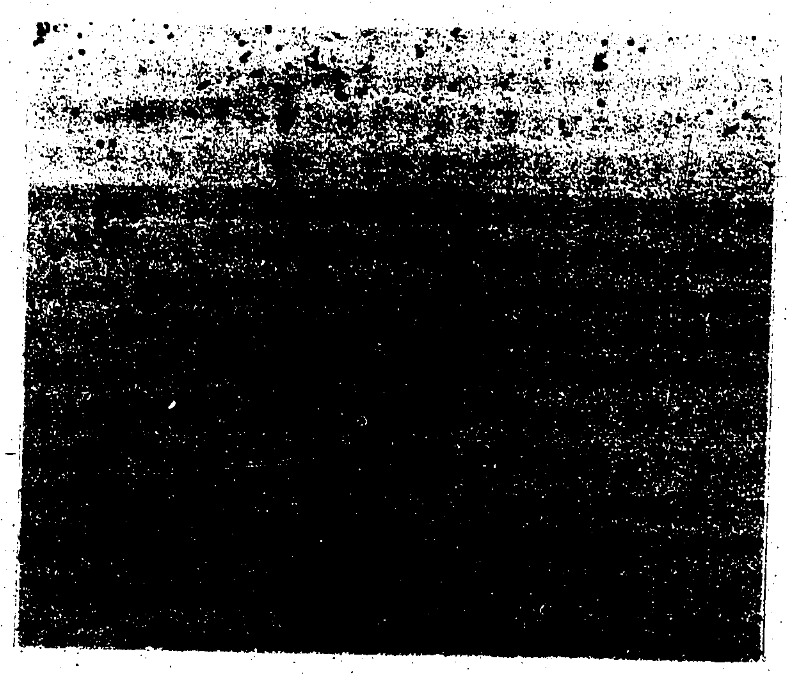

Fig. 73. Binary change picture of Fertilizer area using cubic convolution.

5.4 Discussion

This chapter is concerned with detecting various types of changes in scenes. The changes are classified into different categories. An algorithm for a two dimensional histogram is given for computing the deviation in gray levels from one registered picture to the another. Five sets of registered pictures are considered for identifying various kinds of changes. The pictures are Gaussian normalized and subtracted pixel by pixel. The difference pictures are then thresholded to yield binary change pictures. Isolated change pixels are removed by a cleaning operation. The changes identified are the surface of a lake during fall and winter seasons, forestry cut area, fertilizer area, and rural and urban land use. The mean change deviation for all the sets of pictures was found to be higher than the mean no-change deviation.

CHAPTER VI

CONCLUSIONS

The aim of this research was to study the problems of producing an automatic change detection system. This involved normalization, automatic control point selection, registration, template matching and detection of changes.

The picture normalization method has the advantage that it can map the original gray scale [a,b] to any desired range [c,d]. Normalization enhances and generates the desired standard pictures.

In normalization, averaging over the diamond neighborhood was done to compute the average gray level. The neighborhood size can be varied and local information can be used to obtain better normalized pictures.

The picture enhancement method sharpens pictures and the binary gradient pictures of the enhanced pictures have less noise and thin contour lines. The enhancement method was not used directly in detecting changes; however, it can be used in automatic control point selection. The original pictures can first be enhanced and then control points can be selected from the enhanced pictures.

In enhancement, the neighborhood size for defining binary relations, can be varied. The method can also be used for reducing the number of gray levels in the pictures to some desired number.

In automatic control point selection, control points were selected on the edges of the objects of the pictures. The problems encountered in their selection are presented and some basic features of control points are given.

If a measure for a good control point is defined, the measure may vary from picture to picture. Fixed templates or pre-specified objects such as lakes do not offer a general solution. The connected edge elements approach has been successful but some more work has to be done in this area. The pictures can also be averaged to reduce the complexity problem. The control points can then be found on reduced versions.

In the registration of pictures, the iterative feedback technique is fast and accurate, as it unifies the control point and the correlation approaches. The registration accuracy for the first order transformation was about 1 to 2 pixels and for the second order transformation was about 1 pixel. The cubic convolution technique of computing the gray level of the output pixel was better than the nearest neighbor approximation, as the picture generated by the nearest neighbor transformation was non-smooth.

For checking registration accuracy, either the complete pictures can be correlated or small corresponding windows can be correlated at some check points. The number of check points with a confidence interval of more than 95%

may be obtained by applying techniques of statistical sampling theory. It should be pointed out that the potential control points serve as good test points for checking registration accuracy.

In registration with large scale differences, registration of an aerial picture (2048*2048) with a LANDSAT picture (600*600) has been done. The transformed picture was generated by different approaches, e.g., nearest neighbor, averaging and the proposed approximation. The transformed picture obtained by the proposed technique is visually the best, as the new picture is smooth and noise free.

The proposed technique for large scale differences can also be improved by using either bilinear or spline interpolation, instead of zeroth order approximation. To be more accurate, the contribution of the irregular area within the boundary should also be taken into account. This can best be dealt with by spline functions.

In matching templates, the proposed algorithm was about 2 times faster than Barnea and Silverman's method and about 4 times faster than the standard correlation method. The number of operations in computing the normalized correlation coefficient are evaluated and faster algorithms to reduce the number of operations are reported.

The technique developed for fast template matching can also be improved. The values of the pre-specified constants k1,k2 and k3 should either be pre-computed or obtained through some measures. Also, faster algorithms for computing the correlation coefficient's term Sum x*y, should be developed.

In detecting changes between two registered pictures, various types of changes were detected. In the change detection procedure, stress has been laid on the binary change picture and the computation of two dimensional histograms. Because of fixed thresholding in the difference pictures, some information is lost. Techniques should be developed to filter out the information, which is indeed needed. Specific applications of the binary change detection technique include, urban development using visible wavebands (bands 4 and 5 of LANDSAT), forest fire detection systems where fires produce much stronger signals in the thermal infrared region of the spectrum and snow and ice surveys on rivers and lakes in areas where there is no urban development.

Also, due to cleaning and smoothing, some change pixels are lost. Even though, some isolated pixels may be valid change pixels, cleaning would change them to be no-change pixels. In this thesis, isolated pixels have not been considered to be valid. If due to cleaning, change pixels comprising the objects disappear, better techniques should be used to take care of this problem. Texture information was not used in this change detection study. The fertilizer pictures are texture oriented, which cause the binary change pictures to have many random noise points.

The role of texture in discriminating various types of changes should be studied.

REFERENCES

1. H.C. Andrews, Computer Techniques in Image Processing, Academic Press, New York, 1970.
2. H.C. Andrews and C.L. Patterson, Singular Value Decompositions and Digital Image Processing, IEEE Trans. Acous. Sp. and Signal Processing, Vol. ASSP-24, 1976, pp. 26-53.
3. H.C. Andrews, Digital Image Restoration: A Survey, Computer, Vol. 7, May 1974, pp. 36-45.
4. P.E. Anuta, Spatial Registration of Multispectral and Multitemporal Digital Images using FFT techniques, IEEE Trans. on Geoscience Electronics, Vol. GE-8, 1970, pp. 353-368.
5. P.E. Anuta, Digital Registration of Multispectral Video Imagery, SPIE Journal, Vol. 7, 1969, pp. 108-175.
6. D.I. Barnea and H.F. Silverman, A class of Algorithms for Fast Digital Image Registration, IEEE Trans. Computers, Vol. C-21, 1972, pp. 179-186.
7. Bendix Aerospace System Divison, Multispectral Picture Registration Study, Final Rept., BSR-2855, NASA Cont. NAS5-11699, 19[illegible], 159 pp.
8. R. Bernstein, Digital Image Processing of Earth

Observation Sensor Data, IBM Journal of Res. and Development, Vol. 20, Jan. 1976, pp. 40-57.

9. R. Bernstein and D.G. Ferneyhough, Digital Image Processing System, Photogrammetric Engineering and Remote Sensing, Vol. 41, 1975, pp. 1965.

10. F.C. Billingsley, Review of Digital Image Processing, EUROCOMP 75, London, Sept. 1975.

11. J.W. Christenson and H.M. Lachowski, Urban area delination and detection of Change along the urban-rural boundary as derived from LANDSAT digital data, ACSM-ASP Conference, 1976.

12. W.A. Davis, R.N. McPherson, M.W. Smith and N. Tsang, A System for processing of Digital Pictures, Proc. 1973 NRC Man-Computer Communications Conf., Ottawa, Canada, 1973.

13. W.F. Derouchie and R.B. Forrest, Potential positioning accuracy of ERTS1 MSS Imageing, ACSM-ASP Conference, Missouri, 1974.

14. S.D. De Gloria, S.J. Daus, N. Tosta and K. Bonner, First Linutilization of high altitude photography and LANDSAT-1 data for Change Detection and Sensitive area Analysis, 11th Int. Symp. on Remote Sensing of Environment, Michigan, 1975.

15. R. Eberlein, G.J. Vanderburg, A. Rosenfeld and L.S. Davis, Edge and Line Detection in the ERTS Imagery: a comparative study, Rept. TR-312, University of Maryland, June 1974.

16. R. Ellefson, Automatic Digital Processing of ERTS-1 MSS data in an Urban Land Use Mapping experiment, American Society of Photogrammerty meeting, St. Louis, Missouri, March 1974.

17. W. Feller, An introduction to Probability theory and its applications, Vol. I, John Wiley and Sons, New York, 1968.

18. E.A. Fleming, Positioning of Shore Features with the aid of LANDSAT imagery, Photogrammetric Engineering and Remote Sensing, Vol. 43, 1977, pp. 53-59.

19. A. Habibi and G.S. Robinson, A Survey of Digital Picture Coding, Computer, Vol. 7, May 1974, pp. 22-34.

20. E.L. Hall, Almost Uniform Distributions for Computer Image Enhancement, IEEE Trans. Computers, Vol. C-23, 1974, pp. 207-208.

21. D.J. Hall, R. Endlich, D.E. Wolf and A.E. Brain, Objective methods for Registering landmarks and determining cloud motions from satellite, IEEE Trans. Computer, Vol. C-21, 1972, pp. 768-777.

22. M.J. Hannah, Computer matching of Stereo Images, Ph.D thesis, Stanford University, 1974.

23. R.M. Haralick, K. Shanmugam and I. Dinstein, Textural features for Image classification, IEEE Trans. Sys. Man and Cybernetics, Vol. SMC-3, 1973, pp. 610-621.

24. R.M. Hoffer and the LARS staff, Techniques for Computer Aided Analysis of ERTS-1 data, useful in Geologic Forest and Water Resource Surveys, LARS Report: 121073,

Purdue University, 1973.

25. R.M. Hummel, Histogram Modification Techniques, Computer Graphics and Image Processing, Vol. 4, 1975, pp. 209-224.

26. B.R. Hunt, The Application of Constrained Least Squares Estimation to Image Restoration by Digital Computer, IEEE Trans. Computers, Vol. C-22, 1973, pp. 805-812.

27. R.R. Jayroe, J.F. Andrus and C.W. Campbell, Digital Image Registration method based on Binary Boundary Maps, NASA Tech. Note, TN D-7607, March 1974.

28. T. Kaneko, Evaluation of LANDSAT Image Registration Accuracy, Photogrammetric Engineering and Remote Sensing, Vol. XLII, 1976, pp. 1285-1299.

29. D.G. Kawamura, Automatic Recognition of Changes in Urban Development from Aerial Photographs, IEEE Trans. Sys. Man and Cybernetics, Vol. SMC-1, 1971, pp. 230-239.

30. M.D. Kelly, Edge Detection in Pictures by Computers using Planning, in Machine Intelligence, (B. Meltzer and D. Michie, Ed's.), Vol. 6, University Press, Edinburgh, 1971, pp. 379-409.

31. S.K. Kenue and W.A. Davis, Image Processing Software Routines and User's manual, Tech. Rept., Dept. of Computing Science, Univ. of Alberta, 1977.

32. S.K. Kenue and W.A. Davis, Preprocessing for Image Processing, Canadian Society for Computational Studies of Intelligence, Proc. of First CSCSI/SCEIO National

Conf., Univ. of British Columbia, Vancouver, B.C., August 1976, pp. 31-41.

33. H.P. Kramer and J.B. Bruckner, Iterations of a non-lineat Transformation for Enhancement of Digital Images, Pattern Recognition, Vol. 7, 1975, pp. 53-58.

34. R.P. Kruger, Computer Processing of Radiographic Images, Ph.D thesis, Univ. of Missouri-Columbia, June 1971.

35. R.L. Lillestrand, Techniques for Change Detection, IEEE Trans. Computers, Vol. C-21, 1972, pp. 654-659.

36. D.T. Lindgren, R.B. Simpson and W. Goldstein, Land use Change Detection in the Boston and New Haven areas, Tech. Rept.: Dept. of Geography, Dartmouth College Project in Remote Sensing, Jan. 1974.

37. J.A. Leese, C.S. Novak, and B.B. Clark, An Automated technique for obtaining cloud motion from geo-synchronous satellite data using cross correlation, Journal App. Meteorology, Vol. 10, 1971, pp. 118-132.

38. V.A. Milazzo, Some findings on the applications of ERTS and SkyLab Imagery for Metropolitan land Use Analysis, 9th Int. Symp. on Remote Sensing of Environment, Michigan, 1973.

39. R.N. Nagel and A. Rosenfeld, Ordered Search Techniques in Template Matching, Proc. IEEE, Vol. 60, 1972, pp. 242-244.

40. G. Nagy, Digital Image Processing activities in Remote Sensing for Earth Resources, Proc. IEEE, Vol. 60, 1972,

pp. 1177-1200.

41. M. Onoe and M. Saito, Automatic threshold setting for the SSDA, IEEE Trans. Computer, Vol. C-25, 1976, pp. 1052-1053.

42. F.G. Peet, A. Mack and L. Crosson, Affine Transformations from aerial photographs to Computer Compatiable Tapes, 3rd Earth Resources Technical Satellite Symp., NASA, SP-351, 1973, pp. 1719-1724.

43. F.G. Peet, The Use of Invariants in the Transformation and Registration of Images, The Canadian Surveyor, Vol. 29, 1975, pp. 501-506.

44. K. Price and R. Reddy, Change Detection in Multispectral Images, Dept. of Computer Science, Carnegie-Mellon University, Tech. Report, 1976.

45. L.H. Quam, Computer Comparison of Pictures, Ph.D thesis, Stanford University, May 1971.

46. S.S. Rifman, K.W. Simon, and R.H. Caron, Second generation digital techniques for processing LANDSAT MSS data, Earth Resources Survey Symp., Texas, June 1975.

47. S.S. Rifman, Evaluation of Digitally Corrected ERTS Imagery, Symp. on Management and Utilization of Remote Sensing Data, Sioux Falls, South Dakota, Oct. 20-Nov. 1973.

48. A. Rosenfeld, Picture Processing: 1974, Computer Graphics and Image Processing, Vol. 4, 1975, pp. 133-155.

49. A. Rosenfeld, Picture Processing: 1975, Computer Graphics and Image Processing, Vol. 5, 1976, pp. 215-237.

50. A. Rosenfeld, Picture Processing: 1976, Computer Graphics and Image Processing, Vol. 6, 1977, pp. 157-183.

51. A. Rosenfeld and A.C. Kak, Digital Picture Processing, Academic Press, New York, 1976.

52. A. Rosenfeld and J. Pfaltz, Distance Functions on Digital Pictures, Pattern Recognition, Vol. 1, 1968, pp. 33-63.

53. A. Rosenfeld, Automatic Detection of Changes in Reconnaissance Data, Conf. Proceeding of the 5th. Int. Convention on Military Electronics, 1961, pp. 492-499.

54. S. Shlein, Geometric Correction, Registration and Resampling of LANDSAT Images, Remote Sensing Science and Technology Symp., Ottawa, Canada, Feb. 21-23, 1977.

55. A.J. Scott, Automatic Registration of Imagery, Technical Report, Dept. of Geography, Carleton University, Ottawa, 1976.

56. E.A. Smith and D.R. Phillips, Automated Cloud Tracking using precisely Aligned Digital ATS pictures. IEEE Trans. Computers, Vol. C-21, 1972, pp. 715-729.

57. M.M. Sondhi, Image Restoration: The Removal of Spatially Invariant Degradation, Proc. IEEE, Vol. 60, 1972, pp. 842-853.

58. D. Steiner, Digital Geometric Picture Correction using a Piecewise Zero-order Transformation, Remote Sensing of Environment, Vol. 3, 1974, pp. 261-283.

59. J.E. Taber, Evaluation of Digitally Corrected ERTS Images, 3rd Earth Resources Tech. Satellite Symp., NASA, SP-351, 1973, pp. 1837-1843.

60. M. Tasto and V. Block, Locating Objects in Complex Scenes using a Spatial Distance Measure, Proc. of the Second Int. Joint Conf. Pattern Recognition, Copenhagen, 1974.

61. M.S. Ulstad, An Algorithm for estimating small scale differences between two Digital Images, Pattern Recognition, Vol. 5, 1973, pp. 323-333.

62. R. Welch and C.W. Pannell, LANDSAT investigations of recent Urban Land Use Changes in Northeast China, 10th Int. Symp. on Remote Sensing of Environment, Michigan, 1975.

63. J.S. Weszka, E.J. Carton, J.A. Verson, J.M. Mohr and A. Rosenfeld, Some Basic Edge Degradation and Enhancement techniques, Rept. TR-278, University of Maryland, Dec., 1973.

64. P.M. Will, R. Bakis, and M.A. Wesley, On an All Digital approach to Earth Resources Satellite Image Processing, IBM Research Report: RC 3027, New York, 1970.

65. P.A. Wintz, Transform Picture Coding, Proc. of IEEE, Vol. 60, 1972, pp. 809-872.

66. J.R. Wray, A Remote Sensing System for detecting Land

Use Change in Metropolitan areas, *7th Int. Symp. on Remote Sensing of Environment*, Michigan, 1971.

APPENDIX A

LIST OF SOFTWARE PROGRAMS

The user's manual and the source listing for FORTRAN programs used in various chapters is given in [31]. A list of programs with reference to each chapter [illegible] page number of the technical report [31] is give[illegible]

Name	Page	Chapter
AER.TRANSRC	6	III
AVE44	8	III
AVERAGE	9	II-V
BANDBIN	10	V
BILINEAR	11	III
BINARY	13	V
BYTE.ERTS	13	II-V
CLEAN	15	V
CONSEND	16	II-V
CORR.ERTS	17	III and IV
CPDUMP2	19	III
DIFERTS	21	V
DIFF	22	V
EDHIST	23	V
ENHANCE	24	II

APPENDIX B

SUMMARY OF PICTURE DETAILS

This appendix gives a summary of picture sizes, how they were obtained, filters used, output medium, etc. The Spinach and Neuron pictures were obtained by digitizing two covers from the IBM Journal of Research and Development. This was done by a TV camera, the output of which was converted to digital by a video quantizer [12]. For output, the digital data was then converted into an analog image by a 35mm camera and a Tektronix 611 storage scope operating in non-store mode. Other pictures were first produced on an electrostatic printer and then photographed to reduce the scale. Histograms of the pictures were drawn by a Calcomp plotter.

The LANDSAT pictures in the form of computer compatible tapes were obtained from the Canada Centre for Remote Sensing, Ottawa. The band numbers used are given in the figure description, together with the line and pixel numbers in the following format (first row, last row, first col, last col).

The aerial picture of Edmonton was obtained from the National Air Photo Library and has the following details:

NAPL Roll no: RSA 30376 Frame 3

Date: 5 Oct. 1971

Film: Kodak aerochrome IR

Filter: Pan 520

Altitude: 38600 feet

Scale: 1:12500

The fertilizer pictures were obtained from Dr. Jim Lee of the Pacific Forest Research Centre, and are of the Shawnigan Lake Fertilzer Plots on Vancouver Island, B. C. Both pictures were then digitized on a scanning microdensitometer using red, blue and green filters by the Canada Centre for Remote Sensing. Detailed information about the pictures along with other comments are listed as follows:

PICTURE DETAILS

PICTURE DETAILS

PICTURE DETAILS

PICTURE DETAILS

PICTURE DETAILS