1

Introduction

1.1 Introductory remarks

In considering image processing problems, it is commonly required that certain types of basic patterns be extracted from a noisy and/or complex scene. For digital image processing techniques considered in this book, one generally implies the processing of two-dimensional data that leads to applications in such a broad spectrum of problems as picture processing, medical data interpretation, underwater and earth sounding, trajectory detection, radargram enhancement, etc. If the noise conditions are favorable, involving high signal-to-noise ratios (SNR), and the corrupting noise is Gaussian with independent identically distributed (i.i.d.) samples (pixels), the classical techniques based on the matched filter theory are applicable. On the other hand, even a small deviation from the Gaussian assumption or the variability of SNR in various parts of the image will severely deteriorate the performance of the matched filter. In image formation by certain optical systems, such as infrared sensors and detectors, unfavorable noise conditions will prevail. As a matter of fact, the distribution function of the image pixels contaminated by noise is seldom known in imaging problems.

Another difficulty associated with processing of large two-dimensional images is that the SNR may vary significantly from region to region. In the latter situation the use of a simple thresholded matched filter will yield false alarm rates and probabilities of detection that will vary unpredictably over the scene. In addition, in analyzing a scene one may encounter unwanted background patterns (structured noise) that will further degrade the performance of a linear matched filter. If one uses a histogram of the image data from the entire scene to adjust the threshold, it may reduce the average false alarm rate, yet the probability of detection will vary from region to region, resulting in an unacceptable performance. In addition, the variable threshold approach will also require additional storage and an increase in processing time.

The difficulties associated with the use of matched filters in processing real images led to the development of techniques based on the analysis of variance

(ANOVA) in the framework of experimental designs, which result in simple nonlinear statistical operators (masks) easy to implement and adjust to various needs in processing digital images. For the case of testing a null hypothesis against alternatives for Gaussian noise of unknown variance, the F-statistic associated with the ANOVA model is optimum. As a matter of fact, the test involving the F-statistic maximizes the probability of detection for all alternatives and among all invariant tests with respect to shifting, scaling, and orthogonal transformation of the data. Thus, tests based on the F-statistics involving ANOVA are uniformly most powerful (UMP) in Gaussian noise (Lehmann[1], p. 67). Even if the noise variance is unknown and/or variable throughout the image, the probability of detection remains high even for a fixed threshold operation. Based on many years of experience with various imaging data, it was found that a small mask (5×5 or 3×7) is sufficient to obtain excellent results in applying ANOVA techniques to image processing. Also, this type of processing may be adjusted to suppress naturally structural noise.

If the noise deviates from Gaussian, the F-statistic remains robust.[1] It has been the experience of the senior author that contamination of Gaussian noise up to ten percent does not require any adjustment in the processing of imaging data. Hampel[84] indicates that the ANOVA model is sensitive to noise contamination. This may be true for other types of data and for imaging data with larger than ten percent rate of contamination. A procedure suggested by Hampel[84] to robustize the ANOVA model is impractical in imaging problems. If the imaging data are not statistically independent, the Hampel approach breaks down. It was indicated by Box and Anderson[79] that if the kurtosis $K_u = (E(x-\mu)^4/\sigma^4)-3$ of the Gaussian noise is between ±1, which corresponds to a peaked probability density function (pdf) with long and thin tails for positive kurtosis and flat pdf that drops off rapidly to zero for negative kurtosis, the F-statistic remains robust. This is also true if the skewness $s = (E(x-\mu)^3/\sigma^3)$ is positive and less than unity (long tail to the right). If the contamination of noise is severe, or if the noise is non-Gaussian, one needs to use a robustized version of the ANOVA model estimators. From a practical point of view, estimators based on stochastic approximation algorithms are useful. An exposition of the stochastic approximation methodology is given in Chapter 8. The chapter may also introduce the reader to the stochastic approximation methodology if he or she is not familiar with it.

An alternate approach to processing of imaging data in non-Gaussian noise involves the application of partition tests. Readers interested in this methodology are referred to Kurz[72] and Kersten and Kurz[73] and numerous other references on the subject.[86, 87] Since standard ANOVA theory (independent sampling) is extremely

[1]By robustness, we mean low sensitivity to deviations from a nominal distribution. The concept was first introduced by Box.[78]

sensitive to the dependence of noise samples, a generalization of the model to include statistical dependence is presented in the book. This complicates some of the derivations and expressions, yet the image processing schemes remain simple. The assumption of dependent sampling is realistic if one considers the fast sampling rate in modern data processing systems. If the dependent noise samples are contaminated by non-Gaussian noise, the stochastic approximation methodology may be extended as suggested by Lomp[82] and Kowalski.[83] Alternately, one may use generalized versions of partition tests.[74, 77, 82, 83]

1.2 Decision mechanism in image processing

From a practical point of view the most important aspect of the methodology based on the ANOVA model is the use of small local operators (masks). These masks must be large enough to ensure high probability of detection, yet small enough not to miss some intricate aspects of the image patterns. As indicated before, many years of experience with imaging data suggest 5×5 and 3×7 masks are useful for most applications, though nonrectangular masks may be used in special applications. The specific type of ANOVA design will be determined by specific applications. This statement will be clarified in subsequent chapters. The choice of the masks in conjunction with the use of computers leads to simple calculations, minimal storage for single local masks, and significant reduction in processing time, which in most cases leads to efficient real-time processing. The process of scanning the image by a mask and calculating the test statistic is analogous to a nonlinear time convolution between a signal function and the impulse response of a matched filter. Depending on the type of mask, a number is generated that is compared to a preselected threshold resulting in acceptance or rejection of the point as belonging to the image feature under study, that is, line or edge element. In evaluating the performance of each procedure, subjective and objective measures are used. The subjective evaluation of the procedures is based on numerous simulation results involving known images and various types of noise. In addition, objective evaluations of the various procedures are accomplished using Monte-Carlo simulations to obtain plots of the probability of detection as a function of correlation coefficients for fixed values of false alarm rates. It is interesting to note that all curves of this type are of V-form, with the dip occurring in the midrange of the correlation coefficient. The deterioration of performance in this region results from false patterns being generated by the dependence of noise. It should be noted that even this dip in performance still yields acceptable results.

It is recommended that the reader becomes familiar with the various algorithms by applying the available software (*SIMEX*) to imaging data. The software then may be used to expand the algorithms to include special variations, modifications, and additions of interest to the user.

1.2.1 Organization of the book

The book is presented in nine chapters. After having been introduced to the motivating concepts in this chapter, the reader may proceed to Chapter 2, which presents the basic mathematical notions of the ANOVA model. First, the problem of parameter estimation and linear hypothesis testing is reviewed. This is followed by a detailed presentation of the one- and two-way designs of the ANOVA model. The mathematical developments are related to the physical features of images such as column and row orientations. The two basic designs are then followed by the description of specialized designs such as Latin squares (LS), Græco-Latin squares (GLS), and symmetric balanced incomplete blocks (SBIB). To provide means for testing the validity of the results that are based on multicomparisons, confidence intervals on which to base our decisions are provided. Fundamental to this approach is the concept of contrast functions, which leads naturally to three important multiple comparison techniques: Tukey method, Scheffe method, and Bonferroni method. Because the Scheffe method has an interesting interpretation that if a test based on the F-statistic is rejected for a confidence level α, then a test based on the contrast function will provide means to locate the specific effect for which the effect is rejected, the latter property makes the Scheffe method particularly attractive in image processing applications. In Chapter 2, whenever the proofs are omitted, the reader is referred to Scheffe[2] for proofs involving contrast functions and related topics. It should be noted that unlike the mathematical literature, which treats only the statistically independent case, the theory presented in this chapter is extended to include dependent noise.

The problem of line detection is the basis for the material presented in Chapter 3. Techniques for the detection of lines of various orientations are presented. Specifically ANOVA techniques, starting with the unidirectional detector, which is capable of discriminating only in one direction at a time, are presented. To sharpen the line discrimination capabilities of the procedure, a shape statistic is introduced based on the Scheffe multicomparison test. The process of line detection is then generalized to include a two-way ANOVA model. Next, the GLS designs permit the detection of lines in four directions simultaneously. These algorithms yield high probability of detection in extracting line features at the expense of increase in computational complexity. The GLS design also permits removal of background interference (structural noise). The SBIB design and Youden square design are applied in conjunction with a double-edge transition type algorithm to detect and locate trajectories in noisy scenes. Various simulations as well as Monte-Carlo evaluations of the procedures in this chapter are presented.

The edge detection problem is considered in Chapter 4 (see references [13], [19]). Like the line problem of Chapter 3, the tests for statistical significance are performed in two steps. First, the test statistic is formed to estimate the pertinent parameters for a given ANOVA model. This is followed by a hypothesis test to detect the presence or absence of a particular edge element associated with a preselected direction. Both gray level and texture edges are considered. At first, uni- and bidirectional edge detectors are described followed by multidirectional edge detectors. In areas where edges are corrupted beyond recognition, edge reconstruction algorithms based on directional masks are introduced by using a gradient operation in the context of a GLS mask. The operation of the texture edge detectors is formulated in terms of hypothesis testing of the model parameter variance. Finally, as in the line detection problem, simulations of the various procedures are included.

In Chapter 5 the approaches presented in Chapters 3 and 4 are extended to include radial processing of imaging data. Radial processing results in higher power (improved detectability) in most image processing operations, accompanied by a small cost in increased processing time. In all other aspects the developments in this chapter parallel the material in Chapters 3 and 4.

The important problem of object detection is considered in Chapter 6. The concept of visual equivalence within the framework of standard arrays leads directly to a transformation-based object detection procedure. The procedure is then generalized to include correlated data. Permutation matrices are used to recover the transformation of elements location between original and transformed data spaces. The pertinent test statistics are based on the two-way layout parameterization in modeling the data. Since for large objects inversion of large matrices would be required, an alternate contrast function procedure defined in terms of the effects in the target and background arrays is used. It is applied together with the column and row statistics to test for heterogeneity of the global array. Next, a contrast-based procedure in which the object to be detected is partitioned into background and target arrays is introduced. Elementary contrasts assigned to each array of the partition are then used to generate a contrast function representing a linear combination of these elementary contrasts. The procedure is then generalized involving the use of orthogonal contrasts. This allows for better discrimination among similar objects. In the final part of this chapter a rotation invariant procedure for object detection is introduced. A complex comformal mapping process is used to reduce rotation to translation, which is followed by any of the detection procedures described above. The procedure also allows us to detect the angle between the test and reference objects. Considering that the process is accomplished in the presence of noise, this represents a significant improvement over existing procedures where noise-free or nearly noise-free conditions are assumed.

In many problems involving image data it is necessary to partition the image into various regions to separate features from the background. This partition process is known as image segmentation and is addressed in Chapter 7. The segmentation procedure is based on a split-merge procedure. In general, image elements are merged sequentially to form larger regions. The process is started with a hierarchical decomposition of the image. Using a nested design approach, regions are merged if they satisfy similarity measures defined in terms of an F-test on region effects.

In Chapter 8 certain complementary mathematical and algorithmic concepts are introduced to aid the reader in implementing the procedures presented in previous chapters. In the first part of the chapter the properties of the F-statistic are discussed. This is followed by the description of the Monte-Carlo simulations needed to generate power (probability of detection) for various procedures. Next, an introductory exposition of the stochastic approximation methodology is presented. Robustized versions of the algorithms are also included. This section may be of interest not only for image processing problems but also for many other estimation problems in engineering and science.

Finally, the material on stochastic approximation introduced in Chapter 8 is applied in Chapter 9 to the image restoration problem. Two classes of restoration procedures based on a two-stage approach are introduced. The first stage consists of an edge detector of GLS or Youden square type. The edge detection preprocessor is then followed by a recursive least square estimator described in Chapter 8. If the computing noise is of high variance, a robustized version of the recursive estimator is used. If the corrupting noise is of the "salt and pepper" type, the second stage of the restoration process is based on the "missing value" approach to the restoration process.

2

Statistical linear models

2.1 Introductory remarks

The problem of the application of the linear model in image processing, as developed by Aron and Kurz,[8] involves the interpretation of the experimental data in terms of effects or treatments. Thus, the initial stage is always the selection of the important features, that is, factors to be taken into account and eventually interpreted following the results of any statistical test based on the linear model. The next stage is the introduction of the hypotheses to be tested based on the model that best fits the objectives, the selected factors, and the available data. Finally, the importance to be attached to the eventual results by means of confidence intervals is delineated.

Test statistics based on the theory of the ANOVA within the context of experimental design have been shown to maximize power for all alternatives among all invariant tests with respect to shifting, scaling, and orthogonal transformation of the data.[1] Used in this context, they are generally referred to as UMP.

Several books are devoted to the subject of the ANOVA and it would be rather meaningless to dwell on the theory in this chapter. Instead, we concentrate on the subject of applying ANOVA in image processing problems by providing simple steps to be followed to extract meaningful information from the available data.

We first describe the general model along with the parameter estimation problem that arises in dealing with ANOVA. We then specialize the results to the one-way and two-way layouts. Also included is a detailed analysis of other commonly used designs, namely, Latin squares, Græco-Latin squares, and the balanced incomplete block designs. Next, we review the techniques that are available in the area of multicomparisons among effects. Finally, a detailed analysis of Scheffe's multicomparison technique is presented.

For an easy introduction to the linear model approach the reader may wish to begin with the excellent book by Neter and Wasserman.[36] Besides numerous examples showing the application of ANOVA to various fields, it also contains descriptions of most of the ANOVA models discussed in the literature.

2.2 Linear models

2.2.1 Parameter estimation

The statistical method of analysis of variance is based on the theory of the general linear model. Each collected observation is partitioned into essentially two basic components. The first term is due to causes that are of interest in the study. These causes can be either controlled or measured and represent physical features that are termed "factors" in the experiment. In image processing applications we always assume that the factor effects have already taken place; therefore, any further control during the processing is precluded. The second term in the partition is the random variation or noise component, essentially due to measurement error and any extraneous effect on the experiment.

Data in most cases are collected with a scanning window (mask) with typical size of 5×5. The number of factors that can be assumed depends to a great extent on the objectives to be accomplished. In addition, factors are chosen to represent physical features within the window. There exist two broad classes of models within ANOVA based on the nature of the assumed factors: fixed effects models and random effects models. Before we explain the differences between the models, it is important to keep in mind that a factor in the experiment, which is represented by levels, corresponds to a physical feature that we are interested in detecting by a statistical test on the levels. Hence, it is important to understand the nature of the levels that represent the factor of interest.

When the levels are chosen from a much larger set of levels, the model is referred to as a random effects model. On the other hand, if the levels are specifically selected without being drawn from a population, the model is referred to as a fixed effects model.

The vector space Ω, in which the mean of the observed n-dimensional data vector $\mathbf{y}$ lies, is described by the following model

$$\Omega : \mathbf{y} = \mathbf{X}^T \boldsymbol{\beta} + \mathbf{e} \tag{2.1}$$

where $\boldsymbol{\beta}$ is a p-dimensional vector of unknown parameters to be estimated, $\mathbf{X}^T$ is an $n \times p$ matrix with $p < n$ and is referred to in the statistical literature as the design matrix. The n-dimensional vector $\mathbf{e}$ is the noise vector with the characteristics

$$E(\mathbf{e}) = \mathbf{0} \tag{2.2}$$

and

$$E(\mathbf{e}\mathbf{e}^T) = \sigma^2 \mathbf{K}_f \tag{2.3}$$

where σ^2 is the unknown variance and $\mathbf{K}_f$ is the correlation matrix.[1] For the independent data case, $\mathbf{K}_f$ reduces to the identity matrix. The space Ω also represents the parameter space of $\boldsymbol{\beta}$, which is transformable to the space of the mean of $\mathbf{y}$ by $\mathbf{X}^T$. Thus,

$$\Omega : \begin{cases} E(\mathbf{y}) = \mathbf{X}^T \boldsymbol{\beta} \\ E\left((\mathbf{y} - \mathbf{X}^T \boldsymbol{\beta})(\mathbf{y} - \mathbf{X}^T \boldsymbol{\beta})^T\right) = \sigma^2 \mathbf{K}_f \end{cases} \tag{2.4}$$

Central to our discussion is the vector of parameters $\boldsymbol{\beta}$ that is ordinarily unknown and must be estimated from the available data. An estimate of $\boldsymbol{\beta}$ can be found using the least squares approach or, in the case of Gaussian noise models, by the maximum likelihood method.

First, we consider the independent data case where the correlation matrix $\mathbf{K}_f$ is the identity matrix. Given Eq. (2.1), the least squares estimate (LSE) of $\boldsymbol{\beta}$, which is denoted by $\widehat{\beta}$, is obtained by the well-known normal equation

$$\mathbf{X}\mathbf{X}^T \widehat{\beta} = \mathbf{X}\mathbf{y} \tag{2.5}$$

Here $\widehat{\beta}$ is not unique unless the p column vectors of $\mathbf{X}^T$ are linearly independent or rank $(\mathbf{X}^T) = p$. In case independence is not satisfied, we must place linearly independent side conditions on the parameter vector $\boldsymbol{\beta}$ in the form

$$\mathbf{B}^T \boldsymbol{\beta} = \mathbf{0} \tag{2.6}$$

The model (2.1) is thus modified to reflect the additional assumption, in which case we have

$$\begin{pmatrix} \mathbf{X}^T \\ \mathbf{B}^T \end{pmatrix} \widehat{\beta} = \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} \tag{2.7}$$

The parameter vector estimate can now be calculated and is given by

$$\widehat{\beta} = (\mathbf{X}\mathbf{X}^T + \mathbf{B}\mathbf{B}^T)^{-1} \mathbf{X}\mathbf{y} \tag{2.8}$$

where the rank of $\mathbf{X}^T$ is p and, thus, $\mathbf{X}\mathbf{X}^T$ is invertible.

When the data are dependent, Eq. (2.8) is no longer valid, and we must derive another estimator based on the dependency of the data as expressed by the correlation matrix $\mathbf{K}_f$. The usual approach in similar cases is to proceed with a prewhitening

[1] The mathematical literature treats only the statistically independent case.

transformation so that the transformed data is uncorrelated. This approach has been suggested in references [2], [10], [14], [70].

Consider the correlation matrix $\mathbf{K}_f$;[2] we can always find a nonsingular $n \times n$ matrix $\mathbf{P}$ such that the condition $\mathbf{P}^T\mathbf{K}_f\mathbf{P} = \mathbf{I}$ is satisfied. For the purpose of finding the estimate of the effect vector, we do not normally need to find $\mathbf{P}$ but instead we proceed by deriving the expression for the sum of squares using the least squares approach. Consider the vector of orthogonal data obtained by using the transformation $\mathbf{P}$ on the correlated data. We can write

$$\tilde{\mathbf{y}} = \mathbf{P}^T\mathbf{y} \tag{2.9}$$

and the correlation matrix of the new vector of uncorrelated data is

$$E(\tilde{\mathbf{y}}\tilde{\mathbf{y}}^T) = \mathbf{P}^T E(\mathbf{y}\mathbf{y}^T)\mathbf{P} = \mathbf{P}^T\mathbf{K}_f\mathbf{P} = \mathbf{I} \tag{2.10}$$

The resulting error sum of squares for the dependent case is then derived by first considering the alternative for the orthogonal data

$$\Omega : \tilde{\mathbf{y}} \text{ is } N(\mathbf{P}^T\mathbf{X}^T\boldsymbol{\beta}, \sigma^2\mathbf{I}) \tag{2.11}$$

where the mean is obtained by using Eqs. (2.4) and (2.9). Thus, the error sum of squares is

$$SS_e(\mathbf{y}, \boldsymbol{\beta}) = (\tilde{\mathbf{y}} - \mathbf{P}^T\mathbf{X}^T\boldsymbol{\beta})^T(\tilde{\mathbf{y}} - \mathbf{P}^T\mathbf{X}^T\boldsymbol{\beta}) \tag{2.12}$$

Using Eq. (2.9), we have

$$SS_e(\mathbf{y}, \boldsymbol{\beta}) = (\mathbf{P}^T\mathbf{y} - \mathbf{P}^T\mathbf{X}^T\boldsymbol{\beta})^T(\mathbf{P}^T\mathbf{y} - \mathbf{P}^T\mathbf{X}^T\boldsymbol{\beta}) \tag{2.13}$$

By factoring out the terms in $\mathbf{P}$ and using the fact that $\mathbf{P}\mathbf{P}^T = \mathbf{K}_f{}^{-1}$, the resulting sum of squares is

$$SS_e(\mathbf{y}, \boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}^T\boldsymbol{\beta})^T\mathbf{K}_f{}^{-1}(\mathbf{y} - \mathbf{X}^T\boldsymbol{\beta}) \tag{2.14}$$

The calculation of the effect vector estimate can now be obtained by differentiating $SS_e(\mathbf{y}, \boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ using vector differentiation or any other approach that takes into account the specific nature of the correlation matrix.

2.2.2 *Linear hypothesis testing*

For the general linear model, the linear hypothesis to be tested is generally the test of the means of populations from which samples were taken. In the case of the

[2] $\mathbf{K}_f$ is assumed to be a symmetric positive definite matrix.