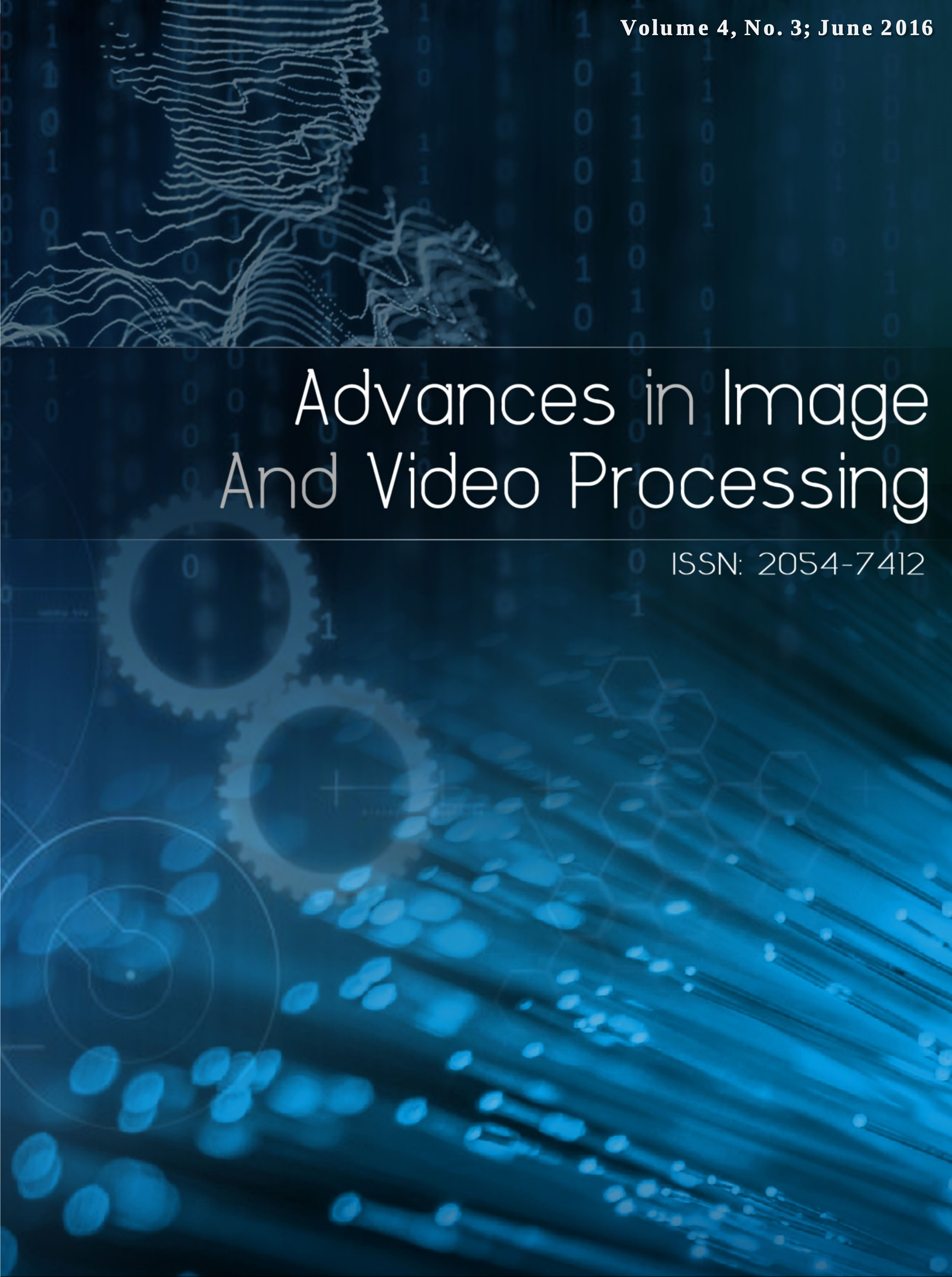
Volume 4, No. 3; June 2016



Advances in Image And Video Processing

ISSN: 2054-7412

Society for Science and Education - United Kingdom

TABLE OF CONTENTS

EDITORIAL ADVISORY BOARD

DISCLAIMER

Society for Science and Education United Kingdom
AIVP Advances in Image and Video Processing
Volume 4, Issue 3
ISSN 2054-7412

A New Approach to Constructing Tolerance Limits on Order Statistics in Future Samples Coming from a Normal Distribution

[1]Nicholas A. Nechval, [2]Konstantin N. Nechval, and [3]Vladimir F. Strelchonok
[1,3]Department of Mathematics, Baltic International Academy, Riga, Latvia;
[2]Department of Applied Mathematics, Transport and Telecommunication Institute, Riga, Latvia;
nechval@junik.lv; konstan@tsi.lv; str_riga@inbox.lv

ABSTRACT

Although the concept of statistical tolerance limits has been well recognized for long time, surprisingly, it seems that their applications remain still limited. Analytic formulas for the tolerance limits are available in only simple cases, for example, for the upper or lower tolerance limit for a univariate normal population. Thus it becomes necessary to use new or innovative approaches which will allow one to construct tolerance limits on future order statistics for many populations. In this paper, a new approach to constructing lower and upper tolerance limits on order statistics in future samples is proposed. Attention is restricted to invariant families of distributions under parametric uncertainty. The approach used here emphasizes pivotal quantities relevant for obtaining tolerance factors and is applicable whenever the statistical problem is invariant under a group of transformations that acts transitively on the parameter space. It does not require the construction of any tables and is applicable whether the past data are complete or Type II censored. The proposed approach requires a quantile of the *F* distribution and is conceptually simple and easy to use. For illustration, the normal distribution is considered. The discussion is restricted to one-sided tolerance limits. A practical example of finding a warranty assessment of image quality is given.

Keywords: Order Statistics, *F* Distribution, Lower Tolerance Limit, Upper Tolerance Limit, Normal Distribution.

1 Introduction

Statistical tolerance limits are an important tool often utilized in areas such as engineering, manufacturing, and quality control for making statistical inference on an unknown population. As opposed to a confidence limit that provides information concerning an unknown population parameter, a tolerance limit provides information on the entire population; to be specific, one-sided tolerance limit is expected to capture a certain proportion or more of the population, with a given confidence level. For example, an upper tolerance limit for a univariate population is such that with a given confidence level, a specified proportion or more of the population will fall below the limit. A lower tolerance limit satisfies similar conditions.

It is often desirable to have statistical tolerance limits available for the distributions used to describe time-to-failure data in reliability problems. For example, one might wish to know if at least a certain proportion, say β, of a manufactured product will operate at least *T* hours. This question can not usually be answered exactly, but it may be possible to determine a lower tolerance limit $L(X_1, \ldots, X_n)$,

DOI: 10.14738/aivp.43.2014
Publication Date: 23rd July, 2016
URL: http://dx.doi.org/10.14738/aivp.43.2014

based on a preliminary random sample (X_1, …, X_n), such that one can say with a certain confidence γ that at least 100β % of the product will operate longer than $L(X_1, …, X_n)$. Then reliability statements can be made based on $L(X_1, …, X_n)$, or, decisions can be reached by comparing $L(X_1, …, X_n)$ to T. Tolerance limits of the type mentioned above are considered in this paper. That is, if $f_\theta(x)$ denotes the density function of the parent population under consideration and if S is any statistic obtained from the preliminary random sample (X_1, …, X_n) of that population, then $L(S)$ is a lower γ probability tolerance limit for proportion β if

$$\Pr\left(\int_{L(S)}^{\infty} f_\theta(x)dx \geq \beta\right) = \gamma, \quad (1)$$

and $U(S)$ is an upper γ probability tolerance limit for proportion β if

$$\Pr\left(\int_{-\infty}^{U(S)} f_\theta(x)dx \geq \beta\right) = \gamma, \quad (2)$$

where θ is the parameter (in general, vector).

The common distributions used in life testing problems are the normal, exponential, Weibull, and gamma distributions [1]. Tolerance limits for the normal distribution have been considered in [2], [3], [4], and others.

Tolerance limits enjoy a fairly rich history in the literature and have a very important role in engineering and manufacturing applications. Patel [5] provides a review (which was fairly comprehensive at the time of publication) of tolerance limits for many distributions as well as a discussion of their relation with confidence intervals for percentiles and prediction intervals. Dunsmore [6] and Guenther, Patil, and Uppuluri [7] both discuss 2-parameter exponential tolerance intervals and the estimation procedure in greater detail. Engelhardt and Bain [8] discuss how to modify the formulas when dealing with type II censored data. Guenther [9] and Hahn and Meeker [10] discuss how one-sided tolerance limits can be used to obtain approximate two-sided tolerance intervals by applying Bonferroni's inequality. Tolerance limits on order statistics in future samples coming from a two-parameter exponential distribution have been considered in [11].

In contrast to other statistical limits commonly used for statistical inference, the tolerance limits (especially for the order statistics) are used relatively rarely. One reason is that the theoretical concept and computational complexity of the tolerance limits is significantly more difficult than that of the standard confidence and prediction limits. Thus it becomes necessary to use new or innovative approaches which will allow one to construct tolerance limits on future order statistics for many populations.

In this paper, a new approach to constructing lower and upper tolerance limits on order statistics in future samples is proposed. For illustration, the normal distribution is considered. It is a commonly used model in reliability and risk theory. Although the concept of statistical tolerance limits has been well recognized for long time, surprisingly, it seems that their applications remain still limited.

2 Mathematical Preliminaries

2.1 Probability Distribution Function of Order Statistic

Theorem 1. If there is a random sample of m ordered observations $Y_1 \leq \ldots \leq Y_m$ from a known distribution (continuous or discrete) with density function $f_\theta(y)$, distribution function $F_\theta(y)$, then the probability distribution function of the kth order statistic Y_k, $k \in \{1, 2, \ldots, m\}$, is given by

$$P_\theta(Y_k \leq y_k \mid m) = \int_{\frac{1-F_\theta(y_k)}{F_\theta(y_k)} \frac{2k}{2(m-k+1)}}^{\infty} f_{2(m-k+1),2k}(x)dx, \tag{3}$$

where

$$f_{2(m-k+1),2k}(x) = \frac{1}{\mathrm{B}\left(\frac{2(m-k+1)}{2}, \frac{2k}{2}\right)} \left(\frac{2(m-k+1)}{2k}\right)\left(\frac{2(m-k+1)}{2k}x\right)^{2(m-k+1)/2-1}$$

$$\times \left(1 + \frac{2(m-k+1)}{2k}x\right)^{-[2(m-k+1)+2k]/2}, \quad x > 0, \tag{4}$$

is the probability density function of an F distribution with $2(m-k+1)$ and $2k$ degrees of freedom.

Proof. Suppose an event occurs with probability p per trial. It is well-known that the probability P of its occurring k or more times in m trials is termed a cumulative binomial probability, and is related to the incomplete beta function $I_x(a, b)$ as follows:

$$P \equiv \sum_{j=k}^{m} \binom{m}{j} p^j (1-p)^{m-j} = I_p(k, m-k+1). \tag{5}$$

It follows from (5) that

$$P_\theta\{Y_k \leq y_k \mid m\} = \sum_{j=k}^{m} \binom{m}{j} [F_\theta(y_k)]^j [1 - F_\theta(y_k)]^{m-j} = I_{F_\theta(y_k)}(k, m-k+1)$$

$$= \frac{1}{\mathrm{B}(k, m-k+1)} \int_0^{F_\theta(y_k)} u^{k-1}(1-u)^{(m-k+1)-1} du = \frac{\left(\frac{2(m-k+1)}{2k}\right)^{2(m-k+1)/2}}{\mathrm{B}\left(\frac{2k}{2}, \frac{2(m-k+1)}{2}\right)} \int_0^{F_\theta(y_k)} u^{\frac{2(m-k+1)+2k}{2}}$$

$$\times \left(\frac{1-u}{u} \frac{2k}{2(m-k+1)}\right)^{2(m-k+1)/2-1} \frac{-2k}{2(m-k+1)}\left(-\frac{du}{u^2}\right)$$

$$= \frac{\left(\frac{2(m-k+1)}{2k}\right)^{2(m-k+1)/2}}{\mathrm{B}\left(\frac{2(m-k+1)}{2}, \frac{2k}{2}\right)} \int_{\frac{1-F_\theta(y_k)}{F_\theta(y_k)} \frac{2k}{2(m-k+1)}}^{\infty} x^{2(m-k+1)/2-1}\left(1 + \frac{2(m-k+1)}{2k}x\right)^{-[2(m-k+1)+2k]/2} dx, \tag{6}$$

where

$$x = \frac{1-u}{u}\frac{2k}{2(m-k+1)}. \tag{7}$$

This ends the proof.

Corollary 1.1.

$$P_\theta(Y_k > y_k \mid m) = 1 - P_\theta\{Y_k \le y_k \mid m\} = \int_0^{\frac{1-F_\theta(y_k)}{F_\theta(y_k)}\frac{2k}{2(m-k+1)}} f_{2(m-k+1),2k}(x)dx. \tag{8}$$

Corollary 1.2. If $y_{k,m;\gamma}$ is the quantile of order γ for the distribution of Y_k, we have from (8) that $y_{k,m;\gamma}$ is the solution of

$$F_\theta(y_{k,m;\gamma}) = \frac{k}{k+(m-k+1)q_{2(m-k+1),2k;1-\gamma}}, \tag{9}$$

where $q_{2(m-k+1),2k;1-\gamma}$ is the quantile of order $1-\gamma$ for the F distribution with $2(m-k+1)$ and $2k$ degrees of freedom.

2.2 Normal Distribution

The normal distribution is perhaps the most commonly used probability distribution in both statistical theory and applications. For example: 1) many classical statistical tests are based on the assumption that the data follow a normal distribution (this assumption should be tested before applying these tests); 2) in modeling applications, such as linear and non-linear regression, the error term is often assumed to follow a normal distribution with fixed location and scale; 3) the normal distribution is used to find significance levels in many hypothesis tests and confidence intervals.

Physical measurements in areas such as meteorological experiments, rainfall studies, and measurements of manufactured parts are often more than adequately explained with a normal distribution. In addition, errors in scientific measurements are extremely well approximated by a normal distribution.

In 1733, Abraham DeMoivre developed the mathematical equation of the normal curve. It provided a basis for which much of the theory of inductive statistics is founded. The normal distribution is often referred to as the Gaussian distribution, in honor of Karl Gauss (1777-1855), who also derived its equation from a study of errors in repeated measurements of the same quantity. Properties of the normal distribution have been well developed (e.g., see Johnson et al. [12], Patel and Read [13], Balakrishnan and Nevzorov {14], Kotz and Vicari [15].

The normal distribution plays a vital role in many applied problems of biology, economics, engineering, financial risk management, genetics, hydrology, mechanics, medicine, number theory, statistics, physics, psychology, reliability, etc., and has been extensively studied, both from theoretical and applications point of view, by many researchers, since its inception.

Thus, the normal distribution is a widely used and widely known distribution. It is characterized by the probability density function of a continuous random variable *Y*,

$$f_\theta(y) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right), \quad -\infty < y < \infty, \tag{10}$$

that is, $Y \sim N(\mu,\sigma^2)$, where $\theta = (\mu,\sigma)$, $-\infty < \mu < \infty$ is the location parameter and $\sigma > 0$ is the scale parameter. These parameters are assumed to be unknown. The cumulative distribution function of the normal distribution is given by

$$F_\theta(y) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{y} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) dy. \tag{11}$$

It is known (Nechval and Vasermanis [16]) that the complete sufficient statistic for the parameter θ, based on observations in a random sample $(X_1, \ldots, X_n)$ of size n from the normal distribution (10) is given by

$$S = \left(\bar{X} = \sum_{i=1}^{n} X_i / n,\; S_1^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 / (n-1)\right). \tag{12}$$

Here the following theorem takes place.

Theorem 2. Let $(X_1, \ldots, X_n)$ be a preliminary random sample from the normal distribution (10) , where it is assumed that the parameter $\theta = (\mu,\sigma)$ is unknown. Then the joint probability density function of the pivotal quantities,

$$V_1 = \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}, \quad V_2 = \frac{(n-1)S_1^2}{\sigma^2}, \tag{13}$$

is given by

$$f(v) = f_1(v_1) f_2(v_2), \tag{14}$$

where

$$V = (V_1, V_2), \tag{15}$$

$$f_1(v_1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{v_1^2}{2}\right), \quad -\infty < v_1 < \infty, \tag{16}$$

$$f_2(v_2) = \frac{1}{2^{(n-1)/2}\Gamma\left((n-1)/2\right)} v_2^{(n-1)/2-1} \exp(-v_2/2), \quad v_2 \geq 0. \tag{17}$$

Proof. The joint density of $X_1, \ldots, X_n$ is given by

$$f_\theta(x_1, \ldots, x_n) = \prod_{i=1}^{r} f_\theta(x_i) = \prod_{i=1}^{n} (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{1}{2\sigma^2}(x_i-\mu)^2\right)$$

$$= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right). \tag{18}$$

Using the invariant embedding technique [17]-[19], we transform (18) to

$$f_\theta(x_1, \ldots, x_n) d\bar{x} ds_1^2 = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \bar{x} + \bar{x} - \mu)^2\right) d\bar{x} ds_1^2$$

$$= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[(x_i-\bar{x})^2 + 2(x_i-\bar{x})(\bar{x}-\mu) + (\bar{x}-\mu)^2]\right) d\bar{x} ds_1^2$$

$$= (2\pi\sigma^2)^{-n/2}\exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(x_i-\bar{x})^2+2(\bar{x}-\mu)\sum_{i=1}^{n}(x_i-\bar{x})+n(\bar{x}-\mu)^2\right]\right)d\bar{x}ds_1^2$$

$$= (2\pi\sigma^2)^{-n/2}\exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(x_i-\bar{x})^2+n(\bar{x}-\mu)^2\right]\right)d\bar{x}ds_1^2$$

$$= n^{-1/2}(2\pi)^{-1/2}\exp\left(-\frac{n(\bar{x}-\mu)^2}{2\sigma^2}\right)d\left(\frac{\sqrt{n}(\bar{x}-\mu)}{\sigma}\right)$$

$$\times(\pi)^{-(n-1)/2}(n-1)^{-(n-1)/2}(s_1^2)^{-(n-1)/2}\left(\frac{(n-1)s_1^2}{2\sigma^2}\right)^{(n-1)/2-1}\exp\left(-\frac{(n-1)s_1^2}{2\sigma^2}\right)d\left(\frac{(n-1)s_1^2}{2\sigma^2}\right)$$

$$\propto(2\pi)^{-1/2}\exp\left(-\frac{v_1^2}{2}\right)dv_1$$

$$\times\left(\frac{v_2}{2}\right)^{(n-1)/2-1}\exp\left(-\frac{v_2}{2}\right)d\left(\frac{v_2}{2}\right). \quad (19)$$

Normalizing (19), we obtain (14). This ends the proof.

Thus,

$$V_1 \sim N(0,1), \quad V_2 \sim \chi^2_{n-1}, \quad (20)$$

where V_2 is statistically independent of V_1.

Corollary 2.1. If V_1 is a normally distributed random variable with unit variance and zero mean, and V_2 is a chi-squared distributed random variable with $n-1$ degrees of freedom that is statistically independent of V_1, then

$$T=\frac{V_1+\Delta}{\sqrt{V_2/(n-1)}}=\frac{V_1+\Delta}{\sqrt{W}}\sim f_{n-1,\Delta}(t), \quad -\infty<t<\infty, \quad (21)$$

is a non-central t-distributed random variable with $n-1$ degrees of freedom and non-centrality parameter Δ, where

$$W=\frac{V_2}{n-1}=\frac{S_1^2}{\sigma^2}\sim f_{n-1}(w)=\frac{(n-1)^{(n-1)/2}}{2^{(n-1)/2}\Gamma\left((n-1)/2\right)}w^{(n-1)/2-1}\exp(-(n-1)w/2), \quad w\geq 0, \quad (22)$$

$$f_{n-1,\Delta}(t)=\frac{(n-1)^{(n-1)/2}}{\sqrt{\pi}\Gamma\left((n-1)/2\right)2^{n/2}}\frac{\exp\left(-\frac{(n-1)\Delta^2}{2(t^2+n-1)}\right)}{(t^2+n-1)^{n/2}}$$

$$\times\int_0^\infty w_\bullet^{n/2-1}\exp\left(-\frac{1}{2}\left[w_\bullet^{1/2}-\frac{t\Delta}{\sqrt{t^2+n-1}}\right]^2\right)dw_\bullet, \quad -\infty<t<\infty, \quad (23)$$

is the probability density function of T,

$$F_{n-1,\Delta}(t)=\Pr(T\le t)=\frac{(n-1)^{(n-1)/2}}{2^{(n-1)/2}\Gamma\big((n-1)/2\big)}\int_0^\infty w^{(n-1)/2-1}\exp(-(n-1)w/2)\Phi(t\sqrt{w}-\Delta)dw, \qquad (24)$$

is the cumulative distribution function of *T*. $\Phi(x)$ is the standard normal distribution function. Note that the non-centrality parameter Δ may be negative.

3 Tolerance Limits on Order Statistic

3.1 Lower Tolerance Limit

Theorem 3. Let $X_1, \ldots, X_n$ be observations from a preliminary sample of size *n* from a normal distribution defined by the probability density function (10). Then a lower one-sided β-content tolerance limit at level γ, $L_k \equiv L_k(S)$ (on the *k*th order statistic Y_k from a set of *m* future ordered observations $Y_1 \le \ldots \le Y_m$ also from the distribution (10)), which satisfies

$$\Pr\big(P_\theta(Y_k > L_k \mid m) \ge \beta\big) = \gamma, \qquad (25)$$

is given by

$$L_k = \overline{X} + \eta_L S_1, \qquad (26)$$

where

$$\eta_L = -\frac{t_{r,\Delta;\gamma}}{\sqrt{n}}, \qquad (27)$$

is the lower tolerance factor, $t_{r,\Delta;\gamma}$ is the quantile of order γ for the non-central *t*-distribution with $r=n-1$ degrees of freedom and non-centrality parameter $\Delta = -z_{1-\delta_\beta}\sqrt{n},\quad z_{1-\delta_\beta}$ denotes the $1-\delta_\beta$ quantile of a standard normal distribution,

$$\delta_\beta = \frac{(m-k+1)q_{2(m-k+1),2k;\beta}}{(m-k+1)q_{2(m-k+1),2k;\beta}+k}, \qquad (28)$$

$q_{2(m-k+1),2k;\beta}$ is the quantile of order β for the *F* distribution with $2(m-k+1)$ and $2k$ degrees of freedom.

Proof. It follows from (8), (11) and (25) that

$$\Pr\big(P_\theta(Y_k > L_k \mid m) \ge \beta\big) = \Pr\left(\int_0^{\frac{1-F_\theta(L_k)}{F_\theta(L_k)}\frac{2k}{2(m-k+1)}} f_{2(m-k+1),2k}(x)dx \ge \beta\right)$$

$$=\Pr\left(\frac{1-F_\theta(L_k)}{F_\theta(L_k)}\frac{2k}{2(m-k+1)} \ge q_{2(m-k+1),2k;\beta}\right) = \Pr\left(F_\theta(L_k) \le \frac{k}{k+(m-k+1)q_{2(m-k+1),2k;\beta}}\right)$$

$$=\Pr\left(\frac{1}{\sigma\sqrt{2\pi}}\int_{L_k}^\infty \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)dy \ge \frac{(m-k+1)q_{2(m-k+1),2k;\beta}}{(m-k+1)q_{2(m-k+1),2k;\beta}+k}\right)$$

$$= \Pr\left(\frac{1}{\sqrt{2\pi}} \int_{\frac{L_k-\mu}{\sigma}}^{\infty} \exp\left(-\frac{z^2}{2}\right) dz \geq \delta_\beta \right) = \Pr\left(\frac{1}{\sqrt{2\pi}} \int_{\infty}^{\frac{L_k-\mu}{\sigma}} \exp\left(-\frac{z^2}{2}\right) dz \leq 1-\delta_\beta \right)$$

$$= \Pr\left(\frac{L_k-\mu}{\sigma} \leq z_{1-\delta_\beta} \right) = \Pr\left(\frac{L_k-\bar{X}+\bar{X}-\mu}{\sigma} \leq z_{1-\delta_\beta} \right)$$

$$= \Pr\left(\frac{L_k-\bar{X}}{S_1}\sqrt{n}\frac{S_1}{\sigma} + \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \leq z_{1-\delta_\beta}\sqrt{n} \right) = \Pr\left(\eta_L\sqrt{n}\sqrt{W} + V_1 \leq z_{1-\delta_\beta}\sqrt{n} \right)$$

$$= \Pr\left(\frac{V_1 - z_{1-\delta_\beta}\sqrt{n}}{\sqrt{W}} \leq -\eta_L\sqrt{n} \right) = \Pr\left(\frac{V_1+\Delta}{\sqrt{W}} \leq -\eta_L\sqrt{n} \right) = \Pr\left(T \leq -\eta_L\sqrt{n} \right) = F_{r,\Delta}(t), \tag{29}$$

where

$$\eta_L = \frac{L_k - \bar{X}}{S_1}, \tag{30}$$

is the lower tolerance factor,

$$\Delta = -z_{1-\delta_\beta}\sqrt{n}, \quad r = n-1, \quad t = -\eta_L\sqrt{n}. \tag{31}$$

It follows from (25), (29) and (31) that the lower tolerance factor η_L should be chosen such that

$$F_{r,\Delta}(t) = F_{r,\Delta}(-\eta_L\sqrt{n}) = F_{r,\Delta}(t_{r,\Delta;\gamma},) = \gamma, \tag{32}$$

where $t_{r,\Delta;\gamma}$ is the quantile of order γ for the non-central *t*-distribution with *r* degrees of freedom and non-centrality parameter Δ. It follows from (32) that

$$\eta_L = -\frac{t_{r,\Delta;\gamma}}{\sqrt{n}}. \tag{33}$$

It follows from (30) that $L_k = \bar{X} + \eta_L S_1$. This completes the proof.

Corollary 3.1. It follows from (29) that $\Pr\left(\eta_L\sqrt{n}\sqrt{W} + V_1 \leq z_{1-\delta_\beta}\sqrt{n} \right)$ can be transformed as follows:

$$\Pr\left(\eta_L\sqrt{n}\sqrt{W} + V_1 \leq z_{1-\delta_\beta}\sqrt{n} \right) = \Pr\left(V_1 \leq -\eta_L\sqrt{n}\sqrt{W} + z_{1-\delta_\beta}\sqrt{n} \right)$$

$$= \int_{-\infty}^{-\eta_L\sqrt{n}\sqrt{W}+z_{1-\delta_\beta}\sqrt{n}} f_1(v_1)dv_1 = \Phi(-\eta_L\sqrt{n}\sqrt{W} + z_{1-\delta_\beta}\sqrt{n}) = \Phi(t\sqrt{W} - \Delta), \tag{34}$$

where

$$t = -\eta_L\sqrt{n}, \quad \Delta = -z_{1-\delta_\beta}\sqrt{n}. \tag{35}$$

Then it follows from (25) and (34) that *t* has to be found such that

$$t=\arg\left(E\left\{\Phi(t\sqrt{W}-\Delta)\right\}=\gamma\right)=\arg\left(\int_0^\infty \Phi(t\sqrt{w}-\Delta)\,f_r(w)dw=\gamma\right)$$

$$=\arg\left(\frac{r^{r/2}}{2^{r/2}\Gamma\left(r/2\right)}\int_0^\infty w^{r/2-1}\exp(-rw/2)\Phi(t\sqrt{w}-\Delta)\;dw=\gamma\right)$$

$$=\arg\left(F_{r,\Delta}(t)=\gamma\right)=t_{r,\Delta;\gamma},\qquad (36)$$

where $t_{r,\Delta;\gamma}$ is the quantile of order γ for the non-central *t*-distribution with $r=n-1$ degrees of freedom and non-centrality parameter Δ,

$$F_{r,\Delta}(t)=\Pr(T\le t)=\frac{r^{r/2}}{2^{r/2}\Gamma\left(r/2\right)}\int_0^\infty w^{r/2-1}\exp(-rw/2)\Phi(t\sqrt{w}-\Delta)dw.\qquad (37)$$

is the cumulative distribution function of *T*,

$$f_{r,\Delta}(t)=F'_{r,\Delta}(t)=\frac{r^{r/2}\exp\left(-r\Delta^2/[2(t^2+r)]\right)}{\sqrt{\pi}\Gamma\left(r/2\right)2^{(r+1)/2}(t^2+r)^{(r+1)/2}}$$

$$\times\int_0^\infty w_\bullet^{(r-1)/2}\exp\left(-\frac{1}{2}\left[w_\bullet^{1/2}-\frac{t\Delta}{\sqrt{t^2+r}}\right]^2\right)dw_\bullet,\quad -\infty<t<\infty,\qquad (38)$$

is the probability density function of *T*, where

$$W_\bullet=W(t^2+r).\qquad (39)$$

Corollary 3.2. If

$$W_{\bullet\bullet}=\frac{W(t^2+r)}{2},\qquad (40)$$

then

$$f_{r,\Delta}(t)=F'_{r,\Delta}(t)=\frac{r^{r/2}\exp(-\Delta^2/2)}{\sqrt{\pi}\Gamma\left(r/2\right)(t^2+r)^{(r+1)/2}}\int_0^\infty w_{\bullet\bullet}^{(r-1)/2}\exp\left(-\left[w_{\bullet\bullet}-\frac{t\Delta\sqrt{2}}{\sqrt{t^2+r}}w_{\bullet\bullet}^{1/2}\right]\right)dw_{\bullet\bullet}$$

$$=\frac{r^{r/2}\exp(-\Delta^2/2)}{\sqrt{\pi}\Gamma\left(r/2\right)(t^2+r)^{(r+1)/2}}\sum_{j=0}^\infty\frac{\Gamma((r+j+1)/2)}{j!}\left(\frac{t\Delta\sqrt{2}}{\sqrt{t^2+r}}\right)^j,\quad -\infty<t<\infty.\qquad (41)$$

This form of the density function is derived in Rao [20] and appears in Searle [21]. In both Rao and Searle, $\sqrt{\pi}$ is incorrectly omitted from the denominator. It should also be noted that the central *t*-distribution is just a special case of the non-central *t* with $\Delta = 0$.

Corollary 3.3. If $k=m=1$, then

$$\delta_\beta=\beta,\quad \Delta=-z_{1-\beta}\sqrt{n}.\qquad (42)$$

3.2 Upper Tolerance Limit

Theorem 4. Let $X_1, \ldots, X_n$ be observations from a preliminary sample of size *n* from a normal distribution defined by the density function (10). Then an upper one-sided β-content tolerance limit at level γ, U_k

$\equiv U_k(S)$ (on the kth order statistic Y_k from a set of m future ordered observations $Y_1 \le \ldots \le Y_m$ also from the distribution (10)), which satisfies

$$\Pr\left(P_\theta(Y_k \le U_k \mid m) \ge \beta\right) = \gamma, \tag{43}$$

is given by

$$U_k = \overline{X} + \eta_U S_1, \tag{44}$$

where

$$\eta_U = \frac{t_{r,\Delta;1-\gamma}}{\sqrt{n}}, \tag{45}$$

is the upper tolerance factor, $t_{r,\Delta;1-\gamma}$ is the quantile of order $1-\gamma$ for the non-central t-distribution with $r=n-1$ degrees of freedom and non-centrality parameter $\Delta = -z_{1-\delta_{1-\beta}}\sqrt{n}$, $z_{1-\delta_{1-\beta}}$ denotes the $1-\delta_{1-\beta}$ quantile of a standard normal distribution,

$$\delta_{1-\beta} = \frac{(m-k+1)q_{2(m-k+1),2k;1-\beta}}{(m-k+1)q_{2(m-k+1),2k;1-\beta} + k}, \tag{46}$$

$q_{2(m-k+1),2k;1-\beta}$ is the quantile of order $1-\beta$ for the F distribution with $2(m-k+1)$ and $2k$ degrees of freedom.

Proof. It follows from (3), (11) and (43) that

$$\Pr\left(P_\theta(Y_k \le U_k \mid m) \ge \beta\right) = \Pr\left(\int_{\frac{1-F_\theta(U_k)}{F_\theta(U_k)}\frac{2k}{2(m-k+1)}}^{\infty} f_{2(m-k+1),2k}(x)dx \ge \beta \right)$$

$$= \Pr\left(\int_0^{\frac{1-F_\theta(U_k)}{F_\theta(U_k)}\frac{2k}{2(m-k+1)}} f_{2(m-k+1),2k}(x)dx \le 1-\beta \right)$$

$$= \Pr\left(\frac{1-F_\theta(U_k)}{F_\theta(U_k)}\frac{2k}{2(m-k+1)} \le q_{2(m-k+1),2k;1-\beta} \right)$$

$$= \Pr\left(F_\theta(U_k) \ge \frac{k}{k+(m-k+1)q_{2(m-k+1),2k;1-\beta}} \right)$$

$$= \Pr\left(\frac{1}{\sigma\sqrt{2\pi}} \int_{U_k}^{\infty} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)dy \le \frac{(m-k+1)q_{2(m-k+1),2k;1-\beta}}{(m-k+1)q_{2(m-k+1),2k;1-\beta}+k} \right)$$

$$= \Pr\left(\frac{1}{\sqrt{2\pi}} \int_{\frac{U_k-\mu}{\sigma}}^{\infty} \exp\left(-\frac{z^2}{2}\right)dz \le \delta_{1-\beta} \right) = \Pr\left(\frac{1}{\sqrt{2\pi}} \int_{\infty}^{\frac{U_k-\mu}{\sigma}} \exp\left(-\frac{z^2}{2}\right)dz \ge 1-\delta_{1-\beta} \right)$$

$$= \Pr\left(\frac{U_k - \mu}{\sigma} \geq z_{1-\delta_{1-\beta}}\right) = \Pr\left(\frac{U_k - \bar{X} + \bar{X} - \mu}{\sigma} \geq z_{1-\delta_{1-\beta}}\right)$$

$$= \Pr\left(\frac{U_k - \bar{X}}{S_1}\sqrt{n}\frac{S_1}{\sigma} + \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \geq z_{1-\delta_{1-\beta}}\sqrt{n}\right) = \Pr\left(\eta_U\sqrt{n}\sqrt{W} + V_1 \geq z_{1-\delta_{1-\beta}}\sqrt{n}\right)$$

$$= \Pr\left(\frac{V_1 - z_{1-\delta_{1-\beta}}\sqrt{n}}{\sqrt{W}} \geq -\eta_U\sqrt{n}\right) = \Pr\left(\frac{V_1 + \Delta}{\sqrt{W}} \geq -\eta_U\sqrt{n}\right) = \Pr\left(T \geq -\eta_U\sqrt{n}\right) = 1 - F_{r,\Delta}(t), \quad (47)$$

where

$$\eta_U = \frac{U_k - \bar{X}}{S_1}, \quad (48)$$

is the upper tolerance factor,

$$\Delta = -z_{1-\delta_{1-\beta}}\sqrt{n}, \quad r = n - 1, \quad t = -\eta_U\sqrt{n}. \quad (49)$$

It follows from (43), (47) and (49) that the upper tolerance factor η_U should be chosen such that

$$F_{r,\Delta}(t) = F_{r,\Delta}(-\eta_U\sqrt{n}) = F_{r,\Delta}(t_{r,\Delta;1-\gamma},) = 1 - \gamma, \quad (50)$$

where $t_{r,\Delta;1-\gamma}$ is the quantile of order $1-\gamma$ for the non-central *t*-distribution with *r* degrees of freedom and non-centrality parameter Δ. It follows from (50) that

$$\eta_U = -\frac{t_{r,\Delta;1-\gamma}}{\sqrt{n}}. \quad (51)$$

It follows from (48) that $U_k = \bar{X} + \eta_U S_1$. This completes the proof.

Remark 1. It will be noted that an upper tolerance limit may be obtained from a lower tolerance limit by replacing β by $1-\beta$, γ by $1-\gamma$.

4 Practical Example of Finding a Warranty Assessment of Image Quality

The image quality assessment (IQA) plays a very crucial role in image and video processing. The aim is to replace human judgment of perceived image quality with a machine evaluation. A large effort has been devoted to developing IQA measures that try to mimic human perception. While many methods and models still rely on simple measures, such as the peak-siqnal-to-noise-ratio (PSNR) and the mean-squared error (MSE), many others use sophisticated signal processing techniques, such as multi-channel filtering [22], [23], discrete cosine transform [24], [25], multi-scale wavelet decompositions [26], [27], and Wigner-Ville distribution [28]. To date, however, it has been very difficult to find a reliable objective measure that correlates very highly with human perception [29].

Digital images are subject to a wide variety of distortions during acquisition, processing, compression, storage, transmission and reproduction, any of which may result in a degradation of visual quality. Quality measuring is needed for many applications, for example if the designer of a medical device want to decide from which device get the better results so he want to measure the quality of the images from those devices. Quality can be measured in two ways subjective and objective. The presence of blur in an image can be easily identified by the human eye but it is difficult for the computer. In practice, however, subjective evaluation is usually too inconvenient, time-consuming and expensive. The goal of research in objective image quality assessment is to develop quantitative measures that can automatically predict perceived image quality. Quality assessment algorithms are needed to monitor the quality for real time applications. Subjective methods are impossible to implement in real time systems, so objective methods are more attracted in recent years. All these methods want to have high correlation with human perception or judgments.

Problem description. An IQA device manufacturer has the data of image quality assessment (in terms of the spearman correlation) obtained from testing *n*=10 IQA devices. These data are given in Table 1.

Table 1. The data of image quality assessment.

Observations (in terms of the spearmen correlation)									
X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
0.913	0.916	0.923	0.926	0.936	0.947	0.961	0.971	0.975	0.992

A buyer tells the device manufacturer that he wants to place two orders for the same type of IQA devices to be shipped to two different destinations. The buyer wants to select a random sample of *m*=5 IQA devices from each shipment to be tested. An order is accepted only if all of 5 IQA devices in each selected sample meet the warranty image quality assessment (in terms of the spearman correlation). What warranty IQA (in terms of the spearman correlation) should the manufacturer offer so that all of 5 IQA devices in each selected sample meet the warranty with probability of 0.95?

In order to find this warranty IQA, the manufacturer wishes to use a random sample of size *n*=10 given in Table 1 and to calculate the lower one-sided simultaneous tolerance limit $L_{k=1}(S)$ (warranty IQA) which is expected to capture a certain proportion, say, β=0.95 or more of the population of selected items (*m*=5), with a given confidence level γ=0.95. This tolerance limit is such that one can say with a certain confidence γ that at least 100β% of the IQA devices in each sample selected by the buyer for testing will give image quality assessment (in terms of the spearman correlation) no less than $L_1(S)$.

Goodness-of-fit testing. It is assumed that the data of Table 1 follow the normal probability distribution

$$F_\theta(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx, \tag{52}$$

where the parameters μ and σ are unknown. Thus, for the above example, we have that *n* =10, *m* =5, *k* = 1, β = 0.95, γ = 0.95,

$$S = \left(\bar{X} = \sum_{i=1}^{n} X_i \,/\, n = 0.946,\ S_1^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 \,/\, (n-1) = 0.000758\right). \tag{53}$$

We assess the statistical significance of departures from the model (52) by performing the Anderson–Darling goodness-of-fit test. The Anderson–Darling test statistic value is determined by (e.g. [30]):

$$A^2 = -\left[\sum_{i=1}^{n} (2i-1)\left(\ln F_\theta(x_i) + \ln\left(1 - F_\theta(x_{n+1-i})\right)\right)\right] \Big/ n - n, \tag{54}$$

where $F_\theta(\ )$ is the cumulative distribution function ,

$$\theta = (\mu = \bar{x}, \sigma = s_1), \tag{55}$$

n is the number of observations.

The result from (54) needs to be modified for small sampling values. For the normal distribution the modification of A^2 is

$$A_{\mathrm{mod}}^2 = A^2(1+0.75/n+2.25/n^2). \quad (56)$$

The A_{mod}^2 value must then be compared with critical values, A_α^2, which depend on the significance level α and the distribution type. As an example, for the normal distribution the determined A_{mod}^2 value has to be less than the following critical values for acceptance of goodness-of-fit (see Table 2):

Table 2. Critical values for A_{mod}^2.

α	0.1	0.05	0.025	0.01
A_α^2	0.631	0.752	0.873	1.035

For this example, α=0.05, $A_{\alpha=0.05}^2 = 0.752$,

$$A^2 = -\left[\sum_{i=1}^{10}(2i-1)\left(\ln F_\theta(x_i)+\ln\left(1-F_\theta(x_{n+1-i})\right)\right)\right]\Big/10-10 = 0.296378, \quad (57)$$

$$A_{\mathrm{mod}}^2 = A^2(1+0.75/10+2.25/10^2) = 0.325275 < A_{\alpha=0.05}^2 = 0.752. \quad (58)$$

Thus, there is not evidence to rule out the normal model (52).

Finding lower tolerance limit (warranty assessment of image quality). Now the lower one-sided simultaneous β-content tolerance limit at the confidence level γ, $L_1 \equiv L_1(S)$ (on the order statistic Y_1 from a set of m = 5 future ordered observations ($Y_1 \le \ldots \le Y_m$)) can be obtained from (26).

Since m=5, k=1, β=0.95,

$$\delta_\beta = \frac{(m-k+1)q_{2(m-k+1),2k;\beta}}{(m-k+1)q_{2(m-k+1),2k;\beta}+k} = 0.989796, \quad (59)$$

$$r = n-1 = 9, \quad \Delta = -z_{1-\delta_\beta}\sqrt{n} = 7.3325, \quad \gamma = 0.95, \quad (60)$$

the quantile of order γ for the non-central t-distribution with r degrees of freedom and non-centrality parameter Δ is given by

$$t_{r,\Delta;\gamma} = \arg\left(F_{r,\Delta}(t)=\gamma\right) = 12.5512, \quad (61)$$

the lower tolerance factor is given by

$$\eta_L = -\frac{t_{r,\Delta;\gamma}}{\sqrt{n}} = -3.96904, \quad (62)$$

it follows from (26) that

$$L_{k=1} = \bar{X} + \eta_L S_1 = 0.837. \quad (63)$$

Statistical inference. Thus, the manufacturer has 95% assurance that at least 100β% of the IQA devices in each sample (m=5) selected by the buyer for testing will give the warranty assessment of image quality (in terms of the spearman correlation) no less than L_1=0.837.

5 Conclusion

Tolerance limits enjoy a fairly rich history in the literature and have a very important role in engineering and manufacturing applications. In contrast to other statistical limits commonly used for statistical inference, the tolerance limits (especially for the order statistics) are used relatively rarely. One reason is that the theoretical concept and computational complexity of the tolerance limits is significantly more difficult than that of the standard confidence and prediction limits. Thus it becomes necessary to use new or innovative approaches which will allow one to construct tolerance limits on future order statistics for many populations.

REFERENCES

[1]. Mendenhall, V., *A bibliography on life testing and related topics*. Biometrika, 1958. XLV: p. 521-543.

[2]. Guttman, I., *On the power of optimum tolerance regions when sampling from normal distributions*. Annals of Mathematical Statistics, 1957. XXVIII: p. 773-778.

[3]. Wald, A. and J. Wolfowitz, *Tolerance limits for a normal distribution*. Annals of Mathematical Statistics, 1946. XVII: p. 208-215.

[4]. Wallis, W. A., *Tolerance intervals for linear regression*, in Second Berkeley Symposium on Mathematical Statistics and Probability, 1951. Berkeley: University of California Press, p. 43-51.

[5]. Patel, J. K., *Tolerance limits: a review*. Communications in Statistics: Theory and Methodology, 1986. 15: p. 2719-2762.

[6]. Dunsmore, I. R., *Some approximations for tolerance factors for the two parameter exponential distribution*. Technometrics, 1978. 20: p. 317-318.

[7]. Guenther, W. C., S. A. Patil, and V. R. R. Uppuluri, *One-sided β-content tolerance factors for the two parameter exponential distribution*. Technometrics, 1976. 18: p. 333-340.

[8]. Engelhardt, M. and L. J. Bain, *Tolerance limits and confidence limits on reliability for the two-parameter exponential distribution.* Technometrics, 1978. 20: p. 37-39.

[9]. Guenther, W. C., *Tolerance intervals for univariate distributions*. Naval Research Logistics Quarterly, 1972. 19: p. 309-333.

[10]. Hahn, G. J. and W. Q. Meeker, *Statistical Intervals: A Guide for Practitioners*. 1991. New York: John Wiley & Sons.

[11]. Nechval, N. A. and K. N. Nechval, *Tolerance limits on order statistics in future samples coming from the two-parameter exponential distribution*. American Journal of Theoretical and Applied Statistics, 2016. 5: p. 1-6.

[12]. Johnson, N. L., S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, vol. 1. Second edition, 1994. New York: John Wiley & Sons.

[13]. Patel, J. K. and C. B. Read, *Handbook of the Normal Distribution*. Second edition, 1996. New York: Marcel Dekker.

[14]. Balakrishnan, N. and V. B. Nevzorov, *A Primer on Statistical Distributions*. 2003. New Jersey: Wiley.

[15]. Kotz, S. and D. Vicari, *Survey of developments in the theory of continuous skewed distributions*. Metron, 2005. 63: p. 225-261.

[16]. Nechval, N. A. and E. K. Vasermanis, *Improved Decisions in Statistics*. 2004. Riga: Izglitibas soli.

[17]. Nechval, N. A., K. N. Nechval, and E. K. Vasermanis, *Effective state estimation of stochastic systems*. Kybernetes (An International Journal of Systems & Cybernetics), 2003. 32: p. 666-678.

[18]. Nechval, N. A., G. Berzins, M. Purgailis, and K. N. Nechval, *Improved estimation of state of stochastic systems via invariant embedding technique*. WSEAS Transactions on Mathematics, 2008. 7: p. 141-159.

[19]. Nechval, N. A., M. Purgailis, K. N. Nechval, and V. F. Strelchonok, *Optimal predictive inferences for future order statistics via a specific loss function*. IAENG International Journal of Applied Mathematics, 2012. 42: p. 40-51.

[20]. Rao, C. R., *Linear Statistical Inference and its Applications*. 1965. New York: Wiley.

[21]. Searle, S. R., *Linear Models*. 1971. New York: Wiley.

[22]. Daly S., *The visual difference predictor: an algorithm for the assessment of image fidelity*, in Human Vision, Visual Processing, and Digital Display, Proc. SPIE, vol. 1666, 1992, San Jose, CA, p. 2-15.

[23]. Heeger, D. J. and P. C. Teo, *A model of perceptual image fidelity*, in Proc. IEEE International Conference on Image Processing, vol. 2, 1995, p. 343-345.

[24]. Malo, J., A. M. Pons, and 5. M. Artigas, *Subjective image fidelity metric based on bit allocation of the human visual system in the DCT domain*. Image und Vision Computing, 1997. 15: p. 535-548.

[25]. Watson, A. B., J. Hu, and J. F. McGowm, *Digital video quality metric based on human vision*. J. of Electronic Imaging, 2001. 10: p. 20-29.

[26]. Lai, Y. K. and C.-C. J. Kuo, *A Haar wavelet approach to compressed image quality measurement*. J. Visual Communication und Image Repres., 2000. 11: p. 17-40.

[27]. Beghdadi, A. and B. Pesquet-Popescu, *A new image distortion measure based on wavelet decomposition*, in Proc. Seventh Intern. Symp. Signal Proces. its Applications (ISSPA-2003), vol. 1, 2003, Paris, p. 485-488.

[28]. Beghdadi, A. and R. Iordache, *Image quality assessment using the joint space/spatial-frequency representation*. EURASIP Journal on Applied Signal Processing, 2006. 6: p. 1-8.

[29]. Wang, Z., A. C. Bovik, and L. Lu, *Why is image quality assessment so difficult?* In Proc. IEEE Inter. Conference Acoustics, Speech, and Signal Processing (ICASSP-2002), vol. 4, 2002, Orlando FL, p. 3313-3336.

[30]. Dodson, B., *Weibull Analysis*. 1994. Milwaukee: ASQ Quality Press.

Society for Science and Education United Kingdom

AIVP Advances in Image and Video Processing

Volume 4, Issue 3
ISSN 2054-7412

Image Matching for Aadhar Card Based on Textural and Geometrical Feature for Identification of Human

Anjali Swamy and Amit Yerpude
Department of Computer Science and Engineering, Rungta College of Engineering and Technology Kohka road , Bhilai(C.G), India
anj.swamy1469@gmail.com; amit.yerpude@gmail.com

ABSTRACT

Face appreciation from picture or film is a trendy topic in biometrics examine. Many civic places habitually have supervision cameras for film confine along with these cameras contain their large value for defense principle. It is broadly accredited that the features appreciation has played an imperative role in scrutiny system as it doesn't require the object's assistance. The genuine reward of features based discovery over added biometrics are rareness and approval. As creature expression is a energetic object having lofty degree of unevenness in its facade, that makes expression recognition a thorny trouble in mainframe idea. In this pasture, precision and rapidity of classification is a key matter.The target of this thesis is to appraise different expression discovery and appreciation method, offer whole clarification for reflection based visage recognition and detection with superior precision, enhanced reaction fee as an early stride for record inspection. resolution is projected based on perform tests on different look affluent database in requisites of subjects, facade, emotion, battle and radiance.

Keywords— SIFT features, Biometric security,Face recognition, Adhar card, Passport size photograph.

1 Introduction

During today's plexus humanity require to preserve the scrutiny of order or significant goods is fetching both gradually more imperative and tricky. periodically we listen regarding the crime of credit card con, processor breakings via hackers or defense crack in a corporation or direction structure.Biometric is automated process of recognize a self based on a physical or else behavioral characterstics.[2] Aadhar card as a verification of character which could be old to authenticate appearances of a person's own identity planned to allow clients that they be society of INDIA.Ballot chains is the vital dilemma of voting this paper is based on this method in which calculated results will be accurate.

DOI: 10.14738/aivp.43.1996
Publication Date: 23rd July, 2016
URL: http://dx.doi.org/10.14738/aivp.43.1996

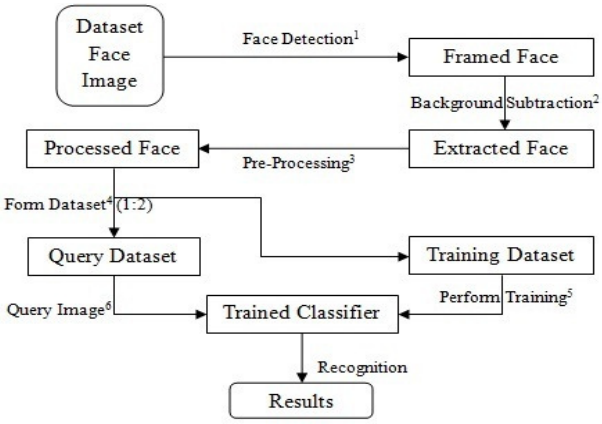


Figure 1. System Overview

Voting as an hour:- Finally ballet is an honour conducted between the citizens by the beginning father. By use their accurate to choose, people disclose their esteem for the the past of the nation. Today e-voting machine(EVM) is used for ballet .This tries to look into e-voting from a diverse approach of the needed people sanction from a diverse slant in its place of concept such as OTP or smart card we endeavor to appear into the pro and con of biometric draw near.

2 The Existing Ranking Methods

Foregoing, The most popular procedure of keeping vote into the polling booth they used voter ID card as Authentication. These schemes desires the user to validate themselves by entering a address to validate that whether they are the citizens of that location or not.But when voter ID card are lost or not yet created at that time vote rigging will be done to make the vote for election which can be done by government.A biometric identification is one in which the user's face becomes the password/PIN biometric characterstics of an individual are unique and therefore can be used to authenticate a user's access to polling booth.

3 User Query Intent and Storage Tags

Sanctuary Experts says that Automatic Teller Machine (ATM) in potential will have biometric verification techniques toward validate identity of client through matter. In South America, near are company that include introduce fingerprint skill as a rooted piece of ATM system, where nation have previously ongoing using fingerprint in set of PIN or Password meant for broad detection with their ID cards.Gregg Rowley said- —Banks will move to smart cards and biometric will be next step after that[17] Electoral fraud can occur in advance of voting if the configuration of electorate is revised. The legitimacy of this type of manipulation varies across authority. Deliberate manipulation of election consequences is widely considered a violation of the principles of democracy.

3.1 Vote Rigging:

Vote rigging is an unauthorized interference with the process of an election acts of manipulation fraud affect vote counts to bring about an voting result. Whether by increasing the vote shared of the approved candidate distressing the vote share of the competitor candidate or both.[3]

3.2 Corruption in democracy:

As per [4] In a narrow election a small amount of crime may be enough to convert the result,Even if the outcome is not affected scam can still have a damaging effect if not a punished as it can diminish votersassurance in democracy.Even the awareness of scam can be damaging as it create people less apt to accept election result.Crooked election can margin to the disruption of democracy and the formulation of ratification of dictatorship.

3.3 Misuse of proxy votes:

As per [5] proxy voting is a particularly endangered to election scam, due to the aggregate of trust placed in the person who throw the vote.In several countries there have been assertion of evacuation home.citizens being asked to fill out Absentee voter form. When the form are signs and accumulated,they are then secretly recast as application for proxy votes, naming party opponent or their friends and relatives gives votes as proxies.But these people stranged to the voter then cast the vote for the party of their choice.This deception relies on elderly care home native typically being careless or suffering from insanity.

4 A New Optimized Ranking Algorithm

Biometric is the science that tries to fetch human biological facial appearance with an mechanized machine either to identify or authenticate. Biometric products eliminates the needs for password and PINS biometric system replace knowledge with entity features such as face recognition or immediacy detection. It makes it relaxed and fast to record features. The investigation of human data using the face recognition termed as the biometric.Initially the application of biometrics.[8]

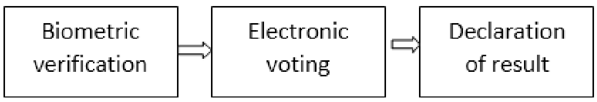
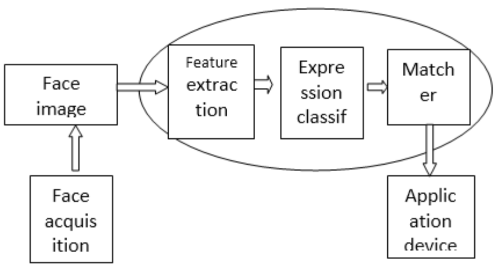


Figure 2: Biometric electronic voting architecture

4.1 Biometric Registration of voters:

This imply the method of capturing an appropriate(18 year of age or beyond with a sound mind a national and tenant in the country) Voter's personal information (Name,DOB, Home town, language,Address,Family,Passport,photography,etc.) with face recognition using recognition system and stored to the voters database be used for confirmation or certification on determination day.



4.2 Candidate Registration:

Each supporting party or sovereign candidate after fleeting all the necessities of the electoral process obligatory for contesting in the selection would then be registered to the scheme after going during the ballot spot balloting apiece nominee individual information.

4.3 Voter's Verification or Authentication:

A face gratitude system operates moreover in verification approach or in recognition approach.

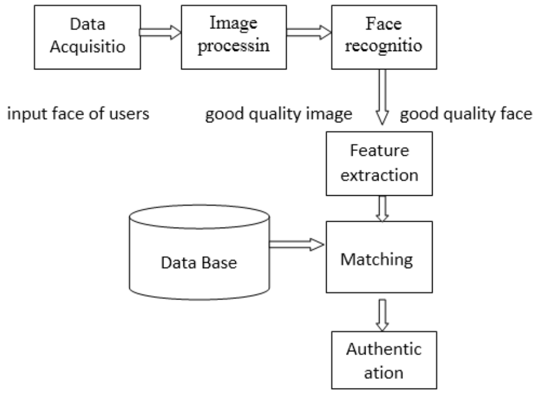


Figure 3: Architecture of face verification system

The first phase is the data acquisition in which a face image is obtained from an individual person.The next phase is the pre-processing phase in which the enter image is processed with some customary image processing algorithms for noise exclusion and smoothening.The pre-processed face icon is then improved via exclusively planned enrichment algorithms which exploit the episodic and directional life of ridges.The augmentation image is then used to dig out prominent features in the feature extraction stage.Finally the extracted textures are used for similar phase.[9]

4.4 Face Verification:

This paper introduces a prototype automatic identity authentication system that is capable of authenticating the identity of an individual using face images. The main algorithms are:

4.4.1 Fiducial Point:

A fiducial marker or fiducial is an object placed in the field of view of an imaging system which appears in the image produced, for use as a point of orientation or a measure. It may be either something placed into or on the imaging subject, or a spot or set of marks in the reticle of an optical instrument.In fiducial point length between eyes and nose will get calculated between similar images to detect the image whether it is matching or not.

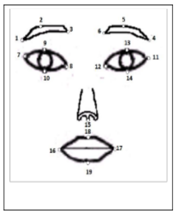


Figure 4: Architecture of fiducial points

Fiducial point is the detection between eyes nose and mouths. The performance of these tasks is usually to a large degree dependent on the accuracy of the facial point detector. It denote those detectors as facial component detectors. However, the cues for tasks like facial expression recognition or gaze detection lie in the more detailed positions of points within these facial components.Vectors V1 and V2 are nodes .The difference between the angles of the two vectors is $^{\alpha}$(alpha) and the ratio between the length of two vectors.

4.4.2 Scale Invariant Feature Transform:

SIFT is one of the most widely used technique of feature extraction. SIFT descriptor can transform image into scale invariant feature keypoints. The SIFT descriptor remain invariant under rotation ,scaling and variation in lightening condition. In original SIFT algorithm Euclidean distance is used for matching keypoints.This SIFT is used to smoothen the image and its points after getting the fiducial point. It will eliminate the unneccessary unstable points from the preceding algorithm.There are four major steps of SIFT algorithm:

- Scale space extrema detection: In this step candidate keypoints are detected.Image get convolved WithGaussianfiltersatdifferentscales
- and we take difference of successive Gaussian blurred image. DoG of image at different scales is-D(x,y,σ)= L(x,y,k_iσ) – L(x,y,kjσ)

  Where L(x,y,kσ) is convolution of original image.

  I(x,y) = Original image.

  G(x,y,kσ)=Gaussian blurred image

  L(x,y,kσ) = G(x,y,kσ) * I(x,y)

After DoG obtained we have to find local maxima/minima of DoG image across scales for this-We have to compute each pixels in the DoG image to its eight neighbour at the same scale and nineth corrosponding neighbouring pixels in each of neighbouring scales. If the pixels is maximum/minimum among all compared pixels than the pixels is selected as a candidate keypoint.

- Keypoint Localization: After 1st step too many candidates points are generated in which some are stable and some are unstable.The main use of this keypoint is localization is to discard those points that have low contrast or poorly localized along an edge.
- Orientation Assignment: Gaussian smoothed image L(x,y,σ) at the keypoint scale σ is taken so that all computations are performed in a scale invariant manner.

  For Example- A two parallel edges are also closed to each other and we use a strong gaussian filter these edge than merged

  Let magnitude = P(i,j) (vertical deviation)

  orientation =Q(I,j) (horizontal deviation).

  Wherever

  P(i,j)=[S(i+1)-S(i,j)+S(i+1,j+1)-S(i,j+1)]/2]

  Where

  Q(i,j)= [S(i,j+1)- S (i,j) + S(i+1,j+1) – S(i+1,j)]/2]

  Magnitude

  m(i,j)= √ (P(i,j))*(P(i,j))+(Q(i,j))*(Q(i,j))

  Orientation K(i,j)=arctan(Q(i,j),P(i,j))

  The Magnitude and direction are calculated for every pixel in a neighbouring region around the keypoint in the Gaussian blurred image L.

- Keypoint Descriptor: There are some steps of keypoint descriptor-
  1. Find the blurred image of closest scale.
  2. Sample the points around the keypoints.
  3. Rotate the gradients and cordinated by the previously computer orientations.

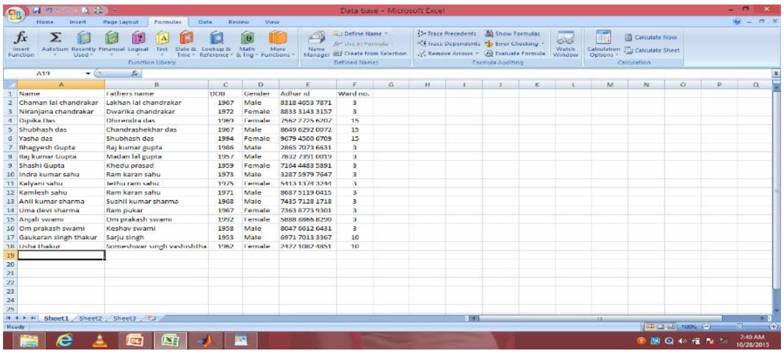
4. Seprate the region into sub region.
5. Create histogram for each subregions with 8 bins.
6. Weight the sample with N(σ)=1.5 Region width.
7. Trilinear interpolation(1-d factor) to place in histogram bins.
8. Actual implementations uses 4*4 descriptors from 16*16 which leads to a 4*4*8=128 element vector.

4.4.3 Sparsity face recognition:

L1- minimization refers to finding the minimum l1- norm minimization solution to an undetermined linear system b=Ax.L1 –norm minimization solution is the sparsest solution.It is used to compress the image and then store it into the database. To compress the image l1- norm minimization technique will be used.

5 Experiment and Analysis

Preprocessing is requisite since nearly everyone of the descriptions in the databases are sloping in some track. Therefore every metaphors are rotate in such a mode that a streak connecting eyes lies in a straight line. For textural system, a subspace be pick by relate features anthropometric calculate to stay away from the computational saddle by means of entire features. Worn to produce the subspace so as to contain needed details in the visage.Images of database tables are given in below Figure:



Name	Fathers name	DOB	Gender	Adhar id	Ward no.
Chaman lal chandrakar	Lakhan lal chandrakar	1967	Male	8318 4653 7871	3
Niranjana chandrakar	Dwarika chandrakar	1972	Female	8833 3143 3157	3
Dipika Das	Dhirendra das	1969	Female	7562 2725 6207	15
Shubhash das	Chandrashekhar das	1967	Male	8649 6292 0072	15
Yasha das	Shubhash das	1994	Female	9679 4500 6709	15
Bhagyesh Gupta	Raj kumar gupta	1986	Male	2865 7073 6631	3
Raj kumar Gupta	Madan lal gupta	1957	Male	7832 7391 0019	3
Shashi Gupta	Khedu prasad	1959	Female	7164 4483 5891	3
Indra kumar sahu	Ram karan sahu	1973	Male	3287 5979 7647	3
Kalyani sahu	Jethu ram sahu	1975	Female	5413 1374 3244	3
Kamlesh sahu	Ram karan sahu	1971	Male	8687 5119 0415	3
Anil kumar sharma	Sushil kumar sharma	1968	Male	7435 7128 1718	3
Uma devi sharma	Ram pukar	1967	Female	7363 8773 9301	3
Anjali swami	Om prakash swami	1992	Female	5888 8866 8290	3
Om prakash swami	Keshav swami	1958	Male	8047 6612 6431	3
Gaukaran singh thakur	Sarju singh	1953	Male	6971 7013 3367	10
Usha thakur	Someshwar singh vashishtha	1962	Female	2422 1082 4851	10

Figure 5: Database

The database contains the details of 17 peoples from which we had coleected the adhar card for their details and passport size photo for face detection with biometrics. The images of collected photos of persons are given below:

Figure 6: Passport images

There are testing and training two features will be taken for biometric face recognition.The experimental result of testing features will be given below

5.1 Testing features:

Figure 7: overview of matching

In testing features the length of eyes nose and lips will be calculated

Length of eyes1 is=25.2982

Length of eyes2 is= 36.4966

Length of nose is= 23.0217

Length of width of nose=45.0111

Length of Lips is= 37.6358

Similarly value of all persons will be taken out and calculated.

5.2 Training Features:

Training features will be calculated from the images of adhar card faces and will be together calculated.17 values calculated and each persong having 5 values i.e length of eyes1 and eyes2 , length of nose , width of nose , and length of lips.

1. 18.682,28.018,30.871,55.009,37.878
2. 30.364,28.018,22.023,40.112,31.56
3. 34.234,40.447,25,67.007,58.128
4. 35.355,38.328,21.024,51.788,27.674
5. 31.575,24.739,31.064,56.719,48.603
6. 37.054,36.346,6.0828,56.009,30.586
7. 38.013,38,38.21,58.078,40.321
8. 37.216,39.051,35,62.129,40.672
9. 37.054,40.792,11.402,63.388,42.695

10. 23.087,30.265,27.074,53.151,41.904

11. 34.713,36.77,40.497,46.011,24.923

12. 35.228,38.079,5.099,57.219,21.516

13. 34.132,39.623,49.659,62.37,34.587

14. 29,41.304,30.017,44.181,44.196

15. 30.414,37.336,22.023,47,36.147

16. 42.012,40.012,30.806,59.414,35.228

17. 39.013,31.145,25.02,52,35.732

5.3 Matching:

Matching will be done by using KNN Algorithm in which min value be considered as a most matched imag.The images of database will get matched one by one with the passport images of the collected persons and then it will show the number of resultant image.The resultant image will be shown below:

Figure 8: Matched image

6 Conclusions

The proposed framework can be analyzed in terms of feasibility and acceptance in the industry.Therefore trying to improve the performance of existing methods and introducing the new methods for face recognition based on today's software project requirement can be future work in this area.Our experiments indicate that better security than other methods like passwords, PIN Number.The benefits of multibiometrics may become even more evident in the case of a large database of users face recognition technology can be used worldwide to access buildings however it had been used in ATM's according to[7] but it can be used in voting machine which would help to reduce the vote rigging electoral fraud.

REFERENCES

[1] Lusekelo Kibona et al. Face Recognition as a Biometric Security for Secondary Password for ATM Users Ruaha catholic university(RUCU) Department of computer science Volume1,Issue2, 2015.

[2] I. Diebold "ATM fraud and security: White paper," Newyork , 2002 department of Computer Science and Engineering.

[3] Anil Pandit and R.C Gangwar " Issues and challenges in electronic voting and direct recording electronic voting system".International Journal of Advanced Research in computer science and software engineering January 2015.

[4] Sarah Birch "Breifing paper on electoral corruption" at 05/11.At IDCR.

[5] Ines Levin, Gabe A. Cohn, Peter, C.Ordershoot and R. Michael Alvarez "Detecting voter fraud in an electronic voting context .An analysis of the Unlimited reelection vote in Venezuela, 23 june 2009.

[6] Riza Aditya Byoung Lee", Colin Boyd and Ed Dawson "Implementation Issues in secure e-voting schemes"at Joongbu University.

[7] S. Pravinthraja and Dr. K Umamaheshwari "Multimodal Biometrics for improving Automatic Teller Machine , December 2011.

[8] M.O Yinyeh, K.A. Gbolagade", Overview of Biometric electronic voting system in Ghana at july 2013.

[9] D.Ashok Kumar and T. Ummal Sariba Begum " A novel design of electronic voting system using fingerprint,International Journal of innovation technology and creative engineering, Vol1 pp.12-09-2011.

[10] Taigun Lee and Sung Kee Park and Mignon park, Department of electrical and electronic engineering "A new facial features and face detection method for Human Robot Iinteraction " IEEE april 2005.

[11] Paola Campaelli and Rafaella Lanzarotti " Localization of facial features and Fiducial points".

[12] Manju James and Shilpa Lizabeth George et al.,An Embedded Platform for Automtic Teller Machine using Finger- Vein Recognition System, International Journal of Industrial Electronics and Electrical Engineering Volume-2,Issue-2,Feb-2014

Society for Science and Education United Kingdom
AIVP Advances in Image and Video Processing
Volume 4, Issue 3
ISSN 2054-7412

Solving Sparsity Problem in Movie Based Recommendation System

Pratibha Bargah and Nitin Mishra
Rungta collage engineering & Technology Bhilai (c.g), India
pratibha23bargah@gmail.com, drnitinmishra10@gmail.com

ABSTRACT

Movie Recommendation is more useful in our community life due to its strength in giving enhanced entertainment. Recommendation system can advise a collection of movies to users depend on their choice, or the popularities of the movies. while, a set of motion picture recommendation systems have been planned, mainly of these either cannot advise a movie to the presented users powerfully.

In this paper we propose a solve the sparsity problem in movie recommendation system that has the ability to recommend movies to a new user as well as the others. It mines movie databases to collect all the important information, such as, popularity and attractiveness, required for recommendation. But in Recommendation system has many problems like sparsity , cold start , first Rater problem , Unusual user problem. K- mean clustering is the most successful method of Recommender System. K- means clustering also K-Means Clustering. The Algorithm K-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori.

Keywords: k-mean clustering, euclidean distance, k-mediod clustering, Harmonic Mean.

1 Introduction

Recommendation system are use for many purpose, it is the type of filtering the information it means that it is used for predict the Rating of item given by the user [1].

In Movie Recommendation system, Recommend the movie for watching or Rating the movie by the user. But in the field of movie Recommendation system it has many problems like cold start problem, sparsity problem etc[2] .there are three points for movie Recommendation system: -

Why:- Movie Recommendation system are Required because of movie information are overload.

Where :- used in social site, box offices and all types of area like bollybood, hollybood etc.

What :- suggest item to users for watching, Rating or purchasing the movie.if the users are interested.

2 The Existing Ranking Methods

2.1 K- Mean Clustering

k-means is used for solving clustering problem. It is the unsupervised leaning. No any classes are define previously. The process follow a straightforward and simple method to categorize a certain data set from side to side a definite amount of clusters (assume k clusters) predetermined apriori[1]. The main idea is to classify k centers, one

DOI: 10.14738/aivp.43.1997
Publication Date: 23rd July, 2016
URL: http://dx.doi.org/10.14738/aivp.43.1997

for each cluster. These center should be located in a craftiness method since of different position cause different result. So, the enhanced choice is to put them as a group as potential far absent beginning each other. The next step is to get each point belong to a known data set and transmit it to the adjacent core. while no point is awaiting, the primary step is finished and an untimely cluster period is complete[2]. on this spot we require to re-calculate k new centroids as barycenter of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be completed between the similar data set point and the adjacent novel center[8].

2.1.1 Steps of K- Mean Clustering

Let X = {x1,x2,x3,……..,xn} be the set of data points and V = {v1,v2,…….,vc} be the set of centers.

1) Randomly select 'c' cluster centers.
2) Calculate the distance between each data point and cluster centers.
3) Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers..
4) Recalculate the new cluster center using:

$$v_i = (1/c_i) \sum_{j=1}^{c_i} x_i$$

where, 'c_i' represents the number of data points in i^{th} cluster.

5) Recalculate the distance between each data point and new obtained cluster centers.
6) If no data point was reassigned then stop, otherwise repeat from step 3).

2.2 Euclidean Distance

In terms of mathematics, the Euclidean distance is the distance between two points in Euclidean space. With this distance, Euclidean space makes a metric space[3]. The associated norm is called the Euclidean norm. In our project first loaded the Rate matrix then Rating matrix and applying euclidean distance in both matrix .

In general, for an n-dimensional space, the distance is

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_i - q_i)^2 + \cdots + (p_n - q_n)^2}.$$

2.3 Harmonic Mean

In mathematics, the harmonic mean contain several types of average and Pythagorean mean. naturally, it is suitable for conditions when the normal of rates is required.

2.4 Two Number

the special case of just two numbers, $x_1$and x_2, the harmonic mean can be written

$$H = \frac{2x_1 x_2}{x_1 + x_2}.$$

In this special case, the harmonic mean is related to the arithmetic mean $A = \frac{x_1 + x_2}{2}$and the geometric mean $G = \sqrt{x_1 x_2}$,by

$$H = \frac{G^2}{A} = G \cdot \left(\frac{G}{A}\right).$$

In the above figure shows the process of solving sparsity problem in rating based Recommendation system. In the first step the dataset collect from IMDb(Internet Movie Database) ,the all require information of movie available .all information of movie and user are presents,in the first step we can gather the all data set that requires for solving sparsity problem in movie recommendation system[4]. Afterthat the process of Rating and Review are started we can generate the review and rating matrix and apply k- mean(Object clustering) clustering in both matrix . the k-mean clustering is simply solving clustering problem. It make cluster of similar object but it has no any predefined classes. And classification of Reviews is based on good ,bad and average comments of movies , we can take the 29*100 matrix for Rating and 29*3 are matrix for Review. Both data of matrix is convert into relational data using Euclidean distance .euclidean distance used to find the distance between two points in Euclidean distance. then apply the harmonic mean for calculate the average set of number[5] .

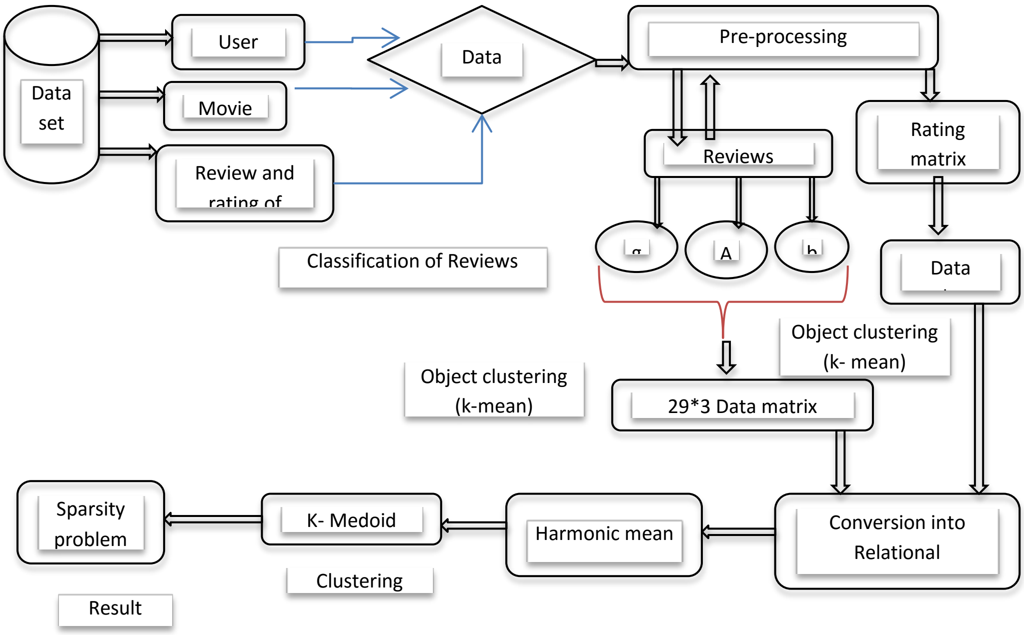


Figure 1: The Process for Solving Sparsity Problem in Rating Based Movie Recommendation System

In fig. shows steps of Solving Sparsity Problem In Recommendation System for movie recommendation system . in this steps show the dataset from IMDb. and using k- mean clustering for better movie Recommendation. In this steps it contain Review and Rating matrix. The Review matrix has 3 *29 matrix and Rating matrix contain 29*100 matrix. The Review are based on good comments, ba comments and average comments. and for generate the sparsity by applying the k- mean clustering. Afterthat we can apply Euclidean distance for converting Relational data . then for solving sparsity problem we apply Harmonic mean and for n*n Matrix you can apply the k- medoid clustering. Afterthat the sparsity problem are solved in Rating based movie Recommendation system. .

2.5 K – Mediod Clustering

The *k*-medoids algorithm is a similar as a k- mean clustering. *K-means* and *k*-medoids algorithms are splits in to some parts (breaking the dataset up into groups) and both challenge to decrease the distance between points labeled to the center of the cluster. *k*-medoid is a classical partitioning technique of clustering that divided the n object in to k cluster[10].

It is more robust to sound and outliers as well as k-means because it minimizes a sum of pair wise dissimilarities instead of a sum of squared Euclidean distances.A k-mediod can be defined as the object of a cluster whose average dissimilarity to all the objects in the cluster is minimal. i.e. it is a most centrally located point in the cluster.

2.5.1 Algorithm of K-Mediod clustering

The most common realization of *k*-medoid clustering is the **Partitioning Around Medoids (PAM)** algorithm. PAM uses a greedy search which may not find the optimum solution, but it is faster than exhaustive search. It works as follow.

1. Initialize: k randomly select(without replacement) from the n data points as the medoids
2. Each data point associate with the closest medoid.
3. While the configuration cost decreases:
    1. For each medoid m, for each non-medoid data point o:
        1. Swap m and o, recompute the cost (sum of distances of points to their medoid)
        2. If the total cost of the configuration increased in the previous step, undo the swap

Other algorithms than PAM have been suggested in the literature, including the following Voronoi iteration method.

1. Select initial medoids
2. Iterate while the cost decreases:
    1. In each cluster, make the point that minimizes the sum of distances within the cluster the medoid
    2. Reassign each point to the cluster defined by the closest medoid determined in the previous step.

3 User Query Intent and Storage of Tags

In the simplest technique used to reduce the sparsity of the user-item matrix, we simply insert a default rating, *d*, for appropriate items for which there exist no explicit ratings. ”Appropriate” is the key word here, meaning that it is wise to choose the matrix entries where the default ratings would be inserted. Nevertheless in where this techniqueis proposed. In recommendation system Yu Rong Xiao,Wen Hong Chenj the Chinese university of hongkong Year 2014b had done best performance.This authors already worked on Monte Carlo Method which is get working on future enhancement also

3.1 Cold start

Its very complicated to give the recommendation to new customer as his profile is empty and he is not rated any item over the available item . this is called cold start problem. And this problem is solved by combination of k – mean clustering, k-mediod clustering and Euclidean distance and harmonic mean.

3.2 Scalability

When increase the number of customer and items ,the system require more number of processing the information of the users and items for recommendation . many number of resource are used for determining the user with similar teste ,goods and similar description. This types of problem is also solved by various types of method used in this paper.

3.3 Sparsity

In online shopping, there are the more number of users rated the few number of items over the total number of available items. Using another approaches like collaborative filtering and association retrieval .In this approach generally created neighbourhood of the user according to their profile .if the user evaluate the few number of item , its the difficult to evaluate similar taste with users. sparsity is a problem they occur for lack of information.

4 K- Mediod Clustering Algorithm

4.1 K- Mediod Clustering Algorithm

The most common realization of *k*-medoid clustering is the **Partitioning Around Medoids (PAM)** algorithm. PAM uses a greedy search which may not find the optimum solution, but it is faster than exhaustive search. It works as follow.

4. Initialize: k randomly select(without replacement) from the n data points as the medoids
5. Each data point associate with the closest medoid.
6. While the configuration cost decreases:
    1. For each medoid m, for each non-medoid data point o:
        1. Swap m and o, recompute the cost (sum of distances of points to their medoid)
        2. If the total cost of the configuration increased in the previous step, undo the swap

Other algorithms than PAM have been suggested in the literature, including the following Voronoi iteration method.

3. Select initial medoids
4. Iterate while the cost decreases:
    1. In each cluster, make the point that minimizes the sum of distances within the cluster the medoid
    2. Reassign each point to the cluster defined by the closest medoid determined in the previous step.

5 Experiments and Analysis

The experiments are performed as follows:

- Initially, submit the data of movie from IMDb, and obtain the original Rating of movie results.
- Now, submit the original Rating of movie result for obtain the accurate rating of movie.
- Re-rank the Rating of movie results according to our algorithm.
- Compare the Rating of movie results with our algorithm.

5.1 Data Set

5.1.1 Original Movie Data Set

Actual data are given by the no. of users given the rating of movie, in this project we can divide in to three clustering of rating the range of 0 %to 33%, 34% to 65% and 66 to 100%. In this group we can easily classify the how no. of users rating the same data . in the actual data after observation the data give [3,13,13] clustering.

5.1.2 Object Data Set

Object data is classify by the k – mean clustering, k-means is used for solving clustering problem. It is the unsupervised leaning. No any classes are define previously. The process follow a straightforward and simple method to categorize a certain data set from side to side a definite amount of clusters (assume k clusters) predetermined apriori. The main idea is to classify k centers, one for each cluster. These center should be located in a craftiness method since of different position cause different result.

5.1.3 Relational Data Set

Relational data are given by k – mediod clustering, The *k*-medoids algorithm is a similar as a k- mean clustering. *K-means* and *k*-medoids algorithms are splits in to some parts (breaking the dataset up into groups) and both challenge to decrease the distance between points labeled to the center of the cluster.

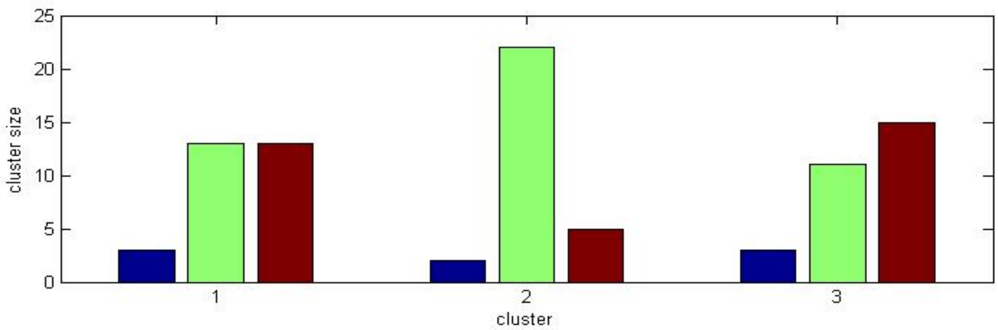


Figure 2:-Graph of result of sparsity problem solved

In this graph shows the result of sparsity problem solved, in this graph x label shows the three no. of clustering and y label shows the cluster size . In the first three bar shows the actual data in second cluster shows the object data and third cluster shows the Relational data, the cluster of actual data are [3,13,13] and relational data are[3,11,15]. So the actual data and relational data are adjacent value. So the k-mediod clusteing is the better than the k- mean clustering.

6 Conclusion

In this paper ,we aimed to solving sparsity problem in rating based movie recommendation system and improve the performance of movie Recommendation system.we use the k- mean clustering , k-mediod clustering and combination of harmonic mean and Euclidean distance method to solving sparsity probem. The effectiveness of the approach was evaluated experimently using data from IMDb Dataset . the experiment indicated that our approach solve the sparsity problem and achieved significantly better Recommendation quality then the other sparsity problem solving method.

REFERENCES

[1] Yibo chen et al progress of solving sparsity problem in recommendation system using association retrival .progress journal of computers vol 6 9september year:2011.

[2] Lalita sharma, anju gera et al progress of hybird approaches to reduce the sparsity problem . Progress on Mtech. Scholar BSAITM faridabad. vol 6 _july 2013.

[3] yu rong ,xiao wen, hong cheng, et al an monte carlo algorithm for cold start problem International world wide web conference committee April 7-11-2014.

[4] Mohammed mahmuda rahumen rahumen lecture, et al contextual recommendation system using multidimensional approch. International journal of intelligent information system august20,2013.

[5] Zuping liu sichuon et al recommendation algorithm based on user intrest ,advanced science and technology letters vol. 53, 2014.

[6] Manos papagelis, dimitris plexousakis Alleviating the sparsity problems of collaborative filtering using trust inferences Institutes of computer science , foundation for research and technology- hellas Years:2004.

[7] Andy yuanxue, jianzhong Qi , Solving the data sparsity problem in destination prediction University of Melbourne , Australia Year: 2013.

[8] Beau piccart, jan struf Alleviating the sparsity problem in collaborative filtering by using an adapted distance and a graph based method. IEEE computer technology Year:2007.

[9] Badrul sarwar, george karypis , joseph konstan Item based collaborative filtering recommedation algorithm . Department of computer science and engineering, University of Minnesota Year:2006 .

[10] Badrul sarwar, ,joseph konstan john riedl Using filtering agent to improve prediction quality in the gruoplen research collaborative filtering Department of computer science and engineering , University of minnesota year:2008 .

[11] Chrsistian Desrosiers, George Karypis. Solving the Sparsity Problem: Collaborative Filtering via Indirect Similarities. Technical Report. Department of Computer Science and Engineering University of Minnesota 4-192 EECS Building 200 Union Street SE Minneapolis, MN 55455-0159 USA. 2008.

[12] Zan Huang, Hsinchun Chen, et al. Applying Associative Retrieval Techniques to Alleviate the Sparsity Problem in Collaborative Filtering. ACM Transactions on Information Systems, Vol. 22, No. 1, January 2004, 116–142. http://dx.doi.org/10.1145/963770.963775.

[13] Sanghack Lee and Jihoon Yang and Sung-Yong Park, Discovery of Hidden Similarity on Collaborative Filtering to Overcome Sparsity Problem, Discovery Science, 2007.

[14] Rong Jin, Luo Si, et al. Collaborative Filtering with Decoupled Models for Preferences and Ratings. CIKM ’03, New Orleans, Louisiana, USA, November 3-8, 2003.

[15] Liu Jianguo, Zhou Tao, et al. Overview of the Evaluated Algorithms for the Personal Recommendation Systems. Complex System and Complexity Science. 2009, Vol.6, No.3, 1-10

.

Society for Science and Education
United Kingdom

AIVP Advances in Image and Video Processing

Volume 4, Issue 3
ISSN 2054-7412

An Efficient Algorithm for Forward Collision Warning Using Low Cost Stereo Camera & Embedded System on Chip

DOI: 10.14738/aivp.43.1997
Publication Date: 23rd July, 2016
URL: http://dx.doi.org/10.14738/aivp.43.1997

AIVP ADVANCES IN IMAGE AND VIDEO PROCESSING

SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM

VOLUME 4, ISSUE 3
ISSN 2054-7412

DOI: 10.14738/aivp.41.1922
Publication Date: 29st,February 2016
URL: http://dx.doi.org/10.14738/aivp.41.1922