International Journal of Computer science engineering Techniques-– Volume 5 Issue1, February 2020

RESEARCH ARTICLE OPEN ACCESS

Superiority Measure of Pupil Center Localization Algorithms based on Statistical Tests

Ibrahim Furkan Ince*, Huseyin Kusetogullari**
*(Nisantasi University, Turkey
Email: ibrahim.ince@nisantasi.edu.tr)
** (Blekinge Institute of Technology, Sweden
Email: huseyin.kusetogullari@bth.se)

Abstract:
Pupil center localization is a key requirement for any robust eye gaze tracking system. But vibration on the pupil center location deviates the desired accuracy of gaze on the eye tracking systems. To solve this problem, a lot of algorithms have been proposed in the literature.We cannot show the superiority of any algorithm without statistical tests. In this paper, the superiority measure of miscellaneous algorithms has been evaluated based on statistical tests.

Keywords —**Eye Gaze Tracking System, Pupil Center Localization, Human-Computer Interface.**

I. INTRODUCTION

Surveillance systems can detect and track objects using either laser scanned data points [1–5] or videos [6, 7]. The detection of abnormal (e.g., [8–17]) and normal (e.g., [18–21]) video events is a cardinal chore of a surveillance camera system. But camera movement and vibration of objects affect the detection of true events. For example, camera movements avoid true recognition of events due to ego-motion [22]. Pupil center localization is a must for robust eye gaze tracking systems. As vibration on the pupil center location defects the accuracy and precision [23] of gaze on the eye tracking systems, the stability on the pupil center location is an important metric for all gaze trackers. The automatic detection and tracking of human eyes specially pupils are a widely debated topic in the international scientific community. The applications of pupils' detection and tracking include advanced interfaces, control of the level of human attention, biometrics, gaze estimation, and early screening of neurological pathologies. Nowadays, human computer interface (HCI) [24] and brain computer interface [25] technologies are developing rapidly. Rather than keyboard and mouse inputs, HCI technologies focuses on various things e.g., human eyes, gestures, handshakes, body movements, and voices. Eye gaze estimation is the one of the most important HCI tools. The robust detection of pupil center locations would be the most difficult part, since the stability on the pupil localization is still a challenging issue in a low-cost paradigm.

Many algorithms have been developed for eye gaze tracking systems and eye gesture detections. For examples, Zhou et al. [26] proposed projection functions for eye detection. Kroon et al. [27] explained applicability of eye localization for face matching. Markus et al. [28] hinted eye pupil localization with an ensemble of randomized trees. Asadifard et al. [29] proposed an automatic adaptive center of pupil detection using face detection and cumulative distribution function analysis. Ponz et al. [30] addressed a topography-based detection of the iris centre using multiple-resolution images. Leo et al. [31] described an unsupervised approach for the accurate localization of the pupils in near-frontal facial images. Bai et al. [32] propose an eye location algorithm based on radial symmetry transform. Yang et al. [33] suggested an eye localization algorithm by means of multi-scale sparse dictionaries. Valenti et al. [34] hinted an eye center location algorithm using invariant isocentric patterns. Kim et al. [35]

estimated eye gaze using a webcam for region of interest detection. Timm et al. [36] placed an algorithm for eye centre localisation by means of gradients. Turkan et al. [37] localized human eye using edge projections. Campadelli et al. [38] proposed a precise eye localization through a general-to-specific model. Niu et al. [39] localized eye using 2D cascaded AdaBoost. Asteriadis et al. [40] detected an eye with the help of pixel to edge information. Hamouz et al. [41] located faces with feature-based affine-invariant. Cristinacce et al. [42] proposed a multi-stage approach to facial feature detection. Behnke [43] used hierarchical recurrent networks for learning face localization. Jesorsky et al. [44] detected face by dint of the Hausdorff distance. Ince et al. [45] proposed a low-cost pupil center localization algorithm based on maximized integral voting of circular hollow kernels.

The BioID images [46] were used to verify the performance of aforementioned algorithms. The normalized error of these methods is adopted as the performance measure for the estimated pupil center locations. The error is calculated by a normalized error formula which is introduced by Jesorsky et al. [44]. The error measure uses the biggest error on both eye estimation and is defined as $e = (max(d_l, d_r))/c_s$, where d_l and d_r are the Euclidean distances between the calculated and the true right and left eye centers. Maximum of the lapses is divided by the distance between the true eye centers c_s in spite of excluding the face size from the error measure. Based on application the value of e can be applied to compare different methods, e.g., for eye tracking applications a high performance for $e \leq 0.05$ is required, whereas for applications that use the overall eyeposition such as face matching comparing the performance for $e \leq 0.25$ will be more suitable. Table 1 compared the performance using $e \in \{0.05, 0.10, 0.15, 0.20, 0.25\}$. But from Table 1, it is extremely hard to find out the superiority of any algorithm. Besides, we cannot demonstrate the superiority of any algorithm without statistical tests.

In this paper, the superiority measure of aforementioned algorithms has been evaluated based on statistical tests. An example of complete statistical analysis can be found in [47]. Statistical tests for paired data are more sensitive than those of unpaired data (independent) because of more data information. We have designed experimental conditions to use statistical tests for paired data and reduce the number of trial runs. We have used the same initialized data for the set of methods at each trial run. Parametric tests which can use information of assumed data distribution. But parametric tests are more sensitive than those of non-parametric. Multiple comparisons with a control algorithm have commonly been used to statistically demonstrate that one approach is better than its alternatives. Non-parametric tests [48] can deal with probabilistic and non-probabilistic methods without any restriction. We have presented non-parametric test results with comparative study among various algorithms. We have performed tests adequate to multiple comparisons together with a set of post-hoc procedures to compare a control algorithm with other algorithms ($1 \times N$ comparisons).

The rest of the paper is organized as: Section 2. illustrates contrast estimation; Section 3. explains multiple comparisons with statistical tests; Section 4. concludes the paper.

II. Contrast Estimation

A contrast is a set of weights that explains a certain comparison over scores or averages. Contrast analysis is a relatively simple, but it serves as the building blocks of many statistical tests. In statistics, particularly in analysis of variance and linear regression, a contrast is a linear combination of variables (parameters or statistics) whose coefficients sum up to zero, allowing comparison of different treatments [49, 50]. For example, all coefficient values of the second column in the Table 2 and all coefficient values of the second row in the Table 2 and Table 3 are zero. Similarly, all coefficient values of the third column in the Table 2 and all coefficient values of the third row in the Table 2 and Table 3 are zero and so on. Despite its numerous advantages, contrast analysis is hardly implemented in a convenient manner in many statistical software packages.

III. Multiple Comparisons

A. z-score

A z-score is known as a standard score and it can be placed on a normal distribution curve. Z-scores range from -3 standard deviations up to +3 standard deviations. To use a z-score, we need to know the mean μ and also the population standard deviation σ. Basic z score of a sample x can be expressed as $z = (x-\mu)/\sigma$. For example, if x = 190, μ = 150, and σ = 25 then z score would be 1.6 standard deviations. A positive z-score tells the data point is above the mean. A negative z-score addresses that the data point is below mean. A z-score close to zero indicates that the data point is close to mean. The z core is used in financial sectors. For example, a lower z score of a company hints that the company is continuously moving towards insolvency or bankruptcy. However, we are more interested in multiple comparisons with statistical tests. We have performed tests adequate to multiple comparisons together with a set of post-hoc procedures for 1×N comparisons. We have demonstrated non-parametric test results with comparative study among algorithms. In conducting a hypothesis test, p-value of the test statistic and the level of significance α are required. Both p-value and α would be easily confused because they are both probabilities and numbers between zero and one. The number α tells us how extreme observed results must be to reject the null hypothesis of a significance test. The p-value of the test statistic is a way of saying how extreme that statistic is for our sample data. The smaller the p-value, the more unlikely the observed sample. In statistical significance testing, the p-value is the probability of obtaining a test statistic result at least as extreme as the one that was actually observed, assuming that the null hypothesis is true [51]. Critics of p-values point out that the criterion used to decide statistical significance is based on an arbitrary choice of level (often set at 0.05) [52]. If significance testing is applied to hypotheses that are known to be false in advance, a non-significant result will simply reflect an insufficient sample size; a p-value depends only on the information obtained from a given experiment.

B. Various Nonparametric Tests

Friedman test [53] and its derivatives (e.g., Iman-Davenport test [54]) are usually referred as one of the most important non-parametric tests for multiple comparisons. First of all, we have performed the Friedman test [53]. A usable characteristic of this test is that it ranks the algorithms from the best performing to the poorest one. However, it can only inform the researcher about the presence of differences among all samples of results com- pared. We have also performed two more alternatives the Friedman Aligned Ranks [55] and the Quade test [56], which differ in the way of computing the rankings and may lead to better results depending on the features of the experimental study considered. After the null-hypotheses have been rejected, we have proceeded with the post-hoc procedures to find the particular pairs of algorithms which produce differences. The post-hoc procedures comprise Bonferroni-Dunn's [57], Holm's [58], Hochberg's [59], Hommel's [60, 61], Holland's [62], Rom's [63], Finner's [64], and Li's [65], procedures in the case of 1 × N comparisons, and Nemenyi's [66], Shaffer's [67], and Bergmann-Hommel's [68] procedures in the case of N × N comparisons. The Bonferroni- Dunn's procedure [57] leads to the statement that the performance of two algorithms is significantly different if the corresponding average of rankings is at least as great as its critical difference.Holm's procedure [58] which checks sequentially hypotheses ordered according to their p-values from the lowest to the highest. All hypotheses for which p-value is less than the significance level α divided by the number of algorithms minus the number of a successive step are rejected. All hypotheses with greater p-values are supported. Holland's [62] and Finner's [64] procedures, also adjust the value of α in a step-down manner as Holm's step-down method [58] does. The Hochberg's procedure [59] operates in the opposite direction to the former, comparing the largest p-value with α, the next largest with α/2, and so on until it encounters a hypothesis it can reject. Rom [63] devised a modification to Hochberg's step-up procedure [59] to increase its power. In turn, Li [65] hinted a two-step rejection procedure.

C. Tools used for Nonparametric Tests

Statistical analysis of the results of experiments was performed using the available software1 and the open source JAVA program calculates multiple comparison procedures: Friedman [53], Iman et al. [54], Bonferroni et al. [57], Holm [58], Hochberg [59], Holland [62], Rom [63], Finner [64], Li [65], Shaffer [67], and Bergamnn et al. [68] tests as well as adjusted p-values. When all possible pairwise comparisons need to be performed, the easiest is the Nemenyi's procedure [66]. It assumes that the value of the significance level α is adjusted in a single step by dividing it merely by the number of comparisons performed. It is a very simple way but has little power. The Shaffer's static routine [67], in turn, follows the Holm's step-down method [58]. At a given stage, it rejects a hypothesis if the p-value is less than α divided by the maximum number of hypotheses which can be true given that all previous hypotheses are false. The Bergmann et al.'s [68] scheme is characterized by the best performance, but it is also the most sophisticated and so difficult to understand and computationally expensive. It consists in finding all the possible exhaustive sets of hypotheses for a certain comparison and all elementary hypotheses which cannot be rejected. The details of the procedure are described in Bergmann et al. [68], Garcia et al. [69] and the rapid algorithm for conducting this test in presented in Hommel et al. [61].

D. Multiple Comparison Nonparametric Tests

Table 4 depicts the average ranks computed using Friedman [53], Friedman Aligned Ranks [55], and Quade [56] non-parametric tests. To achieve the test results Friedman [53], Friedman Aligned Ranks [55], and Quade [56] non-parametric tests are applied to the average number of estimated cost fitness values. The purpose of using Friedman [53], Friedman Aligned Ranks [55], and Quade [56] non-parametric tests is to determine whether there are significant differences among the algorithms considered over given sets of data [56]. These tests obtain the ranks of the algorithms for each individual data set, i.e., the best performing algorithm receives the rank of 1, the second-best rank 2, etc. Here, we have not discussed the non-parametric methods, however, the mathematical equations and further explanation of the non-parametric procedures of Friedman [53], Friedman Aligned Ranks [55], and Quade [56] can be found in the literatures (e.g., Quade [56]).

Based on the obtained results in the Table 4, algorithm of Kim et al. [35] was the best performing algorithm of the comparison, with average rank of 1.59, 22.0, and 1.46 for Friedman [53], Friedman Aligned Ranks [55], and Quade [56] tests, respectively. This demonstrates that algorithm of Kim et al. [35] provides great performance to localize pupils. The second-best algorithm was Ponz et al. [30] with average rank of 3.59, 24.79, and 3.33 for Friedman [53], Friedman Aligned Ranks [55], and Quade [56] tests, respectively. The p-values computed through the statistics of each of the tests considered ($1.580422459\times10^{-10}$, 0.9997200783712 and$1.38749 \times 10^{-7}$).

E. Post-hoc procedures for 1 × N comparisons

The post-hoc procedures comprise Bonferroni-Dunn's [57], Holm's [58], Hochberg's [59], Hommel's [60, 61], Holland's [62], Rom's [63], Finner's [64], and Li's [65], procedures in the case of 1 × N comparisons; and Nemenyi's [66], Shaffer's [67], as well as Bergmann-Hommel's [68] procedures in the case of N × N comparisons. In these statistical analysis tests, multiple comparison post-hoc procedures considered for comparing the control algorithm Kim et al. [35] with the rest of algorithms. The results are shown by computing p-values for each comparison. Tables 5, 6, and 7 demonstrate the p-values obtained, using the ranks computed by the Friedman [53], Friedman Aligned Ranks [55], and Quade [56] non-parametric tests, respectively. Based on the computed results, all tests presented significant improvements of the Kim et al. [35] over its alternative algorithms for all the post-hoc procedures considered.

In brief, based on the results of aforementioned multiple comparisons with statistical tests it would be easy to make an exclusive conclusion that for pupil center localization the algorithm of Kim et al. [35] outperformed over its alternative algorithms. Future work would include all possible pairwise comparisons (i.e., N ×N comparisons). Future study would also include determination of computational complexity deeming various computer hardware implementations [70–73].

IV. CONCLUSIONS

We studied the superiority measure of miscellaneous pupil center localization algorithms based on statistical tests. We performed average rankings of various algorithms using the non-parametric statistical procedures, statistics, and p-values. Future study would include N × N comparisons of different algorithms.

Table 1. Performance comparison of various methods using e. Insufficient information is denoted by dash.

Methods	Accuracy						Learning Required?
	e 0:05	e 0:10	e 0:15	e 0:20	e 0:25	Average	
Zhou et al. [26]	-	-	-	-	94.81%	-	No
Kroon et al. [27]	65.02%	87.01%	-	-	98.78%	83.58%	Yes
Markus et al. [28]	89.90%	97.10%	-	-	99.70%	95.57%	Yes
Asadifard et al. [29]	47.00%	86.00%	89.00%	93.00%	96.00%	82.20%	No
Ponz et al. [30]	82.67%	94.97%	97.07%	98.78%	99.71%	94.64%	N0
Leo et al. [31]	77.01%	86.23%	84.45%	86.86%	90.01%	84.91%	No
Bai et al. [32]	37.03%	64.00%	84.87%	90.05%	96.01%	74.39%	No
Yang et al. [33]	89.60%	95.59%	96.24%	98.39%	99.10%	95.78%	Yes
Valenti et al. [34]	86.09%	91.67%	94.57%	97.09%	97.87%	93.46%	Yes
Kim et al. [35]	86.98%	96.76%	98.68%	99.66%	99.93%	96.40%	N0
Timm et al. [36]	82.50%	93.40%	95.20%	96.60%	98.00%	93.14%	No
Turkan¨ et al. [37]	18.62%	73.70%	94.21%	98.68%	99.59%	76.96%	Yes
Campadelli et al. [38]	62.01%	85.19%	87.58%	91.63%	96.09%	84.50%	Yes
Niu et al. [39]	75.01%	93.02%	95.79%	96.38%	96.98%	91.44%	Yes
Asteriadis et al. [40]	44.01%	81.68%	92.61%	96.02%	97.38%	82.34%	No
Hamouz et al. [41]	58.60%	75.01%	80.78%	87.62%	91.01%	78.60%	Yes
Cristinacce et al. [42]	56.99%	96.02%	96.49%	97.03%	97.09%	88.72%	Yes
Behnke [43]	37.01%	86.01%	95.02%	97.50%	98.01%	82.71%	Yes
Jesorsky et al. [44]	38.00%	78.79%	84.68%	87.22%	91.78%	76.09%	Yes
Ince et al. [45]	81.74%	90.99%	93.63%	95.22%	98.79%	92.07%	No

Table 2. Contrast weights Estimation (1 of 2)

	Zhou et al. [26]	Kroon et al. [27]	Markus et al. [28]	Asadifard et al. [29]	Ponz et al. [30]	Leo et al. [31]	Bai et al. [32]	Yang et al. [33]	Valenti et al. [34]	Kim et al. [35]
Zhou et al. [26]	0.000	-67.59	-76.82	-75.87	-83.91	-73.15	-71.01	-83.96	-82.11	-85.40
Kroon et al. [27]	67.59	0.000	-9.228	-8.281	-16.32	-5.551	-3.415	-16.36	-14.51	-17.81
Markus et al. [28]	76.82	9.228	0.000	0.9470	-7.091	3.677	5.813	-7.137	-5.286	-8.581
Asadifard et al. [29]	75.87	8.281	-0.9470	0.000	-8.038	2.730	4.866	-8.084	-6.233	-9.528
Ponz et al. [30]	83.91	16.32	7.091	8.038	0.000	10.77	12.90	-0.04500	1.806	-1.489
Leo et al. [31]	73.15	5.551	-3.677	-2.730	-10.77	0.000	2.136	-10.81	-8.962	-12.26
Bai et al. [32]	71.01	3.415	-5.813	-4.866	-12.90	-2.136	0.000	-12.95	-11.10	-14.39
Yang et al. [33]	83.96	16.36	7.137	8.084	0.04500	10.81	12.95	0.000	1.851	-1.444
Valenti et al. [34]	82.11	14.51	5.286	6.233	-1.806	8.962	11.10	-1.851	0.000	-3.295
Kim et al. [35]	85.40	17.81	8.581	9.528	1.489	12.26	14.39	1.444	3.295	0.000
Timm et al. [36]	82.15	14.56	5.329	6.276	-1.762	9.006	11.14	-1.808	0.04350	-3.252

Turkan et al. [37]	79.05	11.46	2.233	3.180	-4.858	5.910	8.046	-4.904	-3.053	-6.348
Campadelli et al. [38]	75.16	7.563	-1.665	-0.7175	-8.756	2.012	4.148	-8.801	-6.950	-10.25
Niu et al. [39]	81.70	14.11	4.879	5.826	-2.213	8.555	10.69	-2.258	-0.4070	-3.702
Asteriadis et al. [40]	77.80	10.21	0.9780	1.925	-6.113	4.655	6.791	-6.159	-4.308	-7.603
Hamouz et al. [41]	69.03	1.432	-7.796	-6.849	-14.89	-4.119	-1.983	-14.93	-13.08	-16.38
Cristinacce et al. [42]	82.15	14.56	5.333	6.280	-1.758	9.010	11.15	-1.804	0.04750	-3.248
Behnke et al. [43]	79.64	12.04	2.813	3.761	-4.278	6.490	8.626	-4.323	-2.472	-5.767
Jesorsky et al. [44]	71.07	3.477	-5.750	-4.803	-12.84	-2.073	0.06250	-12.89	-11.04	-14.33
Ince et al. [45]	83.33	15.74	6.508	7.455	-0.5830	10.19	12.32	-0.6280	1.223	-2.072

Table 3. Contrast Estimation (2 of 2)

	Timm et al. [36]	Turkan et al. [37]	Campad. et al. [38]	Niu et al. [39]	Asteriadis et al. [40]	Hamouz et al. [41]	Cristin. et al. [42]	Behnke et al. [43]	Jesorsky et al. [44]	Ince et al. [45]
Zhou et al. [26]	-82.15	-79.05	-75.16	-81.70	-77.80	-69.03	-82.15	-79.64	-71.07	-83.33
Kroon et al. [27]	-14.56	-11.46	-7.563	-14.11	-10.21	-1.432	-14.56	-12.04	-3.477	-15.74
Markus et al. [28]	-5.329	-2.233	1.665	-4.879	-0.9780	7.796	-5.333	-2.813	5.750	-6.508
Asadifard et al. [29]	-6.276	-3.180	0.7175	-5.826	-1.925	6.849	-6.280	-3.761	4.803	-7.455
Ponz et al. [30]	1.762	4.858	8.756	2.213	6.113	14.89	1.758	4.278	12.84	0.5830
Leo et al. [31]	-9.006	-5.910	-2.012	-8.555	-4.655	4.119	-9.010	-6.490	2.073	-10.19
Bai et al. [32]	-11.14	-8.046	-4.148	-10.69	-6.791	1.983	-11.15	-8.626	-0.06250	-12.32
Yang et al. [33]	1.808	4.904	8.801	2.258	6.159	14.93	1.804	4.323	12.89	0.6280
Valenti et al. [34]	-0.04350	3.053	6.950	0.4070	4.308	13.08	-0.04750	2.472	11.04	-1.223
Kim et al. [35]	3.252	6.348	10.25	3.702	7.603	16.38	3.248	5.767	14.33	2.072
Timm et al. [36]	0.000	3.096	6.993	0.4505	4.351	13.12	-0.004000	2.515	11.08	-1.179
Turkan et al. [37]	-3.096	0.000	3.898	-2.646	1.255	10.03	-3.100	-0.5805	7.983	-4.275
Campadelli et al. [38]	-6.993	-3.898	0.000	-6.543	-2.643	6.131	-6.997	-4.478	4.086	-8.173
Niu et al. [39]	-0.4505	2.646	6.543	0.000	3.901	12.67	-0.4545	2.065	10.63	-1.630
Asteriadis et al. [40]	-4.351	-1.255	2.643	-3.901	0.000	8.774	-4.355	-1.836	6.728	-5.530
Hamouz et al. [41]	-13.12	-10.03	-6.131	-12.67	-8.774	0.000	-13.13	-10.61	-2.046	-14.30
Cristinacce et al. [42]	0.004000	3.100	6.997	0.4545	4.355	13.13	0.000	2.519	11.08	-1.175
Behnke et al. [43]	-2.515	0.5805	4.478	-2.065	1.836	10.61	-2.519	0.000	8.564	-3.695
Jesorsky et al. [44]	-11.08	-7.983	-4.086	-10.63	-6.728	2.046	-11.08	-8.564	0.000	-12.26
Ince et al. [45]	1.179	4.275	8.173	1.630	5.530	14.30	1.175	3.695	12.26	0.000

Table 4. Average rankings of algorithms using the non-parametric statistical procedures, statistics, and p-values.

Various Approaches	Multiple Comparison Tests		
	Friedman [53]	Friedman Aligned Ranks [55]	Quade [56]
Kim et al. [35]	1.59	22.00	1.4666666666666666
Ponz et al. [30]	3.59	24.79	3.3333333333333333
Yang et al. [33]	4.20	26.40	4.266666666666666
Ince et al. [45]	5.00	27.80	4.933333333333334
Timm et al. [36]	7.79	31.79	7.799999999999999
Valenti et al. [34]	7.80	32.8	7.733333333333333
Cristinacce et al. [42]	7.80	43.59	6.6000000000000005
Markus et al. [28]	8.60	56.80	1.933333333333332
Niu et al. [39]	9.20	38.19	8.600000000000001
Behnke [43]	10.2	50.2	9.133333333333333
Turkan¨ et al. [37]	10.8	55.4	10.066666666666666
Asteriadis et al. [40]	12.59	58.39	12.333333333333332
Campadelli et al. [38]	13.0	57.19	12.999999999999999
Kroon et al. [27]	13.0	72.6	15.199999999999998
Asadifard et al. [29]	13.4	58.8	12.733333333333334
Leo et al. [31]	14.4	52.6	14.533333333333333
Bai et al. [32]	15.8	68.60	15.466666666666669
Hamouz et al. [41]	16.0	70.19	15.799999999999999
Jesorsky et al. [44]	16.2	69.19	15.866666666666667
Zhou et al. [26]	19.0	92.6	19.199999999999999
Statistics	7.210384356035096	4.542517260779414	5.10947559806229

p-value	1:580422459 10^{10}	0.9997200783712678	1:38749 10^{7}

Table 5. Adjusted p-values for Friedman test (Kim et al. [35] is the control method)

Index	Approaches	Unadjusted p	$1 \times N$ post-hoc procedures							
			One/Two step procedures		Step-down procedures			Step-up procedures		
			Bonf. [57]	Li [65]	Holm [58]	Hol. [62]	Finner [64]	Hoch. [59]	Hom. [60]	Rom [63]
1	Zhou et al. [26]	3.313792394282682E-6	6.296205549137096E-5	8.141531651881968E-6	6.296205549137096E-5	6.296017773765872E-5	6.296017773765872E-5	6.296205549137096E-5	6.296205549137096E-5	5.984711654612923E-5
2	Jesorsky et al. [44]	9.53955866823998E-5	0.001812516146965596	2.343208109642987 5E-4	0.00171712056028319 6	0.001715728919775894	9.058907357581969E-4	0.0017171205602831963	0.0015263293869183967	0.00163217825597070 9
3	Hamouz et al. [41]	1.1881328432232602E-4	0.0022574524021241943	2.918250705891214E-4	0.0020198258334795 42	0.0020179071163990425	9.058907357581969E-4	0.0020198258334795 42	0.0019010125491572163	0.0019199217668995929
4	Bai et al. [32]	1.475782273929768E-4	0.002803986320466559	3.6245091900972634E-4	0.002361251638287629	0.002358639917367622	9.058907357581969E-4	0.002361251638287629	0.002361251638287629	0.002244476764099805
5	Leo et al. [31]	6.240420423385724E-4	0.011856798804432876	0.0015308507620082198	0.009360630635078587	0.00931985101292121	0.002369288777704512	0.009360630635078587	0.008736588592740013	0.008897780850564592
6	Asadifard et al. [29]	0.001612243005572933	0.030632617105885725	0.003945462992442021	0.02257140207802106	0.02233638196532073	0.005096524637408 95	0.02257140207802106	0.019346916066875195	0.0214555449684690 5
7	Kroon et al. [27]	0.0023130835362667745	0.043948587189068716	0.005650860346216375	0.03007008597146807	0.029656277348110716	0.006265928689250 91	0.027757002435201295	0.02544391889893452	0.026385494649048 62
8	Campadelli et al. [38]	0.0023130835362667745	0.043948587189068716	0.005650860346216375	0.03007008597146807	0.029656277348110716	0.006265928689250 91	0.027757002435201295	0.02544391889893452	0.026385494649048 62
9	Asteriadis et al. [40]	0.003283460986069086	0.06238575873531263	0.008002520287214968	0.03611807084675994	0.035530912179900676	0.006919107990588924	0.03611807084675994	0.03611807084675994	0.03433405870615359
10	Turkan et al. [37]	0.01394009226052573	0.26486175294998887	0.03311500487259012	0.1394009226052573	0.13097351679605895	0.02631994891567435	0.1394009226052573	0.12546083034473157	0.13251844148672154
11	Behnke [43]	0.02153637927775721	0.40919120627738703	0.05025332778794284	0.193827413499814 9	0.1779425888335845	0.03690730404216791	0.193827413499814 9	0.15602478770691555	0.18426340843268835
12	Niu et al. [39]	0.04223618695936713	0.8024875522279754	0.09401361049738724	0.33788949567493703	0.29194434877428665	0.06604524343794205	0.33788949567493703	0.21118093479683564	0.321230471432417 1
13	Markus et al. [28]	0.0613688291394022	1.1660077536486417	0.1310211477385455	0.42958180397581536	0.35810379547094895	0.08840833740449894	0.39006196926728887	0.2454753165576088	0.3719303640212528
14	Valenti et al. [34]	0.09751549231682215	1.8527943540196208	0.19327780254591148	0.5850929539009329	0.4596955648831206	0.1299877205404255	0.39006196926728887	0.3900619692672886	0.3719303640212528
15	Cristinacce et al. [42]	0.09751549231682215	1.8527943540196208	0.19327780254591148	0.5850929539009329	0.4596955648831206	0.1299877205404255	0.39006196926728887	0.3900619692672886	0.3719303640212528
16	Timm et al. [36]	0.09751549231682222	1.8527943540196221	0.19327780254591162	0.5850929539009329	0.4596955648831206	0.1299877205404255	0.39006196926728887	0.39006196926728887	0.3719303640212528
17	Ince et al. [45]	0.36351472273644425	6.906779731992441	0.4717694845562627	1.0905441682093326	0.7421512164271273	0.39646197102097447	0.5929800980174268	0.5929800980174268	0.5929800980174268
18	Yang et al. [33]	0.4871309906878604	9.255488823069348	0.5447972983989485	1.0905441682093326	0.7421512164271273	0.5058080068311972	0.5929800980174268	0.5929800980174268	0.5929800980174268
19	Ponz et al. [30]	0.5929800980174268	11.266621862331109	0.5929800980174268	1.0905441682093326	0.7421512164271273	0.5929800980174268	0.5929800980174268	0.5929800980174268	0.5929800980174268

Table 6. Adjusted p-values for Aligned Friedman test (Kim et al. [35] is the control method)

Index	Approaches	Unadjusted p	$1 \times N$ post-hoc procedures							
			One/Two step procedures		Step-down procedures			Step-up procedures		
			Bonf. [57]	Li [65]	Holm [58]	Hol. [62]	Finner [64]	Hoch. [59]	Hom. [60]	Rom [63]
1	Zhou et al. [26]	1.192171117360519E-4	0.002265125122984986	9.819669117006678E-4	0.002265125122984986	0.0022626963890006735	0.0022626963890006735	0.002265125122984986	0.002265125122984986	0.002153061969926103
2	Kroon et al. [27]	0.00582057052253524	0.11059083992816955	Li	0.10477026940563432	0.09974423179588998	0.05394727101658925	0.10477026940563432	0.08875297054837829	0.09958750687130946
3	Hamouz et al. [41]	0.008616138652724509	0.16370663440176567	0.06632729305124911	0.14647435709631665	0.13680013981711991	0.05394727101658925	0.14647435709631665	0.12924207979086763	0.13922948296853263
4	Jesorsky et al. [44]	0.010099018654731064	0.1918813544398902	0.07686511282525955	0.16158429847569702	0.14990379638790252	0.05394727101658925	0.16158429847569702	0.15148527982096596	0.1535932034906191
5	Bai et al. [32]	0.011094121318547286	0.21078830505239843	0.08380426242336746	0.1664118197782093	0.1540895709842195	0.05394727101658925	0.1664118197782093	0.1664118197782093	0.15818334907707055
6	Asadifard et al. [29]	0.04489724653961347	0.8530476842526559	0.2701651085488825	0.6285614515545885	0.4743443227601356	0.1353799397969505	0.6145930527609194	0.414382731808425	0.5842166483127845
7	Asteriadis et al. [40]	0.047276388673916876	0.8982513848044207	0.28046615083796084	0.6285614515545885	0.4743443227601356	0.1353799397969505	0.6145930527609194	0.4254874980652519	0.5842166483127845
8	Campadelli et al. [38]	0.05505838680897174	1.046109349370463	0.3122186335170395	0.6607006417076609	0.4931730527414464	0.1353799397969505	0.636666441826339	0.49552548128074564	0.6052190074782021
9	Markus et al. [28]	0.0578787674387581	1.099696581336404	0.32304549405690763	0.6607006417076609	0.4931730527414464	0.1353799397969505	0.636666441826339	0.5209089069488229	0.6052190074782021
10	Turkan et al. [37]	0.06871113804195002	1.3055116227970502	0.3616406791193495	0.6871113804195002	0.5092683272801923	0.1353799397969505	0.6871113804195002	0.5594166879413738	0.6531874220002328
11	Leo et al. [31]	0.0953726732265185	1.8120807913038515	0.4401953143618538	0.8583540590386666	0.5942791524265874	0.1589703543379749	0.8583540590386666	0.762981385812148	0.8160003876884461
12	Behnke [43]	0.1243148195425275	2.3619815713080223	0.5061636126629152	0.99451855634022	0.654232634959099	0.18956597365054972	0.8787127722351161	0.8787127722351161	0.8787127722351161
13	Cristinacce et al. [42]	0.23911196978476784	4.543127425910589	0.6634642124912769	1.6737837884933748	0.8523459836124079	0.3292731150615177	0.8787127722351161	0.8787127722351162	0.8787127722351161
14	Niu et al. [39]	0.37728652677900215	7.168444008801041	0.7567316236374255	2.263719160674013	0.9416918081427397	0.47420096510679766	0.8787127722351161	0.8787127722351162	0.8787127722351161
15	Valenti et al. [34]	0.5561265072752963	10.56640363823063	0.8209554641556097	2.7806325363764817	0.9827695724974083	0.6425667088221046	0.8787127722351161	0.8787127722351162	0.8787127722351161
16	Timm et al. [36]	0.5932694244890134	11.272119065291255	0.8302622648850689	2.7806325363764817	0.9827695724974083	0.6564010324339653	0.8787127722351161	0.8787127722351162	0.8787127722351161
17	Ince et al. [45]	0.7519247073483286	14.286569439618244	0.8611021873525367	2.7806325363764817	0.9847331113841044	0.7894486437562722	0.8787127722351161	0.8787127722351162	0.8787127722351161
18	Yang et al. [33]	0.8104838351671617	15.399192868176073	0.8698315148538484	2.7806325363764817	0.9847331113841044	0.8272112368995155	0.8787127722351161	0.8787127722351162	0.8787127722351161
19	Ponz et al. [30]	0.8787127722351161	16.695542672467205	0.8787127722351162	2.7806325363764817	0.9847331113841044	0.8787127722351161	0.8787127722351161	0.8787127722351162	0.8787127722351161

Table 7. Adjusted p-values for Quade test (Kim et al. [35] is the control method)

Index	Approaches	Unadjusted p	$1 \times N$ post-hoc procedures							
			One/Two step procedures		Step-down procedures			Step-up procedures		
			Bonf. [57]	Li [65]	Holm [58]	Hol. [62]	Finner [64]	Hoch. [59]	Hom. [60]	Rom [63]
1	Zhou et al. [26]	0.1642620060602802	3.120978115145324	0.5852642522177417	3.120978115145324	0.9669377214909873	0.9669377214909873	0.8835990312795	0.8835990312795	0.8835990312795
2	Jesorsky et al. [44]	0.5852642522177417	4.915566006607508	0.6896925521004276	4.6568520062597445	0.9954318078664239	0.9669377214909873	0.8835990312795	0.8835990312795	0.8835990312795
3	Hamouz et al. [41]	0.2609252124463252	4.957579036480179	0.691511020092623	4.6568520062597445	0.9954318078664239	0.9669377214909873	0.8835990312795	0.8835990312795	0.8835990312795
4	Bai et al. [32]	0.27217712462798566	5.1713653679317275	0.7004438214273986	4.6568520062597445	0.9954318078664239	0.9669377214909873	0.8835990312795	0.8835990312795	0.8835990312795
5	Kroon et al. [27]	0.28141451358930275	5.346875758196752	0.7073996013311228	4.6568520062597445	0.9954318078664239	0.9669377214909873	0.8835990312795	0.8835990312795	0.8835990312795
6	Leo et al. [31]	0.3054299625008019	5.803169287515236	0.7240577679225864	4.6568520062597445	0.9954318078664239	0.9669377214909873	0.8835990312795	0.8835990312795	0.8835990312795
7	Campadelli et al. [38]	0.36568025805261695	6.947924902999722	0.7585449043522245	4.753843354684021	0.9973084220231605	0.9669377214909873	0.8835990312795	0.8835990312795	0.8835990312795
8	Asadifard et al. [29]	0.37686951659847445	7.160520815371014	0.7640220281064881	4.753843354684021	0.9973084220231605	0.9669377214909873	0.8835990312795	0.8835990312795	0.8835990312795
9	Asteriadis et al. [40]	0.3940443951904625	7.486843508618788	0.7719619435297604	4.753843354684021	0.9973084220231605	0.9669377214909873	0.8835990312795	0.8835990312795	0.8835990312795
10	Markus et al. [28]	0.4116846792262007	7.822008905297814	0.7795793747224725	4.753843354684021	0.9973084220231605	0.9669377214909873	0.8835990312795	0.8835990312795	0.8835990312795
11	Turkan et al. [37]	0.4999780982996281	9.499583867692934	0.8111535985749191	4.753843354684021	0.9980461048834213	0.9669377214909873	0.8835990312795	0.8835990312795	0.8835990312795
12	Behnke [43]	0.5476268487975553	10.404910127153551	0.8247046800605835	4.753843354684021	0.9982462221945301	0.9669377214909873	0.8835990312795	0.8835990312795	0.8835990312795
13	Niu et al. [39]	0.57582798544917	10.940731723534231	0.8318461427836065	4.753843354684021	0.9982462221945301	0.9669377214909873	0.8835990312795	0.8835990312795	0.8835990312795
14	Timm et al. [36]	0.6193705517916386	11.768040484041133	0.8417973984104762	4.753843354684021	0.9982462221945301	0.9669377214909873	0.8835990312795	0.8835990312795	0.8835990312795
15	Valenti et al. [34]	0.6230631178037839	11.838199238271894	0.842587394247067	4.753843354684021	0.9982462221945301	0.9669377214909873	0.8835990312795	0.8835990312795	0.8835990312795
16	Cristinacce et al. [42]	0.6872256024877363	13.057286447266991	0.8551554006663897	4.753843354684021	0.9982462221945301	0.9669377214909873	0.8835990312795	0.8835990312795	0.8835990312795
17	Ince et al. [45]	0.7856979635157874	14.92826130679996	0.870966515355534	4.753843354684021	0.9982462221945301	0.9669377214909873	0.8835990312795	0.8835990312795	0.8835990312795
18	Yang et al. [33]	0.8261729290236295	15.69728565144896	0.8765073284979741	4.753843354684021	0.9982462221945301	0.9669377214909873	0.8835990312795	0.8835990312795	0.8835990312795
19	Ponz et al. [30]	0.8835990312795	16.7883815943105	0.8835990312795	4.753843354684021	0.9982462221945301	0.9669377214909873	0.8835990312795	0.8835990312795	0.8835990312795

REFERENCES

[1] F. Galip, M. Caputcu, R. H. Inan, M. H. Sharif, A. Karabayir, S. Kaplan, M. Ozuysal, B. Sengoz, A. Guler, and S. Uyaver, “A novel approach to obtain trajectories of targets from laser scanned datasets,” in Int. Conf. on Computer and Information Technology (ICCIT), 2015, pp. 231–236.

[2] F. Galip, M. H. Sharif, M. Caputcu, and S. Uyaver, “Recognition of object from laser scanned data points using SVM,” in ICMIP, 2015, pp. 231–236.

[3] M. H. Sharif, H. Shehu, F. Galip, I. F. Ince, and H. Kusetogullari, “Object tracking from laser scanned dataset,” International Journal of Computer Science Engineering Techniques, vol. 3, no. 6, pp. 19–27, 2019.

[4] M. H. Sharif, “Particle filter for trajectories of movers from laser scanned dataset,” in Third Mediterranean Conference on Pattern Recognition and Artificial Intelligence (MedPRAI), Istanbul, Turkey, 2019, pp. 133–148.

[5] Z. Su, S. Li, H. Liu, and Z. He, “Tree skeleton extraction from laser scanned points,” in International Geoscience and Remote Sensing Symposium (IGARSS), 2019, pp. 6091–6094.

[6] M. H. Sharif, F. Galip, A. Guler, and S. Uyaver, “A simple approach to count and track underwater fishes from videos,” in Int. Conf. on Computer and Information Technology (ICCIT), 2015, pp. 347–352.

[7] M. Sharif, “A numerical approach for tracking unknown number of individual targets in videos,” Digital Signal Processing, vol. 57, pp. 106–127, 2016.

[8] N. Ihaddadene, M. H. Sharif, and C. Djeraba, “Crowd behaviour monitoring,” in Int. Conf. on Multimedia, 2008, pp. 1013–1014.

[9] M. H. Sharif, N. Ihaddadene, and C. Djeraba, “Covariance matrices for crowd behaviour monitoring on the escalator exits,” in Advances in Visual Computing, 4th International Symposium (ISVC), Las Vegas, NV, USA, December 1-3, 2008, pp. 470–481.

[10] M. H. Sharif, N. Ihaddadene, and C. Djeraba, “Crowd behaviour monitoring on the escalator exits,” in 2008 11th International Conference on Computer and Information Technology, 2008, pp. 194–200.

[11] S. A. Mahmoudi, M. H. Sharif, N. Ihaddadene, and C. Djeraba, “Abnormal event detection in real time video,” in International Workshop on Multimodal Interactions Analysis of Users in a Controlled Environ-ment, ICMI, 2008, pp. 1–4.

[12] M. H. Sharif and C. Djeraba, “A simple method for eccentric event espial using Mahalanobis metric,” in CIARP, Jalisco, Mexico, 2009, pp. 417–424.

[13] R. Mehran, A. Oyama, and M. Shah, “Abnormal crowd behavior detection using social force model,” in Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2009, pp. 935–942.

[14] M. H. Sharif and C. Djeraba, “Exceptional motion frames detection by means of spatiotemporal region of interest features,” in ICIP, 2009, pp. 981–984.

[15] M. H. Sharif, S. Uyaver, and C. Djeraba, “Crowd behavior surveillance using Bhattacharyya distance metric,” in CompIMAGE, Buffalo, NY, USA, 2010, pp. 311–323.

[16] M. H. Sharif, N. Ihaddadene, and C. Djeraba, “Finding and indexing of eccentric events in video emanates,” Journal of Multimedia, vol. 5, no. 1, pp. 22–35, 2010.

[17] M. H. Sharif and C. Djeraba, “An entropy approach for abnormal activities detection in video streams,” Pattern Recognition, vol. 45, no. 7, pp. 2543–2561, 2012.

[18] N. Imran, J. Liu, J. Luo, and M. Shah, “Event recognition from photo collections via pagerank,” in Inter-national Conference on Multimedia, 2009, pp. 621–624.

[19] M. H. Sharif and C. Djeraba, “Pedved: Pseudo euclidian distances for video events detection,” in Int. Sym. on Visual Computing (ISVC), 2009, pp. 674–685.

[20] M. H. U. Sharif, A. K. Saha, K. S. Arefin, and M. H. Sharif, “Event detection from video streams,” Inter-national Journal of Computer and Information Technology, vol. 1, no. 1, pp. 108–114, 2011.

[21] M. H. U. Sharif, S. Uyaver, and M. H. Sharif, "Ordinary video events detection," in CompIMAGE, 2012, pp. 19–24.

[22] Y. S. Erdem, F. Galip, I. F. Ince, and M. H. Sharif, "Estimation of camera ego-motion for real-time computer vision applications," International Journal of Scientific Research in Information Systems and Engineering, vol. 1, no. 2, pp. 1–6, 2015.

[23] M. H. Sharif, "An eigenvalue approach to detect flows and events in crowd videos," Journal of Circuits, Systems and Computers, vol. 26, no. 07, p. 1750110, 2017.

[24] I. S. MacKenzie, Human-Computer Interaction: An Empirical Research Perspective, 1st ed. Morgan Kaufmann, 2013.

[25] M. H. Sharif, R. A. T. Ramadan, and S. Uyaver, "Estimation of camera ego-motion for real-time computer vision applications," Sci.Int.(Lahore), vol. 31, no. 6, pp. 847–849, 2019.

[26] Z. H. Zhou and X. Geng, "Projection functions for eye detection," Pattern Recognition, vol. 37, no. 5, pp. 1049–1056, 2004.

[27] B. Kroon, A. Hanjalic, and S. M. P. Maas, "Eye localization for face matching: is it always useful and under what conditions?" in Proceedings of the 7th ACM International Conference on Image and Video Retrieval, CIVR 2008, Niagara Falls, Canada, July 7-9, 2008, 2008, pp. 379–388. [Online]. Available: http://doi.acm.org/10.1145/1386352.1386401

[28] N. Markus, M. Frljak, I. S. Pandzic, J. Ahlberg, and R. Forchheimer, "Eye pupil localization with an ensemble of randomized trees," Pattern Recognition, vol. 47, no. 2, pp. 578 – 587, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320313003294

[29] M. Asadifard and J. Shanbezadeh, "Automatic adaptive center of pupil detection using face detection and cdf analysis," in Proceedings of the International MultiConference of Engineers and Computer Scientists, vol. 1, 2010, p. 3.

[30] V. Ponz, A. Villanueva, L. Sesma, M. Ariz, and R. Cabeza, "Topography-based detection of the iris centre using multiple-resolution images," in 2011 Irish Machine Vision and Image Processing Conference, IMVIP 2011, Dublin, Ireland, September 8-9, 2011, 2011, pp. 32–37. [Online]. Available: https://doi.org/10.1109/IMVIP.2011.15

[31] M. Leo, D. Cazzato, T. D. Marco, and C. Distante, "Unsupervised approach for the accurate localization of the pupils in near-frontal facial images," J. Electronic Imaging, vol. 22, no. 3, p. 033033, 2013. [Online]. Available: https://doi.org/10.1117/1.JEI.22.3.033033

[32] L. Bai, L. Shen, and Y. Wang, "A novel eye location algorithm based on radial symmetry transform," in 18th International Conference on Pattern Recognition (ICPR 2006), 20-24 August 2006, Hong Kong, China, 2006, pp. 511–514. [Online]. Available: https://doi.org/10.1109/ICPR.2006.136

[33] F. Yang, J. Huang, P. Yang, and D. Metaxas, "Eye localization through multiscale sparse dictionaries," in Face and Gesture 2011, March 2011, pp. 514–518.

[34] R. Valenti and T. Gevers, "Accurate eye center location through invariant isocentric patterns," IEEE Trans-actions on Pattern Analysisand Machine Intelligence, vol. 34, no. 9, pp. 1785–1798, 2012.

[35] H. Kim, J. Kim, J. Lee, T. Lee, and R. Park, "Gaze estimation using a webcam for region of interest detection," Signal, Image and Video Processing, vol. 10, no. 5, pp. 895–902, 2016.

[36] F. Timm and E. Barth, "Accurate eye centre localisation by means of gradients," in International Conference on Computer Vision Theory and Applications (VISAPP), Vilamoura, Algarve, Portugal, 5-7 March, 2011, 2011, pp. 125–130.

[37] M. Turkan, M. Pardas, and A. E. C¸etin, "Human eye localization using edge projections," in VISAPP 2007: Proceedings of the SecondInternational Conference on Computer Vision Theory and Applications, Barcelona, Spain, March 8-11, 2007 - Volume 1, 2007, pp. 410–415.

[38] P. Campadelli, R. Lanzarotti, and G. Lipori, "Precise eye localization through a general-to-specific model definition," in Proceedings ofthe British Machine Vision Conference 2006,

Edinburgh, UK, September 4-7, 2006, 2006, pp. 187–196. [Online]. Available: https://doi.org/10.5244/C.20.20

[39] Z. Niu, S. Shan, S. Yan, X. Chen, and W. Gao, “2D cascaded AdaBoost for eye localization,” in 18th International Conference on PatternRecognition (ICPR 2006), 20-24 August 2006, Hong Kong, China, 2006, pp. 1216–1219. [Online]. Available: https://doi.org/10.1109/ICPR.2006.1194

[40] S. Asteriadis, N. Nikolaidis, A. Hajdu, and I. Pitas, “An eye detection algorithm using pixel to edge infor-mation,” in Int. Symp. on Control, Commun. and Sign. Proc (ISCCSP), 2006.

[41] M. Hamouz, J. Kittler, J. Kamarainen, P. Paalanen, H. Kalvi¨ainen,¨ and J. Matas, “Feature-based affine-invariant localization of faces,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 9, pp. 1490–1495, 2005.

[42] D. Cristinacce, T. F. Cootes, and I. M. Scott, “A multi-stage approach to facial feature detection,” in British Machine Vision Conference, BMVC 2004, Kingston, UK, September 7-9, 2004. Proceedings, 2004, pp. 1–10. [Online]. Available: https://doi.org/10.5244/C.18.30

[43] S. Behnke, “Learning face localization using hierarchical recurrent networks,” in Artificial Neural Networks - ICANN 2002, InternationalConference, Madrid, Spain, August 28-30, 2002, Proceedings, 2002, pp. 1319–1324.

[44] O. Jesorsky, K. J. Kirchberg, and R. W. Frischholz, “Robust face detection using the Hausdorff distance,” in International Conference onAudio-and Video-Based Biometric Person Authentication. Springer, 2001, pp. 90–95.

[45] I. F. Ince, Y. S. Erdem, F. Bulut, and M. H. Sharif, “A low-cost pupil center localization algorithm based on maximized integral voting of circular hollow kernels,” Comput. J., vol. 62, no. 7, pp. 1001–1015, 2019.

[46] BioID, “Bioid face database,” 2017, https://www.bioid.com/About/a-Face-Database.

[47] H. Kusetogullari, M. H. Sharif, M. S. Leeson, and T. Celik, “A reduced uncertainty-based hybrid evolutionary algorithm for solvingdynamic shortest-path routing problem,” Journal of Circuits, Systems, and Computers, vol. 24, no. 5, 2015.

[48] Z. T. B. Trawinski, M. Smetek and T. Lasota, “Nonparametric statistical analysis for multiple comparison of machine learning regressionalgorithms,” Int. J. Ap. Mat. Com., vol. 22, pp. 867–881, 2012.

[49] G. Casella and R. L. Berger, Statistical Inference, 2nd ed. Cengage Learning, 2001.

[50] G. Casella, Statistical design, 1st ed. Springer-Verlag New York, 2008.

[51] S. Goodman, “Toward evidence-based medical statistics. 1: The p value fallacy,” Ann. Intern. Med., vol. 130, pp. 995–1004, 1999.

[52] M. B. T. Sellke and J. Berger, “Calibration of p values for testing precise null hypotheses,” Am. Stat., vol. 55, pp. 62–71, 2001.

[53] M. Friedman, “The use of ranks to avoid the assumption of normality implicit in the analysis of variance,” J. Am. Stat. Assoc., vol. 32, pp. 675–701, 1937.

[54] R. Iman and J. Davenport, “Approximations of the critical region of the Friedman statistic,” Commun. Stat. Theor. M., vol. 18, pp.571–595, 1980.

[55] J. Hodges and E. Lehmann, “Ranks methods for combination of independent experiments in analysis of variance,” Ann. Stat., vol. 33, pp. 482–497, 1962.

[56] D. Quade, “Using weighted rankings in the analysis of complete blocks with additive block effects,” J. Am. Stat. Assoc., vol. 74, pp. 680–683, 1979.

[57] O. Dunn, “Multiple comparisons among means,”J. Am. Stat. Assoc., vol. 56, pp. 52–64, 1961.

[58] S. Holm, “A simple sequentially rejective multiple test procedure,” Scand. J. Stat., vol. 6, pp. 65–70, 1979.

[59] Y. Hochberg, “A sharper bonferroni procedure for multiple tests of significance,” Biometrika, vol. 75, pp. 800–803, 1988.

[60] G. Hommel, “A stagewise rejective multiple test procedure based on a modified bonferroni test,” Biometrika, vol. 75, pp. 383–386, 1988.

[61] G. Hommel and G. Bernhard, “A rapid algorithm and a computer program for multiple

test procedures using procedures using logicalstructures of hypotheses," Comput. Meth. Prog. Bio., vol. 43, pp. 213–216, 1994.

[62] M. Holland, "An improved sequentially rejective bonferroni test procedure," Biometrics, vol. 43, pp. 417– 423, 1987.

[63] D. Rom, "A sequentially rejective test procedure based on a modified bonferroni inequality," Biometrika, vol. 77, pp. 663–665, 1990.

[64] H. Finner, "On a monotonicity problem in step-down multiple test procedures," J. Am. Stat. Assoc., vol. 88, pp. 920–923, 1993.

[65] J. Li, "A two-step rejection procedure for testing multiple hypotheses," J. Stat. Plan. Infer., vol. 138, pp. 1521–1527, 2008.

[66] P. Nemenyi, "Distribution-free multiple comparisons," Ph.D. dissertation, Princeton University, 1963.

[67] J. Shaffer, "Modified sequentially rejective multiple test procedures," J. Am. Stat. Assoc., vol. 81, pp. 826– 831, 1986.

[68] G. Bergmann and G. Hommel, "Improvements of general multiple test proceduresfor redundant systems of hypotheses," in Multiple HypothesesTesting, E. S. P. Bauer, G. Hommel, Ed. Springer, 1988, pp. 100–115.

[69] S. Garcia and F. Herrera, "An extension on "statistical comparisons of classifiers over multiple data sets" for all pairwise comparisons,"J. Mach. Learn. Res., vol. 9, pp. 2677–2694, 2008.

[70] M. H. Sharif, A. Basermann, C. Seidel, and A. Hunger, "High-performance computing of $1/\sqrt{x_i}$ and $exp(\pm x_i)$ for a vector of inputs $x_i$onAlpha and IA-64 CPUs," Journal of Systems Architecture - Embedded Systems Design, vol. 54, no. 7, pp. 638–650, 2008.

[71] E. R. Blem, H. Esmaeilzadeh, R. S. Amant, K. Sankaralingam, and D. Burger, "Multicore model from abstract single core inputs," ComputerArchitecture Letters, vol. 12, no. 2, pp. 59–62, 2013.

[72] M. H. Sharif, "High-performance mathematical functions for single-core architectures," Journal of Circuits, Systems and Computers, vol.23, no. 04, p. 1450051, 2014.

[73] J. Hennessy and D. Patterson, Computer Architecture: A Quantitative Approach, 6th ed. The Mor-gan Kaufmann Series in Computer Architecture and Design, 2017.