Classifying visual objects with the consistency-driven pairwise comparisons method

Waldemar W. Koczkodaj† Nicolas Robidoux‡
Ryszard Tadeusiewicz§
†*Corresponding author: wkoczkodaj cs laurentian ca*
†‡*Laurentian University, Department of Mathematics and Computer Science, Sudbury ON P3E 2C6*
§*Computer Science, AGH University of Science and Technology*

August 28, 2008

Abstract. The classification of the various image features or visual objects can be done by the consistency-driven pairwise comparisons method based on their relative importance. A key issue in the proposed approach is a weight based synthesis for combining various image features. When compared with the traditional experience-based linear assignment method, the proposed approach is more effective and easy to communicate.

Key words: machine vision, visual object, pairwise comparisons, inconsistency analysis, content-based image retrieval.

1. The method of pairwise comparisons

Kato introduced “content-based image retrieval (CBIR)” in 1992 [6] and Condorcet used pairwise comparisons even earlier (in 1785) in [1] so these two methods are not new. The novelty of the present approach lies in the combined use of the two methods.

Let us begin with an important case that highlights the dynamics associated with machine graphics and visual human identification. A person can be identified by a number of features or characteristics. Such features include the face specification (with some more specific sub-features), body shape, height, or even hair. Needless to say, these features do not equally contribute to a successful identification a person. Setting the relative importance of individual features is not a trivial task, especially if there are many such features. This is why we would like to bring to the attention of the machine graphics and vision community the pairwise comparisons method.

The pairwise comparisons method was used by Condorcet in 1785 [1] for his election method in which voters rank candidates in order of preference. A Condorcet method is a voting system which uses matrices for particular pairwise comparisons with rows representing each candidate as a runner and columns representing each candidate as an opponent. However, it was Fechner who specified pairwise comparisons as a scientific method in 1860 although only from the psychometric perspective (see [2]). Thurstone provided a mathematical analysis of this method and called it the “law of comparative judgments" (LCJ) in 1927 (see [3]). LCJ can be used to scale a collection of stimuli based on simple comparisons between stimuli taken two at a time. Although Thurstone referred to it as a law, it can be more appropriately identified as a measurement model which could be of great use for machine vision. This model allows us to synthesize diverse features involved in machine vision, such as the above mentioned human identification problem.

The next milestone in pairwise comparisons was the introduction of a hierarchy in [4]. The hierarchy reduces the number of comparisons from $\Theta(n^2)$ to $\Theta(n \ln n)$, making it finally applicable to a wide variety of problems. For example, a moderate case with 49 features would require 1,176 comparisons without a hierarchy and only 168 comparisons if these 49 features are arranged into hierarchy by grouping seven features.

In the context of human visual identification, the introduction of a hierarchical structure can express, for example, hair identification as one of the nodes with color, volume, hair split, hair line, and shape as its children. Tab. 1 illustrates an example of a hierarchical model for a human identification. The ingenuity of the pairwise comparisons

Tab. 1. Hierarchical model of a visual object identification (example)

Visual merit for human identification			
height	face	hair	body shape
		↓	
		color	
		volume	
		line	
		split	
		shape	

method can be expressed by the old adage "If you want to eat an elephant, do it in small bites." By common sense, comparing features two at a time is easier than doing so all at once. The practical ramifications of this approach is even more poignant in situations where direct measurements are impossible. No one questions the practicality of measurements by length (such as a meter or foot) or by mass/weight (kg or pounds) since they are in common use. We have become so accustomed to having standards that sometimes we find it difficult to imagine anything without a standard measure. In the case of a cancerous tumor, it may be shade gradation or general shape. Although we may not be able to express the exact number of units of a general shape, we may still express preference of one shape when compared with another shape.

We will name the features or criteria $C_1, C_2, \ldots, C_n$ for recognizing visual objects. The pairwise comparison method does not impose any limit on the number of criteria. Setting the maximum on one level to seven is a widely accepted heuristic since seven items gives 21 distinct pairs to compare. The model is shown on the enclosed screen image after entering it into the Concluder system, preliminary rearranging of the attributes and relating them to each other.

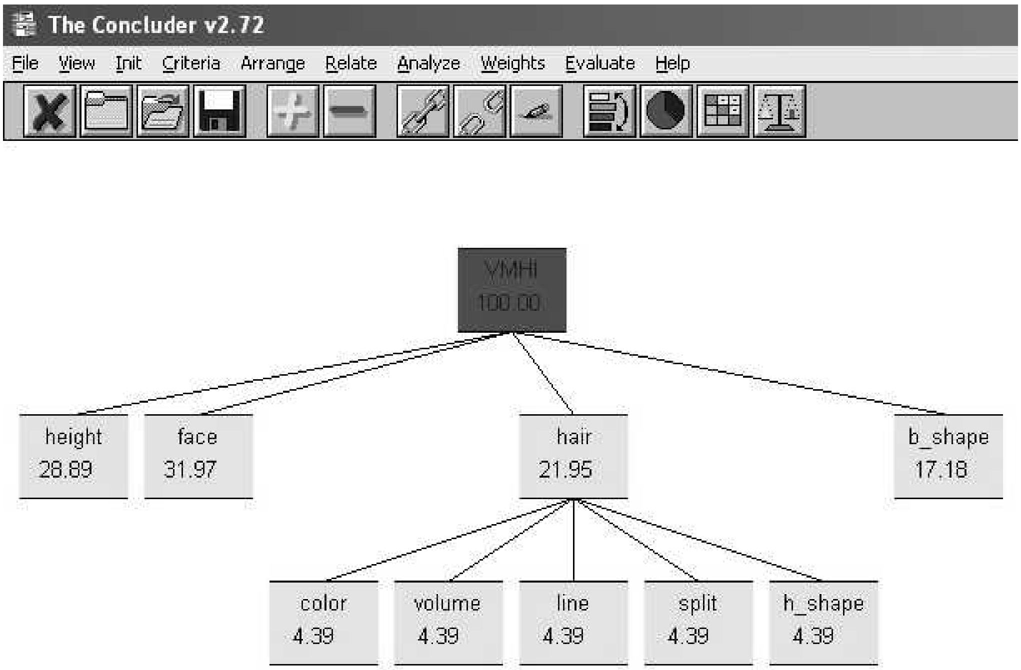

Fig. 1. Screen image of the model

The first step of pairwise comparisons is to establish the relative preference of two criteria for situations in which it is impractical (or meaningless) to provide the absolute estimations of the criteria. The relative comparison coefficients a_{ij} for criteria $C_1, C_2, \ldots, C_n$ are expected to satisfy $a_{ii} = 1$ and $a_{ij} = 1/a_{ji}$. The first constraint is related to comparing a given attribute with itself. The second constraint is a consequence of the obvious fact that $x/y = 1/(y/x)$ for $x, y \neq 0$.

A scale from 1 to 5, presented by Tab. 2, is used for expressing the importance of one criterion over another criterion in a pair. Other scales also exists but all of them are isomorphic.

Tab. 2. Comparison scale

Code	Definition of intensity or importance
1	Equal or unknown importance
2	Weak importance of one over another
3	Moderate to essential importance
4	Demonstrated importance
5	Absolute importance
2.2 etc	Intermediate importance

2. Classifying visual objects by the consistency-driven pairwise comparisons method

It is not our goal to present the entire consistency-driven pairwise comparisons method here, but to demonstrate how this approach can be applied to the classification of visual objects. However, it is necessary to note that the partial assessments of all pairs, entered into the pairwise comparisons matrix, need to be synthesized into weights which can be subsequently used for all objects to be classified or visualized. The solution to the above pairwise comparisons matrix is a normalized vector of geometric means:

$$V = [v_1, v_2, ..., v_n] \text{ where } v_i = \sqrt[n]{\prod_j a_{ij}}. \tag{1}$$

Not only is the vector of geometric means simpler to compute than an eigenvector but geometric means have obvious interpretation as the arithmetic means in the corresponding space obtained by logarithmic mapping. It does make sense to get "average" of partial comparisons for getting the global weights. The eigenvector's lack of interpretation negatively impacts the confidence in a decision making method from the application point of view.

3. Inconsistency analysis

In the pairwise comparisons method, stimuli (for example, criteria or alternatives) are presented in pairs to one or more experts. It is necessary to evaluate individual alternatives, derive weights for the criteria, construct the overall rating of the alternatives, and finally identify the best alternative. Let us denote the stimuli by $A_1, A_2, \ldots, A_n$ (n is the number of compared stimuli), their actual weights by $\gamma_1, \gamma_2, \ldots, \gamma_n$, and the matrix of the ratios of all weights by $\Gamma = [\gamma_i/\gamma_j]$. The matrix of pairwise comparisons $\mathbf{A} = [a_{ij}]$ represents the intensities of assessments between individual pairs of alternatives (A_i versus A_j, for all $i, j = 1, 2, ...n$) chosen usually from a given scale. The elements a_{ij} are considered to be estimates of the ratios γ_i/γ_j, where $\boldsymbol{\gamma}$ is the vector of actual weights of the stimuli (which is what we want to find). All the ratios are positive and satisfy the reciprocity property $a_{ij} = 1/a_{ji},\ i, j = 1, 2, \ldots, n$. Saaty's eigenvector solution of $\mathbf{A}\boldsymbol{\gamma} = \lambda\boldsymbol{\gamma}$ always exists if the consistency (or transitivity) condition $a_{ij}a_{jk} = a_{ik}$ for $i, j, k = 1, \ldots, n$ is satisfied. More details about the problem of inconsistent assessments and definitions of inconsistency can be found, for example, in [4, 7, 12].

The practical challenge confronting the pairwise comparisons method comes from the lack of consistency of the pairwise comparisons matrices. Let us assume that we have decided that the relative importance of considered criteria are as given in the following matrix:

$$\mathbf{A} = \begin{bmatrix} 1 & 1.5 & \boxed{2.5} & 3 & \boxed{3} \\ 2/3 & 1 & \underline{1.5} & \underline{1.5} & 3 \\ 2/5 & 2/3 & 1 & \underline{1.5} & \boxed{2} \\ 1/3 & 2/3 & 2/3 & 1 & 1.5 \\ 1/3 & 1/3 & 1/2 & 2/3 & 1 \end{bmatrix} \tag{2}$$

It is also shown on the second screen image (Fig. 2).

In essence, we need to find a consistent $n \times n$ matrix $\mathbf{B}$ which differs from matrix $\mathbf{A}$ "as little as possible". A possible solution to this problem was proposed by Saaty [4] as the eigenvector of $\mathbf{A}$ corresponding to the largest eigenvalue of $\mathbf{A}$. However, the geometric means method produces results with high accuracy when compared to the eigenvalue method (as evidenced by a Monte Carlo study with ten million cases [8]), and is simpler to use. There is, however, a strong relationship between accuracy and consistency. This is why the inconsistency analysis is the main focus of the consistency-driven approach.

Our conjecture is that in making comparative assessments of intangible criteria (such as body shape to hair), we face not only imprecise or inexact knowledge, but also the inconsistency of our own subjective assessments. More importantly, the improvement of knowledge elicitation by controlling the inconsistency of often highly subjective assessments is not only desirable, but even needed for the refinement of our own expertise.

In practice, inconsistent assessments are unavoidable when at least three criteria are independently compared against each other. For example, let us look closely at the three boxed criteria in matrix A. Eq. 3 gives inconsistency 0.4, which is considered (as a heuristic proposed in [7]) to be too high for most practical cases. So, the most inconsistent triad has to be localized for reconsideration of our assessments. By changing the relative importance of C_1 against C_5 from 3 to 4, we reduce the local inconsistency (of the triad in gray) to 1/3 since there is another triad, shown by the underlined 1.5 values, which has inconsistency indicator $ii = 1/3$.

The inconsistency concept is easier to explain by using three objects A, B, and C and considering their areas. Let us assume the following initial assessments: A/B is 2, B/C is 3, and A/C is 5. Evidently, the above assessments violate $A/C = A/B \times B/C$, so we may try to correct the last assessment to 6 since 2×3 gives 6. Unfortunately, we do

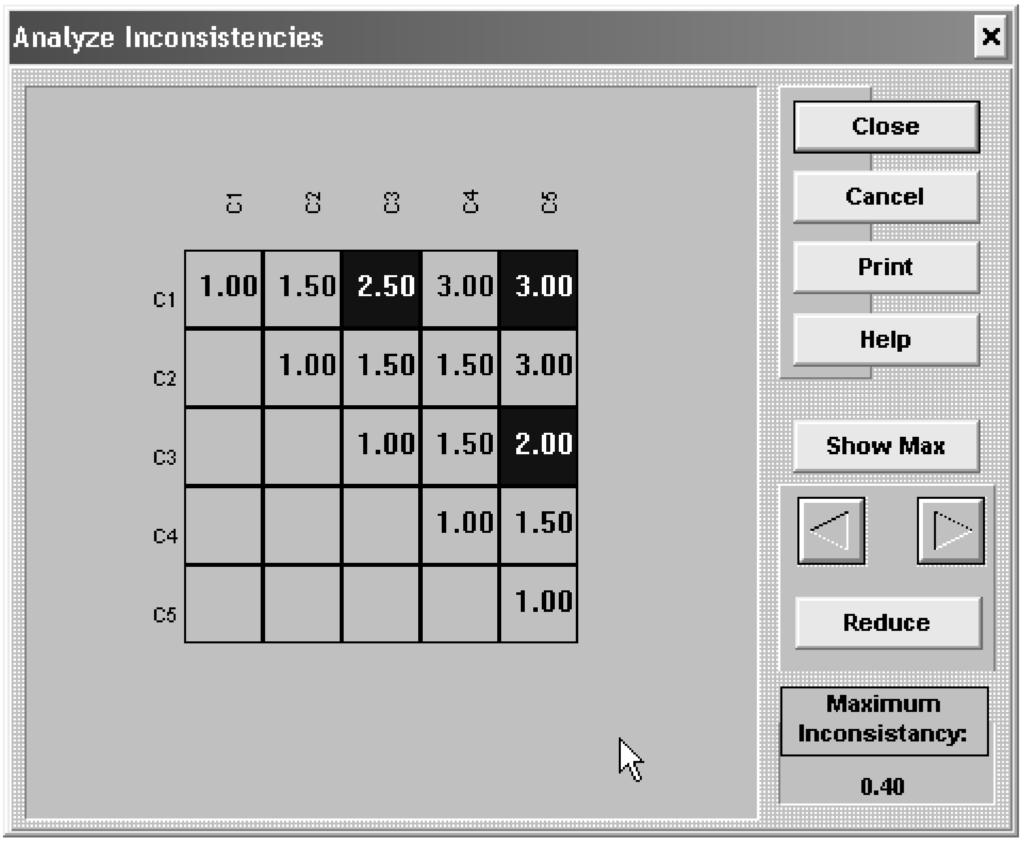

Fig. 2. Screen image with the pairwise comparisons matrix

not know which assessment is inaccurate. In particular, as is frequently encountered in practice, each original assessment might have been (and usually is) just a little inaccurate. In practice, it is safe to assume that every assessment is somewhat inaccurate. Full consistency can also be obtained by, for example, changing 3 to 2.5, since 2×2.5 gives 5.

The eigenvalue-based inconsistency (introduced in [4]) is a *global* characteristic of a matrix and as such, it cannot localize the inconsistency. The distance-based inconsistency (introduced by Koczkodaj in [7]) was independently analyzed and compared with the eigenvalue-based inconsistency in [12]. According to [12], the distance-based inconsistency localizes the most inconsistent triad (or triads) of objects. Basically, the distance-based inconsistency indicator is defined as the maximum over all triads $\{a_{ik}, a_{kj}, a_{ij}\}$ of elements of $\mathbf{A}$ (with all indices i, j, k distinct) of their inconsistency indicators, which in turn are defined as follows:

$$ii = \min\left(\left|1 - \frac{a_{ij}}{a_{ik}a_{kj}}\right|, \left|1 - \frac{a_{ik}a_{kj}}{a_{ij}}\right|\right). \tag{3}$$

Excluding the less common blind comparisons (which was handled separately in [11]), the minimal number of objects which may cause inconsistency is three. Comparing two objects often results in inaccuracy, that is, inexact knowledge, however, it does not involve inconsistency.

The distance-based inconsistency is the minimum distance from three "ideal" triads with no inconsistency when the "third" value is substituted using the consistency condition $a_{ij} \times a_{jk} = a_{ik}$. Since we are not in a position of saying which ratio is incorrect a priori, all three assessments must be reconsidered before we attempt finding a consistent approximation for a given pairwise comparisons matrix. For details related to inconsistency analysis, see [7, 12]. The stress on localizing the most inconsistent assessments is expressed by adding the *consistency-driven* to the name of the method since it is easier to remedy implications of an error (in judgment) when we are able to localize it. Properties of both eigenvalue-based and distance-based inconsistencies were examined by a study published in the Journal of Global Optimization available on line at www.springerlink.com/content/v2x539n054112451 (soon to be published as a hard copy) with a clear conclusion that the distance-based inconsistency is superior because of the localizing property. Basically, when we can find an error, we have good chance to fix it.

In general, there is no practical reason to continue decreasing the inconsistency indicator to zero. Only high values of the inconsistency indicator are considered harmful. A very small value or zero may indicate data doctoring rather than entering honest assessments. We know that "to err is human" but when it is done, it is better to know where it could be hence the need for inconsistency analysis and localization of the most inconsistent assessments.

Inconsistency analysis may look complicated, but the software developed for this analysis (available from the first author's web page) is facilitating it. By decreasing or increasing values in a triad (displayed by the software), one develops a very good orientation quickly.

In the case of the matrix shown in Eq. 2, after consistency analysis, the final weights are calculated from Eq. 3 and the normalized vector is

$$[0.3771, 0.2380, 0.1685, 0.1304, 0.0860]. \tag{4}$$

These weights, corresponding to criteria $C_1, C_2, \ldots, C_5$, should be applied (as multipliers) to the values of attributes for each considered visual object, to establish a merit index for each visual object that will be used for making an identification decision.

Intuitively it is obvious that the "two at a time" approach has chances to be more accurate than the method of assessing "everything at once". However, to show that the pairwise comparisons method is superior to the common sense "by an expert's eye" approach is not entirely a trivial task since there are many hurdles to overcome. At the current stage of pairwise comparisons theory, there is no possibility of proving, or disproving, by analytical means which method is superior. The necessity of using computer technology for Monte Carlo experiments in [9] and [10] may explain why the problem of accuracy had not been properly addressed in the 1950's or 1960's when most of the theoretical work on the pairwise comparisons method took place. However, the drop of estimation error of lengths of randomly generated bars from approximately 15% to 5% [9, 10] is a clear indication of potential gains in precision by using the pairwise comparisons.

4. A fairly realistic example of using the proposed classification for machine vision

There is no universally accepted definition of terrorism. There is, however, a global consensus to curtail it. A potential application of our approach is related to video surveillance. Ideally, individuals should not be differentiated by shape, size, color or similar attributes but on the basis of suspicious behavior such as excessive looking around, uncertain walk etc. Adding weights for these additional "body language" characteristics is likely to help identifying terrorists.

Video surveillance by closed-circuit television (CCTV) in public places is one of the most popular ways of protection. The exact number of CCTV cameras in the UK is not known but their number was estimated in 2002 to be about 500,000 in London and 4,200,000 in the UK. Needless to say, having staff watch each of the 4.2 million CCTV is not practical. Unfortunately, technology, such as face recognition software, has so far been disappointing in helping with this task.

Tracking behavior of suspected individuals by looking for particular types of body movement or particular types of clothing or baggage is likely to be more efficient. The underlying assumption is that in public spaces people behave in a small number of predictable ways, and that terrorists deviate from them. For this reason, this study may be useful for classification of suspected objects and people in supermarkets, etc. (Note: The authors are not experts in automated monitoring or human recognition. For this reason, the following potential application of pairwise comparison is based more on common sense than solid fact.)

According to [5], body language and tone of voice may convey, in some circumstances involving highly emotional situations, as much as 93% of the emotional state of an individual. Although this figure has been criticized as being an overestimation, 50-60% is still a realistic and important contribution as far as suspicious activities are concerned since they are highly correlated with the communication of emotions. For the sake of discussion—subject to review by the true experts on terrorism and/or human recognition—let us shortlist some of the characteristics which CCTV may be able to provide based on the machine vision input:

- excessive stopping,
- looking around,
- withdrawal movements of the head and shoulders,
- rapid body changes,
- number of contacts made,
- body angles,
- posture.

For the sake of exposition, we have selected only the first four of the above characteristics for inclusion in Tab. 3 after comparing them in pairs.

Tab. 3. Selected criteria for recognizing body language with their initial ratings (example)

Criterion	ID	C1	C2	C3	C4
Excessive stopping	C1	1	2	3/2	3
Looking around	C2	1/2	1	2	7/2
Withdrawal movements of head and shoulders	C3	2/3	1/2	1	2
Rapid body changes	C4	1/3	2/7	1/2	1

All combinations of pairs (including the same criterion which gives identity) create the pairwise comparisons matrix. For example, $C2$ to $C3$ according to Tab. 3 is set to 2 (which is "Excessive stopping" and "Looking around") and 1/2 below the main diagonal as a reciprocal value of 2. The inconsistency index ii is computed as the maximum distance as 0.63. By changing $C2/C3$ to 1 (based on new data or an expert's opinion), we get Tab. 4.

Tab. 4. Improved ratings by inconsistency analysis of the input (example)

Criterion	ID	C1	C2	C3	C4
Excessive stopping	C1	1	2	3/2	3
Looking around	C2	1/2	1	1	7/2
Withdrawal movements of head and shoulders	C3	2/3	1	1	2
Rapid body changes	C4	1/3	2/7	1/2	1

The new inconsistency index ii is computed as 0.57. By changing C3/C4 to 3, we get Tab. 5.

Based on equation (3), the final inconsistency indicator ii is computed as 0.25 which is below an acceptable threshold of 1/3 as explained in [7] and [12].

Weights are computed as normalized geometric means of rows as shown in Tab. 6.

Tab. 5. The final ratings (example)

Criterion	ID	C1	C2	C3	C4
Excessive stopping	C1	1	2	3/2	3
Looking around	C2	1/2	1	1	2
Withdrawal movements of head and shoulders	C3	2/3	1	1	2
Rapid body changes	C4	1/3	2	1/2	1

Tab. 6. Weights computed from the pairwise comparison ratings (example)

Criterion	ID	weight
Excessive stopping	C1	0.3987
Looking around	C2	0.2302
Withdrawal movements of head and shoulders	C3	0.2474
Rapid body changes	C4	0.1237

Shopping mall security and movie rating are other potential applications of the proposed method for classifying visual objects. By changing preferences, we may relax the level of acceptability of violence or sexual content depending on the targeted audience.

Conclusions

Finding a magic formula for the machine vision is nothing but a scientific dream. It is more and more evident that it is not just a matter of time before such magic formula is discovered but a long process of gradual improvements. As time passes, the perception has changed from the magic formula expectations to that "a little bit of this and a little bit of that" is needed to have a better machine vision. The transition from the mechanistic approach to the one based on "judgment call" or "expert opinion" is well supported by the pairwise comparisons approach in which the complexity of deciding on "everything at once" is reduced to comparing two features against each other to establish their relative importance. This approach reduces the complexity to the bare minimum. Two features are irreducible in practice since comparing the same feature with itself according to importance or relevance does not much make sense as a trivial case of the identity. However, pairwise comparisons require synthesis of partial solutions into one global solution. This presentation shows that it can be done and should be done.

Ideally, elements of a pairwise comparisons matrix should be based on by measurements and/or statistical studies. For example, in the presented case of detecting potentially criminal activities, values should be based on past observations and statistics. However, relying on opinions of a panel of experts is an acceptable compromise solution until such statistics are available. In fact, a jury in a justice system is an excellent example that such solution works in practice and the society in not even remotely prepared to wait until a certain number of similarly exotic criminal activities occur for an offended to be more accurately punished. So, the authors would not be surprised that the proposed temporary solution may become permanent for at least some cases in machine vision. If adding the weight to certain features of potential criminal activities improves the accuracy of recognizing such activities, why not use it regardless of how the partial assessments have been obtained?

The strongest conclusions are probably induced by the "what if analysis." In the absence of computed weights by the proposed method of pairwise comparisons, one would either add all scores received for the individual characteristics or tried to guess some weights. The authors are quite confident that in both cases, the proposed method would classify visual method more accurately as it is intuitively evident. If not, the calibration of weights by what in medicine is known as "clinical trials" and/or manual tune up would improve the precision (bringing the machine learning and artificial neural networks into the enhanced model) for achieving our goal to improve the machine vision by a better way of classifying visual objects.

Acknowledgments

The research for this study was partially supported by grant R00 035 02 of the Ministry of Education and Research in Poland, based on the policy No. 0019/R/2/T00/06/02.

We also acknowledge invaluable input provided by M. Choras and P. Rotter (listed in the alphabetical order) for the improvement of the example presented in Section 4.

References

1785
[1] de Caritat, M. J. A. N., Marquis de Condorcet: Essai sur l'Application de L'Analyse à la Probabilité des Décisions Rendues à la Pluraliste des Voix. Facsimile reprint of original published in Paris, 1972, by the Imprimerie Royale.

1860
[2] Fechner, G.T.: Elements of Psychophysics, Vol. 1. Translation by H.E. Adler of Elemente der Psychophysik (Leipzig Breitkopf und Hartel, 1860). Holt, Rinehart and Winston, New York, 1965.

1927
[3] Thurstone, L.L., Law of Comparative Judgements, Psychological Review, 34, 273–286.

1977
[4] Saaty, T.L. (1977), A Scaling Methods for Priorities in Hierarchical Structure, Journal of Mathematical Psychology, vol. 15, 234–281.

1981
[5] Mehrabian, A. (1981). Silent messages: Implicit communication of emotions and attitudes (2nd ed.). Wadsworth, Belmont, California.

1992
[6] Kato, T. (1992). Database architechture for content-based image retrieval. In Proceedings of the SPIE – The International Society for Optical Engineering, volume 1662 (pp. 112-113). San Jose, CA, USA.

1993
[7] Koczkodaj, W.W. (1993), A New Definition of Consistency of Pairwise Comparisons. Mathematical and Computer Modelling, Vol. 18, 7, 79–84.

1996

[8] Herman, M.W., Koczkodaj, W.W., A Monte Carlo Study of Pairwise Comparison, Inf. Process. Lett., 57(1), 25–29.

[9] Koczkodaj, W.W., (1996), Statistically Accurate Evidence of Improved Error Rate by Pairwise Comparisons, Perceptual and Motor Skills, 82, 43–48.

1998

[10] Koczkodaj, W.W., Testing the Accuracy Enhancement of Pairwise Comparisons by a Monte Carlo Experiment, Journal of Statistical Planning and Inference, 69(1), 21–32.

1999

[11] Koczkodaj, W.W., M. Orlowski N., Computing a consistent approximation to a generalized pairwise comparisons matrix, Computers & Mathematics with Applications, 37(3), 79–85.

2007

[12] Bozsóki, S., Rapcsák, T., On Saaty's and Koczkodaj's inconsistencies of pairwise comparison matrices, Journal of Global Optimization)to appear), available on line www.springerlink.com/content/v2x539n054112451/