Optimizing feature extraction in image analysis using experimented designs, a case study evaluating texture algorithms for describing appearance retention in carpets

S. A. Orjuela[a,d], R. A. Quinones[c], B. Ortiz-Jaramillo[b], F. Rooms [a], R. de Keyser[c] and W.Philips[a]

[a]Department of Telecommunications and Information Processing (TELIN), IPI, IBBT, Ghent University, Sint-Pietersnieuwstraat 41, B-9000 Gent, Belgium;
[b]Department of Textile, Ghent University, Belgium;
[c]Department of Electrical energy, Systems and Automation, Ghent University, Belgium
[d]Antonio Nariño University, Colombia

ABSTRACT

When performing image analysis, one of the most critical steps is the selection of appropriate techniques. A huge amount of features can be extracted from several techniques and the selection is commonly performed based on expert knowledge. In this paper we present the theory of experimental designs as a tool for an objective selection of techniques in image analysis domain. We present a study case for evaluating appearance retention in textile floor coverings using texture features. The use of experimental design theory permitted to select an optimal set of techniques for describing the texture changes due to degradation.

Keywords: Experimental designs, image analysis, texture analysis

1. INTRODUCTION

One of the most critical steps in image analysis is the selection of the relevant features according to the goal of interest. The variety of techniques available to extract characteristics from images as well as the different processes involved before interpretation may result in a high amount of data to be analysed.

Evaluation of image algorithms performance is traditionally based on human expert by visually comparing side by side the results.[1, 2] In this type of comparisons, the performance of image operators are correlated with the preferences of the observer, resulting in a subjective evaluation (Researcher bias).[3] Besides, particular applications require sophisticated algorithms that are commonly evaluated on few types of images, resulting in less generic algorithms.[4] Furthermore, dependences among techniques, processes and types of images among other factors are commonly not evaluated, decreasing reliability of findings and inferences in the evaluation of image algorithms performance.

Reliable findings and inferences can be obtained analysing quantitative and objective metrics through statistical tests.[4] Statistical tests permits to verify whether a scientific assumption is false. Commonly, experiments are conducted to refute a general assumption called the null hypothesis. The contrary statement to the null

hypothesis is called the alternative hypothesis. An statistical hypothesis test compares both hypothesis based on experimental results.

Planning phases in advance to optimize findings while reduce the development time and overall costs in processes can be conducted combining statistical tests based on experimental design theory.[5,6] Experimental designs have been included in many industrial applications to increase the quality of goods for international competition. The main advantage of this type of evaluation is the identification of dependences between the different elements in a study for detecting optimal combinations. The evaluation is based on identifying characteristics in the application that allow a reliable evaluation of the performance of all elements.

In image processing applications, statistical test have been conducted to improve the efficiency in evaluating differences among algorithms performance.[7–9] Studies include evaluations of the test abilities for rejecting null hypothesis when false and identification of specific differences between pairs of algorithms.[10,11] Few studies have been conducted using current computational tools for statistical analysis.[12] Recently, studies include experimental designs.[1] Most of these investigations are focused on the applications rather than on explaining the statistical procedures.

This paper presents the experimental design theory applied for evaluating appearance retention in textile floor coverings using texture features. The main concepts are discussed in detail for using it in other image applications.

Assessment of appearance retention in textile floor coverings specifies the duration of the products.[13,14] Appearance retention is certified following tests based on simulating traffic exposure by degrading (accelerating the wear) the surfaces of textile floor coverings samples using mechanical devices.[15] Changes in appearance of the degraded samples are currently evaluated subjectively by human inspectors. Therefore, manufacturers are seeking for more objective mechanisms.

With this goal in mind we digitize the 3D structure of textile floor covering surfaces into images, called range images, in which the pixels contain information of depth instead of colours, permitting to evaluate structural changes while being less affected by colour and lighting changes than intensity images.[16]

The experimental design in this study is conducted to evaluate the performance of a predefined set of texture analysis techniques applied on range images of different types of textile floor covering. Due to the nature of the evaluated textures, different types of textile floor coverings are better described at different scales. Therefore, the study requires to identify the scales at which the techniques perform better for each of the evaluated samples.

For reliable inferences, dependences among techniques, textile floor covering types and image scales are evaluated. The best combination is chosen by statistically analysing characteristics from the relation between the texture features and the transitional degrees of texture.

The paper is organized as follows, in Section 2 we describe the elements considered in this study. Then, in Section 3.1 an exploratory analysis is conducted to identify appropriate characteristics for evaluating the performance of the techniques. The characteristics are established from the relation between the texture features and the transitional degrees of texture. Subsequently, response variables are design to quantify these characteristics in Section 3.2. Then, experiments are conducted in Section 3.3 following experimental design models. Afterwards, response variables are compared conducting statistical tests in Section 3.5. Then, inferences and results are clearly stated and communicated by interpreting the probability results in Section 4. Finally, conclusions are drawn in Section 5.

2. DEFINING ELEMENTS OF THE EXPERIMENT

The first step in designing a statistical experiment is to limit the elements, called factors, to be evaluated. The factors are divided into primary factors, which are the main focus of the investigation, and nuisance factors, which are factors that may affect the measured results but are not of primary interest to the investigator. The possible options within a factor are called levels. A particular combination of levels of different factors is called a treatment.

In this approach, one is interested in selecting from a set of texture features, those that are optimal for describing texture changes on textile floor coverings due to degradation. The texture features are computed using texture techniques that are applied at different image scales. The factors of this experiment are established in terms of technique, texture types and scale resolutions with the techniques being the main factor of interest. In the followings the descriptions of the factors and corresponding levels as well as the image features to be used in this study are given.

2.1 Techniques

We have recently shown that algorithms based on Local Binary Pattern (LBP) techniques are appropriate for quantifying the texture appearance changes resulting from different exposure times of accelerated degradation.[17] Analysis in different scales is achieved by considering equidistant points on a concentric circle around each pixel in an image while varying the radius of the circle and the number of the points. LBP techniques characterize local textures in images with code numbers assigned to each pixel.[18] A code number represents the surrounding local texture described with a binary representation, which is constructed from boolean comparisons of intensity values between the pixel and a set of points around it. The binary representation are obtained by comparing each of the points in the circular neighbourhood with the central pixel as follows:

$$s_n = \left\{ \begin{array}{ll} 1, & (I_n > I_c) \\ 0, & otherwise \end{array} \right. , \forall\, n = 1, \ldots, N \tag{1}$$

The basic LBP code number is computed as follows:

$$LBP_{(r,N)} = \sum_{n=1}^{N} s_n \tag{2}$$

with r being the radius of the circle and N the number of points on the neighbourhood.

Pixels in which the intensity changes in the neighbourhoods lead to invariant versions of binary representations are associated with the same code numbers. Particularly, we have proposed an LBP technique extension based on mirror, complement and rotational symmetries, called the LBPSym technique, which have shown to be superior than previous similar techniques in terms of texture accuracy and texture discriminance when used on the Brodatz database.[19] The Brodatz texture database is the most common data set to test texture classification and texture recognition algorithms in the computer vision community, becoming the standard to evaluate the performance of new algorithms based on texture analysis.[20]

In this study we evaluate the performance of the LBPSym technique against some variations resulting from including features of other proposed LBP extensions, which are listed below.

1. The LBP with threshold technique.[21] This technique includes a threshold value (ϵ) to make the LBP technique more robust against negligible changes in pixel values. The boolean comparisons are performed as follows:

$$s_n = \begin{cases} 1, & (I_n - I_c) > \epsilon \\ 0, & otherwise \end{cases} \tag{3}$$

2. The LBP with modified census transform technique.[22] This technique computes the boolean comparisons using the average of the intensity values on the neighbourhood instead of the intensity value of the center pixel as follows:

$$s_n = \begin{cases} 1, & (I_n - \mu) > \epsilon \\ 0, & otherwise \end{cases} \tag{4}$$

with

$$\mu = \frac{1}{9} \times \left(\sum_{n=1}^{8} I_n + I_c \right) \tag{5}$$

Table 1: LBPSym variations

A. Census transform (CT) type	CT			Modified CT		
B. Threshold	0	0.05	0.1	0	0.05	0.1

2.2 Texture types

We selected a set of degraded textile floor covering samples exhibiting four different types of construction. This allows to evaluate the performance of the algorithms for generic purposes. The algorithms are applied on range images obtained from samples of textile floor coverings that have been degraded following the EN1471:1996 European standard.[13] The range images are obtained using our own scanner specifically designed for textile floor coverings.[16]

Each pixel in a range image represents the depth of the surface in an area of 0.24 mm by 0.24 mm. The surface of the samples are subjected to different levels of accelerated wear by using a Vetterman tester.[15] We use textile floor covering samples that exhibit nine transitional changes in texture appearance by being subjected to different degrees of surface degradation. The degrees on surface degradation are specified by numbers called Appearance Retention (AR) grades, varying from 1 to 5 with steps of half point. An AR grade of 1 represents a severe change whereas an AR grade 5 represents an original sample not subjected to degradation. A complete set (1-5 AR grades) of samples from the same textile floor covering type is called an AR reference.

2.3 Scale Resolutions

Degradation of different references is not optimally represented using the same imaging parameters since the area of the pile yarns tufts in textile floor covering samples depends on the AR reference type. Therefore, it is necessary to identify the best spatial resolution scale that optimally stresses the surface changes for each AR reference. We perform this analysis by keeping the LBP parameters fixed while resizing the images. Images are evaluated at spatial resolution scales within the range from 0.6 to 1 with intervals of 0.1. Scale factor smaller than 0.6 are not considered because the details of the texture surface are vanished at those scales.[17]

2.4 Image features

The set of image features to be evaluated results from applying the several treatments on the images. In this study we compute image features based on texture to quantify the degree of texture change in a textile floor covering sample.

The difference between the degraded and its corresponding original texture is quantified by comparing the frequency of appearance of the LBP code numbers between images of both textures. We use a symmetrized adaptation of the Kullback-Leibler divergence, commonly used to measure the difference between two probability distributions.[23] The frequencies of occurrence of the code numbers are first accumulated into one histogram per image, with each bin in an histograms representing a code number. Then both histograms are normalized and their difference is computed as follows:

$$\kappa = \sum_{h=1}^{2} \sum_{j=1}^{N} H(h,j) \log H(h,j) - \sum_{j=1}^{N} H_p(j) \log H_p(j) \tag{6}$$

where $h = 1$ or $h = 2$ represents the histograms, j is the code number index, $H_p(j) = \sum_{h=1}^{2} H(h,j)$ and N is the total number of code numbers.

We compute 18 κ-values for each one of the combinations: technique T on an image at a scale S of sample with an AR grade G from an AR reference R.

3. TECHNIQUES PERFORMANCE EVALUATION

In the followings, we discuss in detail the required steps for finding optimal treatments by using statistical experimental designs. The steps are discussed while performing the case experiment of evaluating variations of the LBPSym technique for quantifying texture changes of textile floor coverings.

3.1 Exploratory Analysis

To objectively evaluate the performance of the features extracted with image techniques on different imaging environments (given in this study by images resulting from combining different AR reference types at different scales) an exploratory analysis is conducted in advance to identify characteristics that allow a reliable comparison of the techniques in terms of the goal of the investigation.

The exploratory analysis involves descriptive and graphical statistical methods, which provide useful visual information to select appropriate assumptions as well as to identify data relationships. Specifically for this study, metrics that objectively measure the description of the κ-values related to the AR grades are required.

For a visual exploration, box plots and error bars of κ-values including complete subsets of AR grades are used. The subsets are randomly selected.

Figure 1 illustrates two box plots (on the left side) and two error bars (on the right side) that show the relationship between the κ-values related to AR grades for four of these randomly subgroups. The graphics indicate that the texture changes due to degradation on the textile floor coverings are transitionally described by the κ-values. This suggests that the optimal κ value features are those that monotonically describe the AR grades while the distinction between consecutive AR grades is maximized.

The box plots evidence the presence of outliers which may seriously affect the AR description. Outliers are detected by measuring the Mahalanobis Distance[6] on the residuals, which are data resulting when subtracting from each κ-value, the sample mean of the corresponding AR group. A κ value is considered an outlier if the
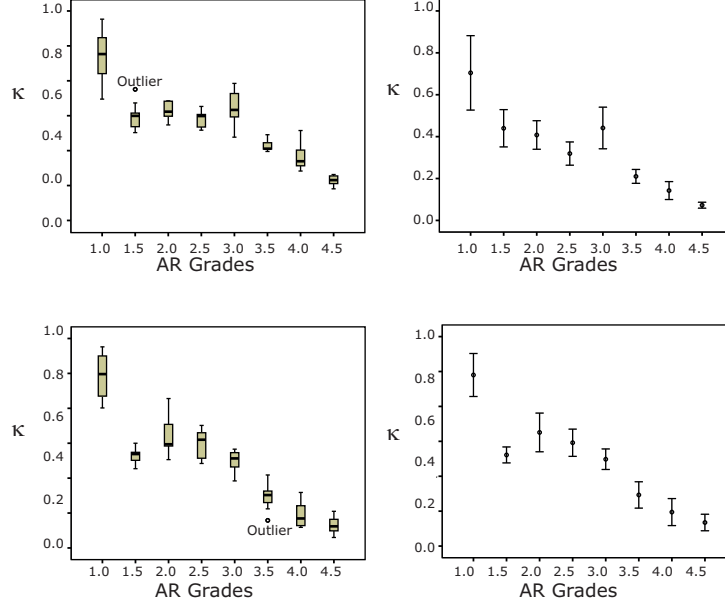
Figure 1: Samples of κ-values vs. AR grades.

Mahalanobis distance of its residual exceeds three standard deviations. In that case it is replaced with the associated sample mean to keep the number of the κ-values fixed.

The relationship between AR grades and κ-values is non linear and a linear description is preferred to give the same probability to each AR grade. A linear combination of the κ-values obtained from the different scale resolutions provide a better linear description of the AR grades as shown as follows:

$$\Gamma = \quad \alpha_0 + \sum_{i=1}^{5} \alpha_i \kappa_i \tag{7}$$

where i is the index of the scale resolution and the α-values are the parameters of the linear model.

In the followings we define response variables that measure these characteristics from the relationship between AR grades and Γ-values.

3.2 Defining response variables

The most difficult step in designing a experiment is to define objective performance measures quantifying characteristics that compliance with the investigation goals. These measures, called response variables, are computed using the image features. A measurement in the experiment using a response variable is called an observation.

In this study, response variables for quantifying characteristics in the AR grades representation with Γ-values are required. The findings in the exploratory analysis suggest that an appropriate representation of the transitional degrees of texture due to degradation may exhibit the following characteristics:

1. The mean value of the Γ-values per AR grade may be linear rank ordered.

2. It must be a clear distinction between Γ-values of consecutive AR grades.

Figure 2 illustrates both examples, one with a non optimal (on the left side) and one with an optimal description (on the right side).
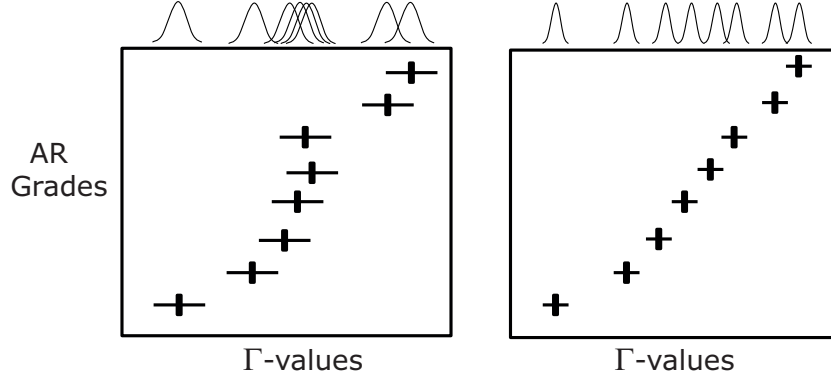

Figure 2: Optimal description of AR grades with Γ-values.

We propose to quantify the following three characteristics from the relationship between AR grades and Γ-values.

1. The monotonicity of the Γ-values related to AR grades.

2. The variability of the relation between Γ-values of each AR grades and the Γ-values of all AR grades in an AR reference.

3. The discriminance between Γ-values of consecutive AR grades.

These characteristics are quantified using the response variables described below.

1. **The monotonicity (M)** is computed by using the Spearman rank correlation, termed ρ.[6] For this, the mean values of the Γ data related AR grades are first ordered from small to large. Then, the Spearman rank correlation is computed using Equation (8).

$$\rho = 1 - \frac{K}{(G)((G)^2 - 1)} \sum_{g=1}^{G} d_g{}^2 \tag{8}$$

Where, $G = 8$ is the number of AR grades; $g = 1, \ldots, G$ is the index of the AR grade ; d_g is the differences between the assigned rank and the expected rank for Γ mean value of the AR grade g; and $K = 6$ is a constant defined by Spearman.

2. **The variability (V)** measures how well a linear relation between the Γ-values and the AR grades is explained. It is quantified by using the coefficient of determination, termed R^2,[6] which is computed as follows:

$$R^2 = \frac{S \sum_{g=1}^{G} \left(\bar{\Gamma}_g - \bar{\Gamma}\right)^2}{\sum_{g=1}^{G} \sum_{s=1}^{S} \left(\Gamma_{gs} - \bar{\Gamma}\right)^2} \tag{9}$$

with $S = 18$, $G = 8$ and $\bar{\Gamma}$ the mean value of all Γ-values of an AR reference.

3. **The discriminance (D)** is calculated based on the Tukey test,[6] which is used to identify whether there are significant statistically differences in the Γ-values between consecutive AR grades. With $S = 18$ being the number of Γ-values per AR grades, the statistical significance is computed as follows:

$$\varsigma = \frac{q_{(\alpha,G,SG-G)}}{\sqrt{(SG-G)S}} \sqrt{\sum_{g=1}^{G} \sum_{s=1}^{S} \left(\Gamma_{gs} - \bar{\Gamma}_g\right)^2} \tag{10}$$

$q_{(\alpha,G,SG-G)}$ is obtained from the studentized range distribution at $100(1-\alpha)$ of confidence. $\bar{\Gamma}_g$ is the Γ mean value of the AR grade g.

The discriminance is computed as the number of times that Equation (11) is satisfied.

$$(\bar{\Gamma}_{(g+1)} - \bar{\Gamma}_g) - \varsigma > 0 \tag{11}$$

The discriminance is represented with the Greek letter τ. The τ value is divided by 7, the maximum possible value, to normalize the response variable from 0 to 1.

In the next Section we explain how to construct a generalized linear model to represent the response variables in function of the treatments.

3.3 Selection of experimental design model

In experimental designs is assumed that the mean values of a response variable due to the treatments can be correctly fitted by a generalized linear model. If this assumption is validated, a reliable statistical analysis can be further conducted to evaluate differences among the treatments.

Table 2: Factor and levels in this experiment.

a_i	Census transform (A)	b_j	Threshold (B)
a_1	CT	b_1	0
a_2	MCT	b_2	0.05
		b_3	0.1

c_l	AR Reference (C)
c_1	Cut
c_2	Loop
c_3	Cut Loop
c_4	Shaggy
c_5	Loop
c_6	Shaggy

We have chosen in this approach a factorial experiment, in which the response variables are modelled in function of parameters quantifying the effects of all possible treatments. Other types of experimental designs

that require less computation costs by modelling using only subsets of treatments (like fractional factorial designs) are based on the same principles explained below.

The main parameters in this study are defined in Table 2. As an example, a model in function of the a set of parameters $\{a_2, b_3, c_5\}$ refers to the treatment in which the LBPSym technique is applied using the modified Census transform and a threshold $\epsilon = 0.1$ on the Loop AR reference type.

The basic model without taking into account dependences for the response variable Monotonicity is constructed using Equation 12. Models for variables Discriminance and Variability are equally constructed.

$$\rho_{ijklm} = \bar{\rho} + a_i + b_j + c_k + \varepsilon_{ijklm} \tag{12}$$

where $\bar{\rho}$ is the mean value computed from all ρ data, m is the index when applying the same treatment several times, which are called replications, and the ε-values are the errors of the model, called residuals, with respect to the expected value.

Equation 13 shows the model including the effect of dependences, called interactions, between levels of different factors. Interactions are indicated with $(X)_x$, where X is a combination of factors and x specifies the levels. I.e $(ab)_{23}$ means interaction between factors CT Type and Threshold for levels $\{MCT, 0.1\}$.

$$\rho_{ijkl} = \bar{\rho} + a_i + b_j + c_k + (ab)_{ij} + (ac)_{ik} + (bc)_{jk} + (abc)_{ijk} + \varepsilon_{ijkl} \tag{13}$$

The experimental error due to nuisance factors can be reduced by dividing their levels into homogeneous groups, called blocks. This type of model do not necessarily require the use of replications. As an example a reduced model, in which blocks are created in terms of AR references, is expressed as follows:

$$\rho_{ijk} = \bar{\rho} + a_i + b_j + c_k + (ab)_{ij} + \varepsilon_{ijk} \tag{14}$$

A model is expressed in matrix form as follows:

$$\vec{\rho} = M\vec{\theta} + \vec{\varepsilon} \tag{15}$$

Where $\vec{\rho}$ is a vector containing all ρ observations, $\vec{\theta}$ is the set of parameters, M a matrix called the matrix design and $\vec{\varepsilon}$ is the set of residuals.

For the particular design given by Equation 14, the vector $\vec{\theta}$ is defined by Equation 16, where $\vec{A}$ is a vector containing the parameters of the levels of a factor A (Same for the other factors), and $(\vec{X})$ is a vector containing the interaction parameters due to a factors combination X.

$$\vec{\theta} = [\bar{\rho}|\vec{A}|\vec{B}|\vec{C}|(\vec{AB})] \tag{16}$$

M is constructed by filling with ones in each row, the columns corresponding to the parameters that explain the associated observation. Other columns are filled with zeros.

The parameter of a factor level is calculated as the difference between the mean value of its associated observations and the overall mean as in Equation 17. Where $\bar{\rho}(X_x)$ is the mean value of the observations in which the level x of the factor X, for any of the experimental factors, is applied.

$$X_x = \bar{\rho}(X_x) - \bar{\rho} \tag{17}$$

The interaction between the levels of two different factors is estimated as in Equation 18. Where $\bar{\rho}((X_x, Y_y))$ is the mean value of the observations in which levels x and y of the factors X and Y are applied.

$$(XY)_{xy} = \bar{\rho}(X_x, Y_y) - X_x - Y_y - \bar{\rho} \tag{18}$$

Thus, the expected observation values $\hat{\rho}$ can be estimated as follows:

$$\hat{\rho} = M\vec{\theta} \tag{19}$$

The set of differences between the expected $\hat{\rho}$-values using the model and the experimental ρ-values are then computed using Equation 20.

$$\vec{\varepsilon} = \vec{\rho} - \hat{\rho} \tag{20}$$

3.4 Model validation

A model is said to be valid if the observations are normally distributed around their expected values. This can be easily evaluated by visually comparing the shape of the residuals distribution with a normal distribution using a Q-Q plot, in which the quantiles of both distributions are plotted against each other. If residuals are normally distributed the plotted points should lie on a straight line.
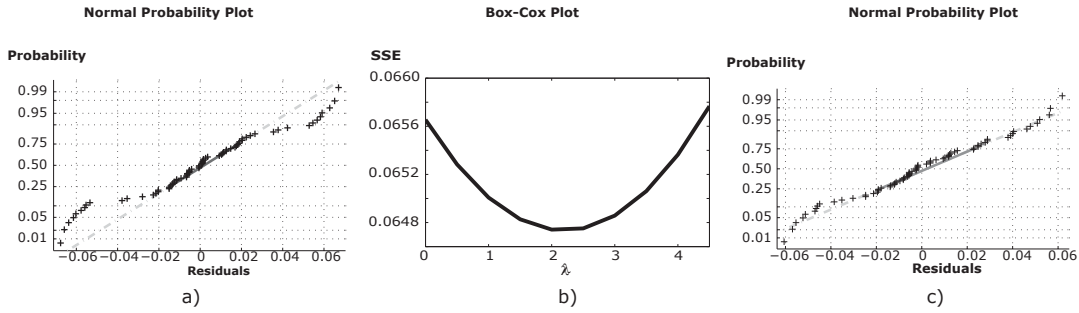

Figure 3: Cox-Box Method for residuals normalization.a) shows the residuals of the model in Equation 13. b) shows the sum of squares of the residuals vs. the variable λ. c) shows the Q-Q plot of the residuals when SSE is minimal.

Figure 3 a) shows the residuals of the model in Equation 13. The graphic shows that the residuals do not lie on a straight line. The residuals can be transformed into normally distributed data by applying a transformation on the observation data.

An appropriate transformation can be found by using the Box-Cox method.[6] The method reduces the sum of squares of residuals, termed SSE, by applying the transformation given by Equation 21.

$$\rho^\lambda = \begin{cases} \frac{\rho^\lambda - 1}{\lambda K^{\lambda-1}} & \lambda \neq 0 \\ K \log \rho & \lambda = 0 \end{cases} \tag{21}$$

with $K = \left(\prod_{i=1}^{t} \rho_i\right)^{1/t}$, the geometric mean of the observations and t the total number of observations. Figure 3 b) shows the sum of squares of the residuals Vs. the variable λ. Figure 3 c) shows the Q-Q plot of

the residuals when SSE is minimal. Once the residuals are normally distributed, the statistical analysis can be conducted.

3.5 Statistical analysis

The statistical comparison allow the identification of undesired dependences between levels of different factors. The comparison of the factors are conducted by analysis of variance ANOVA tests. ANOVA is a powerful test that in its simple form permits to check whether there is at least one significant difference between different factors. ANOVA results are probabilities, termed p-values, where a p-value less than a given α means that there is at least one difference between the comparison with $100(1 - \alpha)\%$ of confidence.

In case of differences (p-value less than a given α), ANOVA does not specify the differences and a test for multiple comparison, like the Scheffé test, must be performed. In the Scheffé test, significant differences are detected also when p-values are less than a given α. For valid p-values, the factors must be independent. The interactions are identified also by ANOVA where a p-value less than a given α reveals interactions between the associated elements. assuring the validity of the p-values.

The p-values of the ANOVA test are obtained using the sum of squares of the parameters of the model and the corresponding degrees of freedom. Table 4 shows the computation of the p-values for the model in Equation 14. In the table, X is any of the sources of variation listed in row 1.

Table 3: Basic samples of sum of squares and degrees of freedom.

Source of Variation (X)	Sum of Squares (SS_X)	Degrees of Freedom (Df_X)	Mean Square	F_0	p-values
A	$jk \sum_{i=1}^{I=2} (a_i)^2$	$I - 1 = 1$			
B	$ik \sum_{j=1}^{J=3} (b_j)^2$	$J - 1 = 2$	$MS_X =$	MS_X/MS_E	From an
C	$ij \sum_{k=1}^{K=6} (c_k)^2$	$K - 1 = 5$	SS_X/Df_X		F Distribution
(AB)	$k \sum_{i=1}^{2} \sum_{j=1}^{3} ((ab)_{ij})^2$	$(I - 1)(J - 1) = 2$			
Residuals (E)	$\sum_{i=1}^{2} \sum_{j=1}^{3} \sum_{k=1}^{6} (\varepsilon_{ijk})^2$	$(IJ - 1)(K - 1) = 25$	MS_E		
Total (T)	$\sum_{i=1}^{2} \sum_{j=1}^{3} \sum_{k=1}^{6} (\rho_{ijk} - \bar{\rho}_{ijk})^2$	$IJK - 1 = 35$			

In the following Section we show how to interpret the p-values for the current study case.

4. RESULTS AND DISCUSSIONS

The analysis is performed by evaluating the p-values on the table from down to top to first identify interactions. In case of interactions (p-value $> \alpha$), the p-values of the factors are not valid. Therefore, the treatments must be divided into subgroups that do not reveal interactions between levels of different factors. Then, an independent statistical analysis on each subgroup to conducted to find optimal treatments.

In this study, we conducted independent statistical tests on the three response variables: monotonicity, discriminance and variability. The response variables were initially split in the effects of the factor levels using

Table 4: ANOVA tests results

Source of Variation (X)	p-values		
	ρ	R^2	τ
A	0.165	0.041	0.197
B	0.179	0.080	0.916
C	< 0.005	< 0.005	< 0.005
(AB)	0.165	0.966	0.995
(AC)	0.908	0.528	0.149
(BC)	0.132	0.895	0.976
(ABC)	0.908	0.851	0.999

the model of Equation 13. The resulting p-values are listed in Table 4. These p-values are used to identify interactions between the AR references and the techniques. In case of interactions the AR references are divided into subgroups and the response variables are evaluated using reduced models without interactions.

In the followings we discuss in detail the interpretation of these p-values.

1. **Statistical analysis of monotonicity (M)** The statistical analysis of the response variable mono-
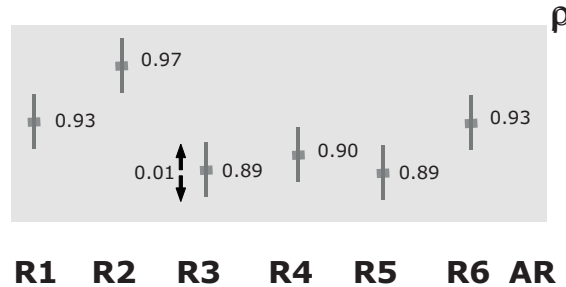

Figure 4: Comparison of monotonicity in AR reference representations.

tonicity do not evidence interactions between factors as well as differences between techniques are not identified, both inferences with with a 99% of confidence (p-values > 0.01).

The p-value of the Factor C evidence differences between its levels (types of AR References). A Sheffé test shows that the AR reference 2 has bigger monotonicity than the others. However, the Sheffé test show mean values in terms of monotonicity above 0.88, which suggest that all the variations of the LBPSym technique describe almost transitional the texture changes due to degradation.

Errors in the transitional description are according to human mistakes. Figure 4 shows the resulting mean values in the Sheffé. The standard deviations are indicated constant and indicated with the symbol $\leftrightarrow$.

2. **Statistical analysis of variability (V)**

The statistical analysis of the response variable monotonicity do not evidence interactions between factors with a 99% of confidence (p-values > 0.01). However, it shows difference between the levels of the techniques A, with 95% of confidence (p-value = 0.041 < 0.05), and B, with 90% of confidence(p-value = 0.080 < 0.1), as well as between the AR references with 99% of confidence(p-value < 0.01). The results of the Sheefé test are shown in Figure 5 in the same way that the results of monotonicity in Figure 4.

The results show that the LBPSym computed with the Census Transform and without applying a threshold performs better that the other alternatives in terms of fitting a linear model with the Γ values. Furthermore, References one, two and three are better modelled than the others.
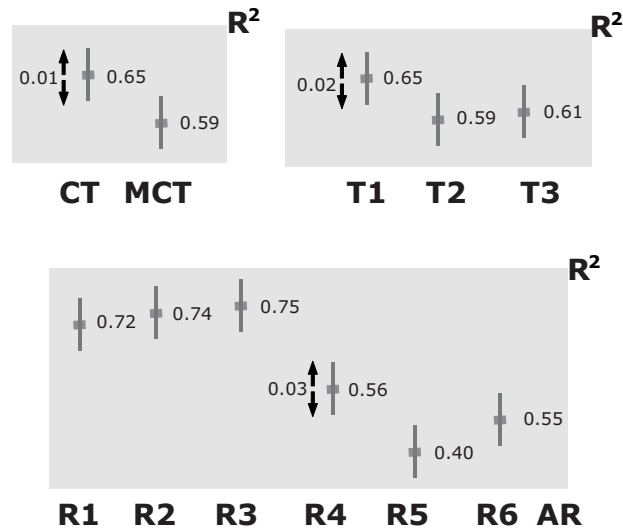
Figure 5: Comparison of Variability in techniques and AR reference representations.
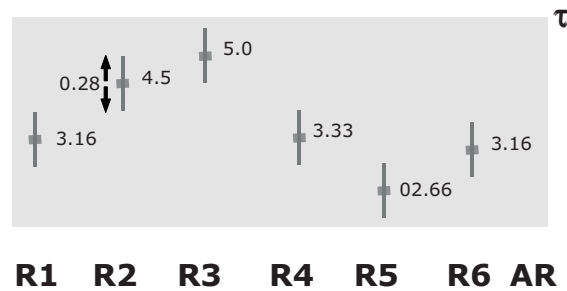
3. **Statistical analysis of discriminance (D)**

Figure 6: Comparison of Discriminance in AR reference representations.

The statistical analysis of the response variable discriminance do not evidence interactions between factors as well as differences between techniques are not identified, both inferences with with a 99% of confidence (p-values > 0.01). It shows only differences between the AR reference types, where AR references two and three discriminate more consecutive AR grades.

5. CONCLUSIONS

In this paper we have presented a methodology for optimizing the feature extraction process when evaluating texture algorithms for describing appearance retention in textile floor coverings. The methodology permitted to identify optimal combinations of the LBPSym techniques using two types of Census Transform and three types of thresholds, applied on different AR references. Results show that the basic Census transform combined with no threshold performs better for describing the texture changes due to degradation. We have discussed the main concepts of the methodology for using it in other image processing applications.

REFERENCES

[1] Z. Pei, Z. Lin, H. Zhou, and H. Wang. A method of image processing algorithm based on orthogonal experimental design. In *In ternational conference on image and graphics*, 2009.

[2] M. Heath, S. Sarkar, T. Sanocki, and K. Bowyer. Comparison of edge detectors, a methodology and initial study. *Computer Vision and Image Understanding*, 38.54(1):69, January 1998.

[3] I. Avcivaş, B. Sankur, and K. Sayood. Statistical evaluation of image quality meausures. *Journal of electronic imaging*, 11(2):206–223, April 2002.

[4] P. Courtney and N. A. Thacker. *Imaging and Vision Systems: Theory, Assessment and Applications*, chapter Performance Characterisation in Computer Vision: The Role of Statistics in Testing and Design, pages 109–128. Nova Science Publishers, 2001.

[5] D. C. Montgomery. *Design and analysis of experiments*. Wiley, 7th edition, 2009.

[6] M. Kutner, C. J. Nachtsheim, J. Neter, and Li W. *Applied Linear Statistical Models*. McGraw-Hill/Irwin, 5 edition, 2004.

[7] R. M. Zucker and O. T. Price. Statistical evaluation of confocal microscopy images. *Cytometry*, 44:295–308, 2001.

[8] S. Singh and K. Bovis. An evaluation of contrast enhancement techniques for mammographic breast masses. *IEEE Transactions on Information Technology In Biomedicine*, 9(1):109–119, March 2005.

[9] H. R. Sheikh, M. F. Sabir, and A. C. bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on image processing*, 15(11):3441–3452, November 2006.

[10] adn J. A. Detre G. K. Aguirre, E. Zarahn, and D. C. alsop. Experimental design and the relative sensitivity of bold and perfusion fmri. *NeuroImage*, 15:488–500, 2002.

[11] A. R. Mcintosh, F. L. Bookstein, J. V. Haxby, and C. L. Grady. Spatial pattern analysis of functional brain images using partial least squares. *NeuroImage*, 3:143–157, 1996.

[12] J. Shiraishi, L. L. Pesce, C. E. Metz, and K. Doi. Experimental design and data analysis in receiver operating characteristic studies: Lesson learned from reports in radiology from 1997 to 2006. *Radiology*, 253(3):822–830, December 2009.

[13] Constructional details of types of textile floor covering available as reference fatigued specimens. EN1471, 1996.

[14] The Carpet and Rug Institute. Cri test method - 101: Assessment of carpet surface appearance change using the cri reference scales. Technical Bulletin, 2003.

[15] Textile floor coverings – production of changes in appearance by means of vettermann drum and hexapod tumbler testers. Tc 219 Floor coverings, ISO 10361:2000, March 2005.

[16] S. A. Orjuela, E. Vansteenkiste, F. Rooms, S. De Meulemeester, R. De Keyser, and W. Philips. Feature extraction of the wear label of carpets by using a novel 3d scanner. In *Proc. of the Optics, Photonics and Digital Technologies for Multimedia Applications conference*, 2010.

[17] S. A. Orjuela, E. Vansteenkiste, F. Rooms, S. De Meulemeester, R. De Keyser, and W. Philips. Evaluation of the wear label description in carpets by using local binary pattern techniques. *Textile Research Journal*, 80(20):2132–2143, December 2010.

[18] T. Ojala and M. Pietikäinen. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, 29(1):51–59, January 1996.

[19] S. A. Orjuela, R. Quinones, B. Ortiz-Jaramillo, F. Rooms, R. de Keyser, and W. Philips. Improving texture discrimination in the local binary patterns technique by using symmetry and group theory. In *17th International Conference on Digital Signal Processing*, 2011.

[20] P. Brodatz. *Textures: A Photographing Album for Artists ans Designers*. Dover Publications, 1999.

[21] M. Heikkilä, M. Pietikäinen, and J. Heikkilä. A texture-based method for detecting moving objects. *Pattern Analysis and Machine Intelligence*, 28(4):657–662, 2003.

[22] B. Froba and A. Ernst. Face detection with the modified census transform. In *Automatic Face and Gesture Recognition*, pages 91–96, 2004.

[23] M. Petrou and P. G. Sevilla. *Image Processing Dealing with Texture*. Wiley, January 2006.