AESTHETIC-ORIENTED CLASSIFICATION OF 2D FREE-FORM CURVES

Petrov A.
LSIS UMR CNRS 7297
Arts et Metiers ParisTech, France
aleksandar.petrov@ensam.eu

Pernot J-P.
Veron P.
LSIS UMR CNRS 7297
Arts et Metiers ParisTech, France
{jean-philippe.pernot, philippe.veron}@ensam.eu

Giannini F.
Falcidieno B.
IMATI-CNR, Genova, Italy
{franca.giannini, falcidieno}@ge.imati.cnr.it

ABSTRACT

Nowadays, it is commonly admitted that the aesthetic appearance of a product has an enhanced role in its commercial success. Therefore, understanding and manipulating the aesthetic properties of shapes in the early design phases has become a very important field of research. There exists an appropriate vocabulary for describing the aesthetic properties of 2D free-form curves that includes terms such as straightness, acceleration, convexity and tension, which are normally used by stylists when describing and modifying shapes. However, the relationships between this vocabulary and the geometric models are not well established. This work investigates the possibility of applying Machine Learning Techniques (MLT) to discover possible classification patterns of 2D free-form curves with respect to the so-called straightness of the curve. First, we verified that MLT can correctly (99, 78%) reapply the classification to new curves. In addition, we verified the abilities of the Attribute Selection methods to identify the most important attributes for the considered classification, among a larger set of attributes. As a result, it was possible to recognize as the most characterizing parameters the same curve attributes previously used to compute the measure of straightness (S). Moreover, Linear Regression (LR) was able to extract automatically an exact mathematical model, which can correlate the geometric quantities with the class of the curve, congruent to one we previously specified. This work indeed demonstrates that MLT are very suitable and can be efficiently used in this context. The work is a first step towards the characterization and classification of free form surfaces giving the general directions on how MLT can be exploited to characterize free-form surfaces with respects to the aesthetic properties.

KEYWORDS

Aesthetic properties, 2D free-form curves, shape characteristics, shape classification, machine learning techniques (MLT)

1. INTRODUCTION

Nowadays, the point of view for analysing the shape of a new product has been moved from production feasibility toward users acceptance. Until recently, all products were designed taking into consideration production and cost constraints therefore designers could not pay enough attention to the shape they prefer, because they had to focus on the shape that can be reached with production equipment as well. It was a closed system (designers-manufacturers) where the customers were somehow excluded.

Now, with the availability of new materials and development of production equipment, especially with the development of five-axis CNC machines, we do not pose anymore the question of which shape can be produced, but which shape do you want to produce. Thus, the designers have more freedom to design what they like. Additionally, in the current competitive market, being faster in reaching the customers expectations is becoming extremely important. Therefore, we can foresee a new product development process in which the customer is an important actor as the other experts in analysing and defining the product. Taking into consideration the fact that customers are more important than before, aesthetic aspects of the product are becoming significantly more important and they become key factors for its acceptance in the market.

Including the customers in the design process emphasizes the difference of the languages used in different activities of the design process [9]. Namely, Marketing Language is referred to as first language and Designer language as second language, and the aesthetic characteristics represent a bridge between both languages. The first one is adopted during the specification of the wished product and during the briefing activity. It represents an emotional and individual description of the aesthetic character. The second one represents its detailed specification during the creation and modification of the product model according to geometric properties and criteria. In order to identify the appropriate aesthetic characterization of shapes and to understand which feature properties stylists consider, the design activities carried out in different industrial field have been deeply analysed within the European Project FIORES II (Character Preservation and Modelling in Aesthetic and Engineering Design). One of the objectives of this project was to find mappings between verbal styling description and geometric parameters and characteristics. Such a mapping can allow the development of modelling tools closer to the stylists mentality. References [9] and [11] present the identified terms and developed measurements for styling properties together with CAD tools based on them. Even though an appropriate aesthetic vocabulary is determined, there is still the need for a verification of the understanding of their meaning by the non-designers. Moreover, it might be more useful to allow a qualitative judgment (e.g. less, more, very, not or not at all) of these aesthetic properties than quantitative measures. In this work, the focus is on the straightness property of 2D curves. Using the results provided in [8] and [12] we aim at analysing possible ways to automatically defining styling properties measures and qualitative categorization of curves according to such properties.

Thus, this work introduces a new framework for classifying curves with respect to aesthetic property. In particular it has the following goals:

- Verifying the interrelation between the curve characteristic quantities and the aesthetic property
- Exploring the capability for selecting the most meaningful geometric quantities for classifying curves with respect to the aesthetic properties.
- Evaluating the ability of extracting a mathematical model (equation) which will correlate the straightness with the geometric quantities.

To do that, a set (ST) of curves has been generated by applying appropriate deformation vectors at two points of initial straight lines. Afterwards, for each curve from the set ST, the characteristic quantity values of all curves are computed.

Finally, having the curves quantities on one hand and the measure of straightness on the other hand, we have the suitable data for applying Machine Learning Techniques to discover some structural patterns and to select most meaningful curve quantities (figure 1).

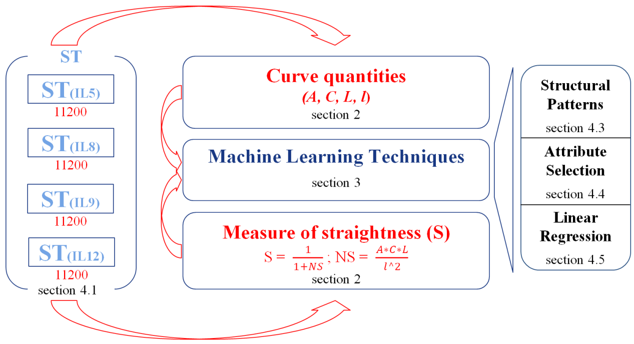

Figure 1 Overall framework and data workflow

The following section details the formulation for the measure of the straightness of 2D free-form curves. Then, after a brief presentation of the Machine Learning Techniques in section 3, the section 4 describes how the experiments were conducted and what the results are. Finally, section 5 discusses our results with a conclusion and some perspective for future works.

2. STRAIGHTNESS OF 2D FREE-FORM CURVES

Before providing the definition of straightness, we note here the distinction between the terms aesthetic and styling properties. The aesthetic property refers primarily to the appearance of the overall shape of the product while styling property refers mostly to the shape of the geometric elements (curves and surfaces). However, styling properties seem to be the best way to build up the aesthetic character (appearance) of the product. In other words, with modifying the shape of its geometric elements, the related styling properties are changed so the overall appearance of the object is changed and it might lead to another aesthetic character.

Among the terms and measures for styling properties identified by the European Project FIORES II, in this paper we focus on the straightness property. While in engineering, a curve is either straight or not (a part some tolerance), for a stylist a curve can be more or less straight, depending on how visually it differs from the line segment which is somehow related to the dimension of the overall curvature radius in relation to the curve length. The bigger the radius is, the straighter the curve. Even curves having inflection points and consequently variable radius can appear somehow straight.

The FIORES II project proposed the following formulation for the straightness measure:

$$straightness(S) = 1 - d_{min}/d_{max} \tag{1}$$

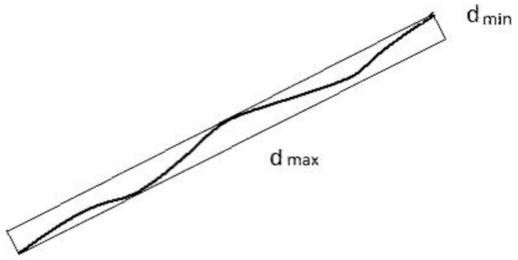

Figure 2 Straightness of a curve

Where dmin and dmax indicate the height and the width of the minimum bounding rectangle. Actually, the bounding rectangle does not provide any supplementary information concerning to the character of the curve, i.e. how it behaves, the derivations or oscillations. In addition, totally different curves can have the same bounding rectangle.

Because of previously mentioned drawbacks, the work described in [12] proposes a revised measure for the straightness.

$$NS = non - straightness = \frac{C \cdot A \cdot L}{l^2} \tag{2}$$

non-straightness (NS) $\in [0,)$

$$S = straightness = \frac{1}{1 + non - straightness} \tag{3}$$

straightness (S) $\in (0, 1]$

Where the curve is given as a function of the parameter u:

$$P(u) = (x(u), y(u)), u \in [0, 1] \tag{4}$$

$$C = \int_0^1 |k(u)| du \tag{5}$$

is the integral of the absolute value of the curvature.

$$A = \int_0^1 \sqrt{y(u) \cdot \dot{x}(u)} du \tag{6}$$

is the value of the area between the curve and the line that joins the two extremes of the curves.

$$L = \int_0^1 \sqrt{\dot{y}(u)^2 + \dot{x}(u)^2} du \tag{7}$$

is the length of the curve

$$l = \sqrt{(x(1) - x(0))^2 + (y(1) - y(0))^2} \tag{8}$$

is the length of the chord between the two end points of the curve.

This measure of straightness better represents the character of the curves than the initial one. The initial qualitative classification of the straightness property according to property range values has been further verified (see Table 1). From the work in [12], it has been concluded that there is a general convergence in the categorisation of curves according the very not, not, fairly and very judgments for the straightness property. Therefore, in this work the measure of straightness expressed with the equ. (3) and the classification given in the table 1 are used in the further analysis to create the model for straightness category prediction.

Table 1 Curve classification with respect to the straightness measure formulation in (eq. 3)

Straightness (S)	Classification
(0 0,7)	very-not-straight(nS)
[0,7 0,9)	not-straight(ns)
[0,9 0,98)	fairly-straight(s)
[0,98 0,999)	very-straight(S)
[0,999 1]	straight-line(SS)

3. USING MACHINE LEARNING TECHNIQUES

This section provides some basic notions related to Machine Learning Techniques, to their usage in applications, and to the environment adopted in the reported experiment.

3.1. Machine Learning

Machine Learning Techniques have been widely used to discover structural patterns for analysing problems in medical and biology application [17, 22, 33, 34, 36]. Next, wider application of the machine learning is in prediction of stock market [16, 20, 26], where its application is used to discover which are the most relevant factors (features selection), that will influence the stock market. Furthermore, most common application of machine learning is in image processing [1, 5, 15, 23, 25, 30, 31], where machine learning is used to discover classification patterns for feature recognition and automatic annotation. Some similar usage is applied for 3D model classification and retrieval [6, 14, 13]. The later refers to discovering classification patterns of shape descriptors of already created 3D CAD models. In design application, various works aim at understanding the association between the shape of a product and the evoked product image, i.e. the most appropriate description such as calm, feminine, aggressive, from customer interviews [18, 19, 21, 29]. These works are mainly devoted to provide tools able to foreseen the evoked product images for new products. Only very few works try to detect shape parameters directly usable for the shape manipulation in the early phases of design process [11]. In FIORES-II, for instance, machine learning capabilities of CBR system have been applied to verify the emotional image of variants of product templates [9]. The the aim of this work is to investigate the possibility of applying machine learning in early phase of design process for mapping of aesthetic properties to geometric entity (2D free-form curves).

Machine learning is a core subarea of Artificial Intelligence (AI) that studies computer algorithms for learning rules from data to deal with some particular tasks. The main domain where Artificial Intelligence (AI) has increasing application is in solving Data Mining (DM) problems. Different definitions of Data Mining exists, two of the most general state: Data Mining is defined as process of discovering patterns in data where the process must be automatic or semiautomatic and the patterns must be meaningful in that they lead to some advantages [37]; Data Mining defines the automated extraction procedures of hidden predictive information from databases [24]. Overall, both interpretations designate that the capability of Data Mining is finding a way to mine (dig) in a huge set of data and getting some knowledge from it [32]. In general, tasks that address Data Mining problems are divided into [37]:

- A Predictive group that refers to Prediction, Classification, and Regression
- A Descriptive group that refers to Association rules, Clustering, Summarization and Attribute Selection.

To get trustable rules form data it is important to have consistent data set on which to proceed in the learning process. Considering this, in our experiment we used set of curves ST consisting of 44.800 curves, which is an acceptable number to be candidate for a Data Mining problem solving.

In our work, the WEKA data mining workbench has been adopted. The WEKA (Waikato Environment for Knowledge Analysis University of Waikato, New Zealand) workbench is a collection of the most used machine learning algorithms and data pre-processing tools designed in a way that provides flexibility of their manipulation. Reference [37] provides a comprehensive and detailed description of WEKA workbench. In addition, WEKA performances and particular functions are described in various papers, such as [2, 3, 4, 7, 35] illustrating its use in different applications.

3.2. Implementation of machine learning algorithms using WEKA workbench

There are three ways of using WEKA. The first one is to apply a learning method to a dataset and analyze its output in order to learn more about the data. The second way is to use a learned model in order to generate prediction of new instances and the third is to apply several different learners and compare their performance in order to choose the most suitable for prediction.

WEKA provides several different applications: Explorer, Experimenter, Knowledge Flow and Simple CLI. For our purposes, the most useful application is the Explorer that gives access, through menu selection, to quickly read a dataset from a ARFF file and build a decision tree from it or to any other algorithm, from all different Data Mining problems to explore. The second WEKA application is Experimenter and it is designed to help to answer basic practical questions when applying classification or regression techniques - such as which methods and parameters values work best for the given problem. In addition, Experimenter provides a good environment to compare a variety of learning technics. The Knowledge Flow application allows the manipulation with boxes that represent learning algorithms and data sources around the screen in order to design the configuration wanted. In other words, it helps the design of the structure of the data flow with simple connection of box components that represent data source, appropriate tools, learning algorithms, evaluation methods and visualization modules.

In our case, the application Explorer is going to be used for reading and preparation of data, applying the learning algorithms and results visualization. The objective of this work is to create a classification model using Machine Learning algorithms and to investigate the possibility of using the Attribute Selection algorithm in order to select most meaningful geometric quantities of curve among of set of various basic quantities. The aim of the later objective is to see whether WEKA could be used for selection of most significant geometric quantities of surfaces in the view of defining styling properties and related measures also for surfaces and objects. A more detailed analysis of previously mentioned objectives is presented in the next section.

3.3. Modeling of the straightness problem into WEKA

WEKAs native data file format is the ARFF (Attribute-Relation File Format). An ARFF file (fig. 4) has two distinct sections: the first section is the *Header* information which is followed by the *Data* information.

```
% Comments
@RELATION <relation-name>

@ATTRIBUTE <attribute-name1> <datatype>
@ATTRIBUTE <attribute-name2> <datatype>
@ATTRIBUTE class <nominal-specification>

@DATA
attribute-name1,attribute-name2,class
attribute-name1,attribute-name2,class
```

Figure 3 Structure of an ARFF input file for WEKA

The *Header* contains the name of the relation (@RELATION), a list of the associated attributes (@ATTRIBUTE) with their types and if necessary comment lines starting with the character %. The *Data* section contains the data declaration (@DATA) line followed with the instance lines. Aiming at modeling the straightness, the measure expressed by the eq. 3 is adopted as a mathematical model of curve straightness and its parameters (C, A, L, l) represent the curves quantities. So as to create the model for straightness class prediction, as many as possible different curves have to be created and for each of them, the measure of straightness and curves quantities have to computed. Later, such computed values have to be stored in appropriate text data file (.arff) suitable for further use of WEKA. In order to get a suitable data an adequate program in Matlab has been implemented. It includes a function for the automatic generation of curves and a function that compute curves quantities and its associated curves class according with the eq. 3 and the table 1. At the end of this Matlab program, the ARFF file containing both *Header* and *Data* section of information is generated with the following *Header* section:

% Header

@RELATION straightness

@ATTRIBUTE Area(A) REAL
@ATTRIBUTE Curvature(C) REAL
@ATTRIBUTE Curve_length(L) REAL
@ATTRIBUTE Base_length(l) REAL
@ATTRIBUTE Class $\{(SS), (S), (s), (ns), (nS)\}$

The *Data* section contains a line for each instance of curves with the computed values of the curves quantities and the corresponding class according to the table 1.

4. EXPERIMENTS

This section describes the experiments carried out in WEKA regarding the categorisation of curves according straightness property measures. First the method adopted for the data set creation is reported, and then the various experiments are reported together with a discussion on the obtained results.

4.1. Generation of an input set of curves (ST)

First, a set of significant number of instances of 2D curves is created. For this purpose, a deformation model (figure 4) has been defined and applied on an initial straight line to create many 2D free-form curves. The adopted deformation model uses three deformation modes (ϵ_1, ϵ_2 and ϵ_3) based on two displacement vectors (blue and yellow) applied on two control points of the parametric curve. 200 displacements are assigned to each vector direction. Figure 4 illustrates the different vector directions and the resulting types of shapes

For example, in the ϵ_1 mode, in the first subset, we selected 11 different directions and then both vectors are equal and parallel with the 11 vector directions and 200 positions for each direction. In the second subset, only the vector modules are different and they are still parallel. The same principle is adopted for the ϵ_2 and ϵ_3 deformation modes with different number of vector directions. In these two cases we provided 6 different initial directions. To summarize, the deformation modes allow to create with ϵ_1 (11 directions * 200 positions * 2 cases) = 4400 curves, with ϵ_2 (6 directions * 200 positions * 2) = 2400 curves and with ϵ_3 4400 curves too. Collecting the curves obtained with the above described deformation modes gives us the set of 11200 curves applied on four different straight initial curve lengths (5, 8, 9 and 12 cm) to produce the global set ST of 44800 2D curves as illustrated in figure 5.

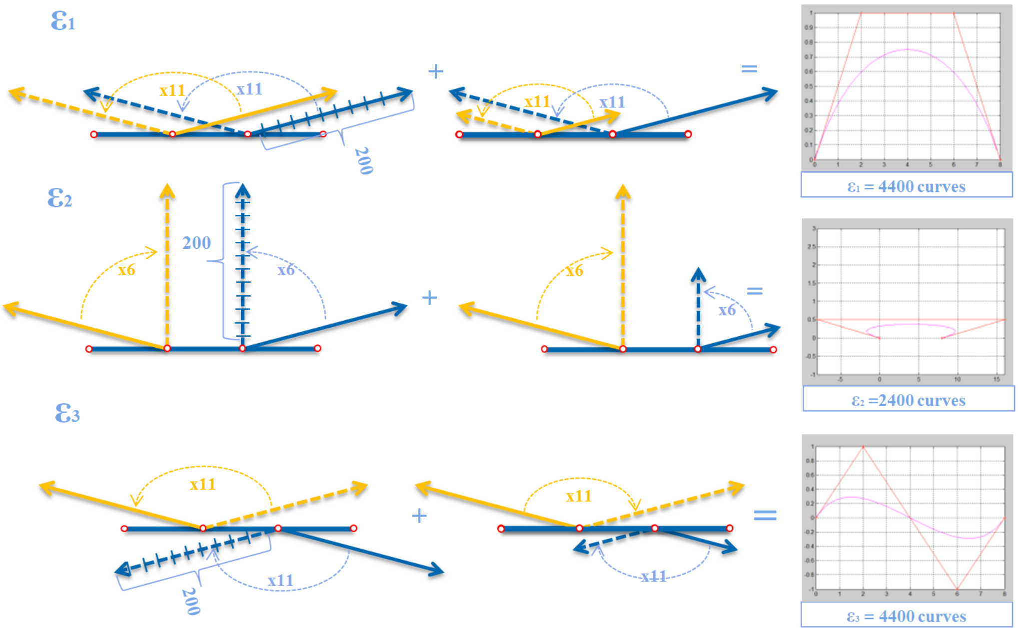

Figure 4 The deformation model used to generate a set containing a variety of 2D curves

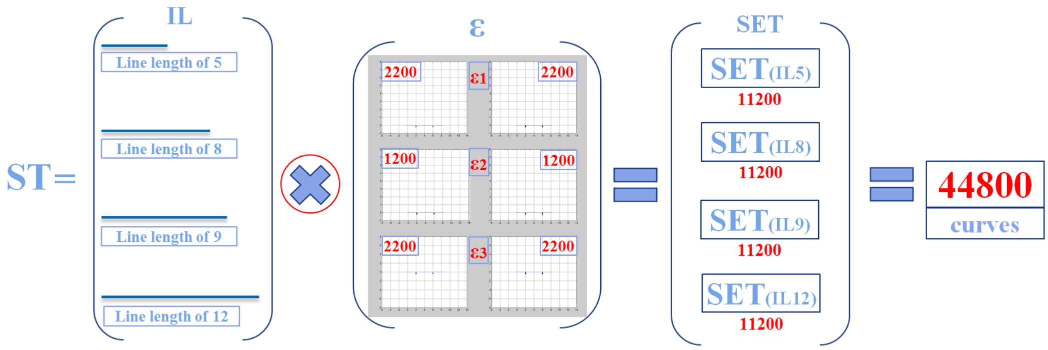

Figure 5 Generating a set of 2D curves for the experimentation in WEKA

4.2. Classification using dimensional attributes

In terms of using machine learning algorithms, many different classification algorithms have been used but only the classification trees (particularly, tree J48) appeared to be most convenient in case of numerical attributes [32]. The process of classification model creation consists of two steps (modeling and validation). In the first step, the classification algorithm is applied on the global input set ST of 2D curves, while in the second step, a subset S-ST (10-30%) of ST only is used to validate the model, giving the percentage of good classified instances (in this step the classification curves belonging to the set ST minus S-ST is considered as unknown). For the validation process, WEKA enables three different ways: cross-validation split test set and supplied test set, performing use of appropriate algorithms. Namely, cross-validation algorithm divides the set of instances in training set and testing set in a ratio that can be chosen by the user. For instance, if 10 folds cross-validation is chosen, it means that whole data set of instances is divided in 10 parts from which 9/10 are used as a training set (for creation of a model) and 1/10 is used as test set (for validation of the model) and this step is repeated 10 times. Unlike the cross-validation, split percentage algorithm splits the data set in training and test set by inserting the percentage (given by the user) of instances that will be contained in the test set and it is done only once. The third way is totally different: the user independently chooses the training and testing sets which are not required to be a part of a common data set.

Results:
When the modelling process is over, learning schemes are automatically generated by WEKA and depending on which validation method has been performed, we obtained a percentage of correctly classified instances varying from **98.63** % - cross validation - to **98.18** % - split test. Moreover, when analysing the incorrectly classified instances, we can observe that they are classified just in the neighboring classification class and the error of the classification is not more than **0.5** %. Having these results into account and the fact that the intervals of the classification class (table 1) are very strict from a numerical value point of view, a change of **0.5** % in the intervals will not make any visible changes in the appearance of curves which allows us to consider that this classification model is relevant.

Discussion:
One important aspect is that the classification in the curve category should be somehow independent on the curve dimension, i.e. it should reflect only shape behaviour characteristics. Thus, to verify whether this model is independent form the initial line length and use it to predict the straightness class of any curve, an evaluation of the model has been performed using curves obtained from other initial lines having different lengths, i.e. 0.8, 80 and 240 cm. The results obtained from the evaluation, show that the model correctly classifies from **21.14** % to **28.42** % of all data instances so it demonstrates that this way of model creation does not guarantee to be size independent. Therefore, another approach for model creation has to be defined that will be more general and independent from initial curves or attributes. In the next section this new approach is presented together with the results obtained.

4.3. Classification using dimensionless attributes

As mentioned before a new approach has been explored in order to overcome the lack of low rate of correctly classified instances. After an analysis of the mathematical model of the straightness (eq. 3), it was concluded that whilst the final measure is independent of the curve size, each single parameter (attribute) in the equation is related to the curve dimension. Thus, this model gives good results in classification only for the range of curves used for its creation. We had therefore to find a way to apply the mathematical model (eq. 3) and to get rid of such size dependency of the constituting parameters. To achieve it, all the parameters of the equation are divided by a dimensional parameter, which is constant for all the deformed curves obtained from a specific initial curve (see section 4.1). Knowing the fact that all curves are obtained by applying deformation model over a straight line, it means that the parameter (eq. l) remains the same for that set of curves. In order to use the same equation (eq. 3) but not to have dimensional parameters in it, a transformation of the equation is obtained as following:

$$NS = \frac{C \cdot A \cdot L}{l^2}(\frac{l}{l}) = (C \cdot l)(\frac{A}{l^2})(\frac{L}{l}) = Cr \cdot Ar \cdot Lr \quad (9)$$

Using this transformed version of the equation (eq. 3) ensures the same values for the measure of straightness but using dimensionless quantities.

Results:
Based on this reformulation of the measure parameters, the new values have been computed for the curve data set ST and stored in an appropriate ARFF data file. Afterward, the same classification algorithm has been applied to all data file curve instances to create the classification model. In order to evaluate this model, the same set of curves obtained by deformation of initial lines (length of 0.8, 80 and 240 cm) has been used and the results show that this model can correctly classify **99.78** % of all curve instances.

Discussion:
As before, the **0.22** % of instances, which are incorrectly classified, are considered belonging to the neighbor class. Considering that at the border of the class interval (the error is less than **0.5** %), and that the mathematical model (eq. 3) for class evaluation is very rigid, a change of **0.5** % in the intervals will not make any visible changes in the appearance of curves. Therefore, *we can conclude that this classification model is reliable and provides good classification of curves.*

4.4. Relevant attribute selection

Another objective of our experiments was to verify the capabilities of MLT in identifying the key elements for entity classification. Thus, we intend to use the attribute selection capability to solve the problem of characterizing free form surfaces with respect to their aesthetic properties. At present, there is no clear specification of styling and aesthetic properties and related measures for surfaces. Exploiting such selection capability, we want to find out which of the various computable surface characteristics are the most significant for the evaluation and modification of their aesthetic and styling properties.

In general, when there is a huge set of data characterised by many different attributes, whose importance in the further analysis or is not known, then Attribute Selection (AS) as a part of Data Mining methodology can be applied. AS allows the identification of which attributes can be omitted without affecting the results of further analysis.

To investigate the Attribute Selection capabilities of WEKA for shape classification, we used the measure of straightness of curves together with its parameters and additional other computable properties for curves. The idea is to see if the Attribute Selection provided by WEKA will select the same parameters already used for computation of the straightness measure (S) among a larger list of parameters (attributes). In other words, if the Attribute Selection using WEKA proposes the same curve parameters used in the computation of (S), then we can consider this methodology (AS) as reliable and use it in further investigation of free form surfaces with respect to the aesthetic properties. For this purpose, using the same deformation model, applied onto an initial straight line, a set of **11200** curves ($SET_{(IL8)}$, fig. 5) is created and for each curve, the **13** different curve parameters shown in table 2 are computed. The first three correspond to the relative (transformed) parameters (C_r, A_r, L_r) of the equation (9) used for the NS measure computation.

Table 2 List of curve quantities used for AS

No.	Parameters	Description
1	*Lr*	Relative curve length
2	*Ar*	Relative area
3	*Cr*	Curvature
4	centrox	x - coordinate of barycenter
5	centroy	y - coordinate of barycenter
6	Ix	Moment of inertia over x-axis
7	Iy	Moment of inertia over y-axis
8	Con	The measure of Convexity
9	acceleration	The measure of acceleration
10	acc	Measure of non-acceleration
11	j	Number of local maximums
12	s	The parameters value of max
13	ks	Local curvature maximum

Con - corresponds to the measure of convexity which is related to the sign of the curvature along the curve, *acceleration* is a measure of acceleration that describes the rise of the curvature along the curve and *acc* is a measure just opposed to the acceleration measure.

WEKA provides two different methods for the AS regarding the attributes evaluation and their representation. The first method uses algorithms that provide independent evaluation of all attributes and then applies search algorithms to rank all attributes in a list. In this case, the *InfoGainAttributeEval* evaluation algorithm is used which calculates the *mutual information* (information entropy) of the attributes and classes; then such calculated values are ranked in decreasing order by the *Ranker* algorithm. The second method uses correlation based algorithm to evaluate a subset of attributes; then it applies appropriate search of algorithms to rank and propose the best subset of attributes. In this case, the *CfsSubsetEval* evaluation algorithm is used and then *BestFirst* search algorithm is applied to propose a subset of attributes which is highly correlated with the classes, but the attributes in the subset are more independent among themselves.

Results:

The following figures show the results of AS using WEKA in which all attributes associated with the values of the *mutual information* (Information Theory) are ranked (figure 6), and a subset of attributes is proposed (figure 7) as most significant with respects to the straightness.

```
Ranked attributes:
 1.298    1 Relative_Curve_Length(Lr)
 0.896    2 Relative_Area(Ar)
 0.826    7 Iy
 0.793    3 Curvature(Cr)
 0.734    6 Ix
 0.706    5 centroy
 0.437    8 Con
 0.167    4 centrox
 0       12 s
 0       13 ks
 0        9 acceleration
 0       10 acc
 0       11 j
```

Figure 6 Ranking of the attributes

Discussion:

The results produced, shown in figure 7 confirm the assumptions previously made: *the AS using WEKA has chosen the same transformed parameters that were used in the computation of the measure of straightness.* Therefore the AS is very promising for

```
Selected attributes: 1,2,3 : 3
                     Relative_Curve_Length(Lr)
                     Relative_Area(Ar)
                     Curvature(Cr)
```

Figure 7 Selection a subset of attributes

the identification of surface properties meaningful for the evaluation of the aesthetic and styling properties of surfaces and objects.

4.5. Linear Regression model

Creation of classification model using a decision tree based on the attribute values can be convenient for visualization of the results. Anyhow, it is appropriate only in the case the decision tree has a little number of branches and leaves, so that the class of the entity can be easily established by following the decision rules. The use of this classification method to classify the 2D free form curves results in generation of decision tree with more than 450 branches and leaves.Thus, we do not show it in this paper. Having this size of the decision tree makes very complex and unclear the determination the class of the curve. Furthermore, it is very difficult to find out the correlation of the shape quantity values with the class of the curves, i.e. how the change of one quantity will affect the class of the curve. In order to reach this correlation, an appropriate mathematical model has to be applied. In our case, i.e. classification of the 2D free form curves with respect to straightness, there is the mathematical model indicated by the equations (eq. 2) and (eq. 3). So, the aim of this part is not to define a mathematical model for straightness but to evaluate the ability of Machine Learning Techniques in WEKA to solve Data Mining problem using Regression. This will give us interesting results that could be then reapplied for surfaces.

After analyzing all available machine learning algorithms in WEKA, the linear regression algorithm has shown as the most suitable for handling only with numerical inputs (attributes) and outputs (classes). The idea of linear regression is to express the class as a linear combination of the attributes:

$$C_l = w_0 + w_1 \cdot A_1 + w_2 \cdot A_2 + w_3 \cdot A_3 + \cdots + w_n \cdot A_n \quad (10)$$

Where, C_l – *numeric value for the class,*
$A_1, A_2, A_3, ..., A_n$ – *the attribute values,*
$w_0, w_1, w_2, w_3, ..., w_n$ – *the weights.*

Regarding the equation for computing the straightness (non-straightness), we can see that the equation (eq. 2) for non-straightness is combination (not linear) of the attributes of curve and yells values form 0 to (infinity). So, if we want to transform this equation (eq. 2) into a linear combination (which is the way how the regression algorithm gives the results) of the attributes, each left and right side of the equation has to be used as an argument of a logarithmic function:

$$log(NS) = log(C) + log(A) + log(L) - 2 \cdot log(l) \quad (11)$$

this holds also for the transformed size independent version of the eq. 2 (eq. 9):

$$log(NS) = log(C_r) + log(A_r) + log(L_r) \quad (12)$$

The reason why a logarithmic function is used is the fact that if we want to compare the results of the regression model with the equation (eq. 2), they have to be in same form . This means, all values (of the class and the attributes) have to be transformed using logarithmic function and then to be recorded in an appropriate ARFF file so that will adjust the input data into a form adequate for the use of regression.

<u>Results:</u>
As it was mentioned before, the values of the curves quantities and the value of the non-straightness for each curve generated using the deformation model (figure 4), are transformed using logarithmic function. After the ARFF file is created, WEKA workbench is ran to create the linear regression model. The obtained results are given in following figures 8 and 9:

```
Linear Regression Model

Non-Straightness(NS) =

      1      * Curve_Length(L) +
      1      * Area(A) +
      1      * Curvature(C) +
     -4.1589
```

Figure 8 linear regression model for the eq. (11)

```
Linear Regression Model

Non-Straightness(NS) =

      1      * Relative_Curve_Length(Lr) +
      1      * Relative_Area(Ar) +
      1      * Curvature(Cr) +
      0
```

Figure 9 linear regression model for the eq. (12)

<u>Discussion:</u>
To prove that using linear regression modeling, a reliable mathematical model can be extracted we have to obtain the same results as with the linearized functions (11) and (12). Thus, equation (11) is compared

with result shown in figure 8 and equation (12) with figure 9. It is evident that regression model in figure 8 completely match with the eq. (11) because *2log(l)* = 4.1589 (where l = 8 cm is the length of the initial line used to generate the set of curves used). The same conclusion can be made if we compare the regression model in figure 9 with the eq. (12). In both cases, the correlation coefficients are **1** and relative absolute error is **0** %. Therefore, we can make a conclusion stating that *Machine Learning Techniques using Linear Regression modeling can extract a reliable mathematical model (equation) that will show the correlation of the geometric quantities with the class of the curve with respect to an aesthetic property straightness.* Such an approach can be then applied for surfaces.

5. CONCLUSION AND PERSPECTIVES

The goal of this work is an attempt to i) introduce Machine Learning Techniques (MLT) as a mean for discovering classification patterns with respects to the aesthetic properties of shapes on 2D free form curves; ii) use Data Mining (DM) methodology, to investigate which of the shape characteristics of a geometric entity (curve, surface) are the most significant with respects to a specific aesthetic property.

To verify that MLT could be suitable and useful for shape classification, we have analyzed its behavior in the case of the straightness of 2D curves. We based our work of the mathematical formula for the straightness measure defined in our previous work. We verified that MLT can correctly reapply the classification to new curves. In addition, we verified the abilities of the Attribute Selection methods to identify the most important attributes for the considered classification: among a larger set of attributes. As a result, it was possible to recognize as the most characterizing parameters the same curve attributes previously used to compute the measure of straightness (S). Moreover, Linear Regression (LR) was able to extract automatically an exact mathematical model, which can correlate the geometric quantities with the class of the curve, congruent to one we previously specified.

Finally, this work indeed demonstrates that MLT is very suitable and can be used efficiently in this kind of mechanical engineering and aesthetic applications. This offers good perspectives for solving the same problem on free form surface.

At present, there is no classification of surfaces with appropriate aesthetic properties. This requires at first the identification of the most meaningful free form surface characteristics (parameters) and reciprocal relations with respects to the aesthetic properties and then their classification patterns to be discovered.

Therefore, this work is considered to be a first step towards the characterization and classification of free form surfaces with respect to their aesthetic properties.

ACKNOWLEDGMENT

This work has been partially supported by the VISIONAIR project funded by the European Commission under grant agreement 262044.

References

[1] Abreu, M., Fairhurst, M., (2008), "An Empirical Comparison of Individual Machine Learning Techniques in Signature and Fingerprint Classification", BIOID 2008, LNCS 5372, pp. 130-139.

[2] Aher, B. S., Lobo, L.M.R.J., (2011), "Data Mining in Educational System using WEKA", International Conference on Emerging Technology trends (ICETT), Proceedings published by International Journal of Computer Applications (IJCA).

[3] Aksenova, S.S., (2004), "Machine Learning with WEKA: WEKA Explorer Tutorial", Scholl of Engineering and Computer Science - Department of Computer Science, California State University, Sacramento.

[4] Caruana, R., Freitag, D., (1995), "Greedy Attribute Selection", School of Computer Science, Carnegie Mellon University, Pittsburgg.

[5] Conilione, C.P., Wang, D., (2011), "Automatic localization and annotation of facial features using machine learning techniques", Soft Compt, 15:1231-1245, DOI 10.1007/s00500-010-0586-y.

[6] desJardins, M., Eaton, E., Wagstaff, L.K., (2006), "Learning User Preferences for Sets of Objects", Proceeding of the 23rd International Conference on Machine Learning, Pittsburgh, PA.

[7] Garner, R.S., (1995), "WEKA: The Waikato Environment for Knowledge Analysis", Department of Computer Science, University of Waikato, Hamilton.

[8] Giannini, F., Montani, E., Monti, M., Pernot, JP., (2011) "Semantic Evaluation and Deformation of Curves Based on Aesthetic Criteria", Computer-Aided Design Applications, Vol. 8(3) pp. 449-464.

[9] Giannini, F., Monti, M., Podehl, G., (2006), "Aesthetic-driven tools for industrial design", Journal of Engineering Design, 17:03, 193-215.

[10] Giannini, F., Monti, M., (2002), "An innovative approach to the aesthetic design", Common Ground Design Research Society International Conference 2002 (London, UK, 5-7 September 2002). Proceeding, pp. 415-426., D Durling, J. Shackleton (eds.), Staffordshire University Press.

[11] Giannini, F., Monti, M., (2010), "A survey of tools for understanding and exploiting the link between shape and emotion in product design", Proceedings of the TMCE 2010 Conference, April 12-16, 2010, Ancona, Italy.

[12] Giannini, F., Monti, M., Pelletier, J., Pernot, JP., (2013), "A Survey to Evaluate how non Designers Perceive Aesthetic Properties of Styling Features", Computer-Aided Design Applications, 10(1), 129-138.

[13] Ip, Y.C., Regli, C.W., (2005), "Content-Based Classification of CAD Models with Supervised Learning", Computer-Aided Design and Application, Vol. 2, No. 5, pp. 609-617.

[14] Laga, H., (2009), "3D Shape Classification and Retrieval Using Heterogenous Features and Supervised Learning", Machine Learning, Abdelhamid Mellouk and Abdennacer Chebira (Ed.), ISBN: 978-953-7619-56-1, InTech

[15] Lattner, D.A., Miene, A., Herzog, O., (2004), "A Combination of Machine Learning and Image Processing Technologies for the Classification of Image Regions", AMR 2003, LNCS 3094, pp. 185-199.

[16] Lee, M.C., (2009) "Using support vector machine with a hybrid feature selection method to the stock trend prediction", Expert Systems with Applications, 36, pp. 10896-10904.

[17] Lemm, S., Blankertz, B., Dickhaus, T., Muller, K.R., (2011) "Introduction to machine learning for brain imaging", Neuroimage, 56, pp. 387-399.

[18] Lesot, J-M., Bauchard, C., Detyniecki, M., Omhover, J-F., (2010), "Product shape and emotional design an application to perfume bottle", International Conference an Kansei Engineering and Emotion Research 2010, KEER2010, March 2-4, 2010, Paris, France.

[19] Lu, X., Suryanarayan, P., Adams, B.R., Li, J., Newman, G.M., Wang, Z.J., (2012), "On Shape and the Computability of Emotions", MM12, October 29 November 2, 2012, Nara, Japan.

[20] Luo, L., Chen, X., (2013) "Integration piecewise linear representation and weighted support vector machine for stock trading signal prediction" Applied Soft Computing, 13, pp. 806-816.

[21] Machwe, A., Parmee, C.I., (2006), "Integrating aesthetic criteria with evolutionary processes in complex, free-form design an initial investigation", 2006 IEEE Congress on Evolutionary Computation, July 16-21, 2006, Vancouver, BC, Canada.

[22] Maddouri, M., Elloumi, M., (2002) "A data mining approach based on machine learning techniques to classify biological sequences", Knowledge-Based Systems, 15, pp. 217-223.

[23] Motaal, G.A., El-Gayar, N., Osman, F.N., (2010), "Different Region Identification in Composite Strain-Encoded (C-SENC) Images Using Machine Learning Techniques", ANNPR 2010, LNAI 5998, pp. 231-240.

[24] Negnevitsky, M., (2005), "Artificial Intelligence: A Guide to Intelligent Systems", Second Edition (2005), chapter 1.1, Pearson Educated Limited, Harlow, England

[25] Negri, G.R., Dutra, V.L., Sant-Anna, S.J.S, (2014), "An innovative support vector machine based method for contextual image classification", ISPRS Journal of Photogrammetry and Remote Sensing, 87, pp.241-248.

[26] Ni, L.P., Ni, Z.W., Gao, Y.Z., (2011), "Stock trend prediction based on fractural feature selection and support vector machine", Expert Systems with Applications, 38, pp. 5569-5576.

[27] Pernot, JP., (2004) "Fully Free Form Deformation Features for Aesthetic and Engineering Designs", PhD Thesis, INP-Grenoble, IMATI-CNR, 2004

[28] Pham, B., (1999), "Design for aesthetics: interactions of design variables and aesthetic properties", In Proceeding of SPIE IST/SPIE 11th Annual Symposium - Electronic Imaging '99, Vol. 3644 pages pp. 364-371, San Jose, USA

[29] Ren, Y., (2012), "Design Preference Elicitation, Identification and Estimation", PhD thesis, The University of Michigan, Michigan, USA

[30] Sajn, L., Kukar, M., (2011), "Image processing and machine learning for fully automated probabilistic evaluation of medical images", Computer Methods and Programs in Biomedicine, I04, e75-e86.

[31] Schwenker, F., Trentin, E., (2014), "Pattern classification and clustering: A review of partially supervised learning approaches", Pattern Recognition Letters, 37, pp. 4-14.

[32] Shhab, A., Guo, G., Neagu D., (2005), "A Study on Application of Machine Learning Techniques in Data Mining", Department of Computing, University of Bradford, UK.

[33] Wang, S., Summers, M.R., (2012) "Machine learning in radiology", Medical Image Analysis, 16, pp. 933-951.

[34] Wang, X.W., Nie, D., Lu, B.L., (2013) "Emotional state classification from EEG data using machine learning approach", http://dx.doi.org/10.1016/j.neucom.2013.06.046i.

[35] War, De R., Neal, L.D., (1994), "WEKA Machine learning Project Cow Culling", Livestock Improvement Corporation (LIC) of New Zealand.

[36] Wasan, P.S., Uttamchandani, M., Moochhala, S., Yap, V.B., Yap, P.H., (2013) "Application of statistics and machine learning for risk stratification of heritable cardiac arrhythmias", Expert Systems with Application, 40, pp. 2476-2486.

[37] Witten, I.H., Frank, E., Hall, M.A., (2011), "Data mining: Practical Machine Learning Tools and Techniques", Third Edition (2011), chapter 1, Morgan Kaufmann, Burlington, USA