Automatic morphological classification with Case-Based Reasoning

Thomas FISCHER*[1], Alexander ARTSCHWAGER[1], Konrad PFLEIDERER[1],
Anke RISSIEK[2], Magdalena MANDALKA[2], Andreas SEIDL[2], Rainer TRIEB[2]
[1] Center for Management Research, German Institutes of Textile
and Fiber Research Denkendorf, Denkendorf, Germany;
[2] Human Solutions GmbH, Kaiserslautern, Germany

Abstract

It is still a challenge for the apparel industry to develop good fitting products and underlying sizing and grading systems. This is due to the diversity of human bodies having the same traditional size but different morphotypes. Additional reasons are differences between different countries and special target groups such as young people or old people.

The objective of the iMorph-approach is the morphological classification based on body scan data to be used for size system development and to provide better fitting clothes. Additional applications include recommendations systems in online business and curated shopping.

iMorph is a unique approach to estimate the morphological classification of individuals based on body scan data. First, a morphological classification scheme was developed. It comprises 10 features and according ordinal scales. The analysis of the available body scan data showed that it is merely impossible to derive rules for the automatic morphological classification. Therefore, human experts visually classified a number of selected scans (data sets) by looking at the scanatars. A simple web-based application allows remote classification for invited experts. The resulting case base of classified scans is the core of a Case-Based Reasoning (CBR) system. It is able to compare the data of a new, unclassified scan with the scans of the case base. The most similar scans are the used derive a good estimation for the classification of the new scan.

The crucial element of this iMorph approach is the applied similarity model. The functions are specific for each morphotype feature because not all measurements are relevant for all features. An example may illustrate this: for the classification of the shoulders of a new data set, only a number of scan data related to the shoulders is relevant, for example breast girth, distance shoulder to buttock. The classification is then derived from the morphological classification of these most similar "shoulder cases".

The described approach has been proved to be valid and comprehensive [1]. It is flexible and extendable because each classification feature has separate similarity and retrieval functions with vast expert knowledge embodied and can be linked to various case-bases. Time consuming individual morphotype classification can be replaced by CBR technology and supports fashion product development as well as size recommendation in online retail.

Keywords: Morphological Classification, Case-Based Reasoning (CBR), morphotypes, size table development, fit optimization, size recommendation

1. Introduction and Outline

As human populations change constantly up-to-date size surveys are the pre-condition for fitting fashion products and a high market share in the apparel industry. Numerous survey results from different countries and special target populations are used to develop standard sizing systems as well as to optimize grading and size ranges. Today survey results are provided in body measurement portals allowing standard statistical analysis as well as company-specific size chart development for key markets [2]. Due to economic reasons the apparel industry tries to maximize their global market potential with a relatively small number of sizes. The standard sizes are often just defined by a few primary dimensions although body shapes, proportions and measurement correlations differ. Body scans from surveys show the variation of body shapes within one size and are therefore an optimal starting point for the definition and differentiation of morphotypes and at the same time for a better fit.

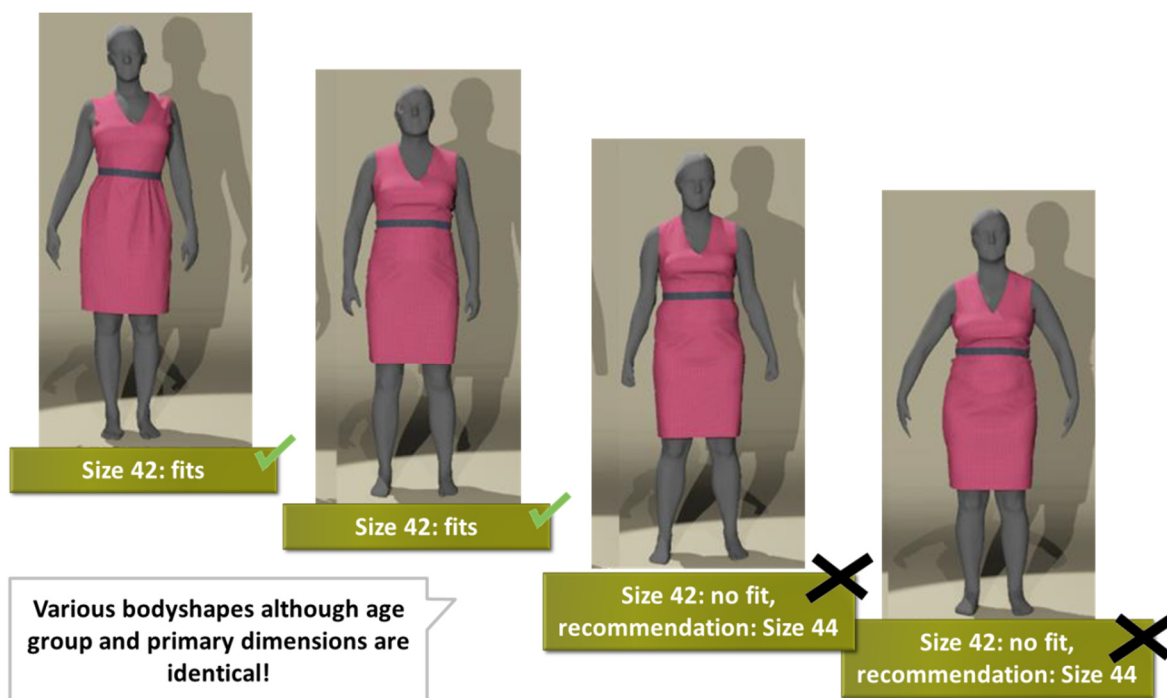* ThomasVFischer@DITF-MR-Denkendorf.de; +49- 711- 9340 - 419; www.DITF-MR-Denkendorf.de

Fig. 1. Women in German Size 42 with different body shapes and size recommendation.

The objective of the project iMORPH is the morphological classification based on female body scan data form the representative German size survey SizeGERMANY to be used for size system development and to provide better fitting clothes. The morphotype classification can be applied in online retail shops for size recommendation and curated shopping. Furthermore future collection development can incorporate the knowledge about morphotypes and provide special collections that optimally fit to individual morphotypes.

The analysis of the available body scan data showed that it is merely impossible to derive rules for the automatic morphological classification. On the other hand, experts are able to classify shapes when looking at different scans (front view, side view etc.). The Case-Based Reasoning (CBR) approach seems therefore rather promising [1]. The idea is that experts evaluate visually a representative number of selected data sets (body scans) according to the newly developed morphological classification scheme. Automated, case-based classification then takes advantage of the expert knowledge by comparing a new scan with already classified ones and by deriving the new classification from the most similar ones.

The paper therefore reflects this approach and is organized as follows: The second chapter after this introduction analyses current morphological classification systems and their deficits. A new scheme is derived and described. Next, the chapter on Case-Based Reasoning for morphological classification goes into the details of how to apply CBR for the classification problem at hand. The next chapter presents some first results from experiments with larger numbers of cases and describes the benefits for garment making industry and for the end users. A brief summary and outlook concludes this paper.

2. Morphological Classification Model

Over the past years there have already been several approaches to characterize morphotypes (without using CBR methodologies) and to develop a classification scheme with respect to the needs of the Fashion industry which have some significant deficits: Most of them are only regarding the full body, and are therefore not appropriate for the classification of single body parts as upper or lower body that are indispensable for fit optimization [3, 4]. Furthermore they classify only by analyzing length and width dimensions and not regarding the complete 3D shape from body scanners [5, 6]. Their classification schemes have only been developed by regarding subjective and visual assessments. Most of them consider only single measurements, but no measurement relations and correlations between different classifications are accounted. As they altogether relied on expert classification which comes along with an enormous expenditure of time they are not validated on a higher number of test persons [7].

Morphological classification on base of CBR philosophy requires a careful setup of the morphological classification system: The case base for expert classification has to include numerous body shapes to be appropriate for global markets as well as for specific target groups. The body shapes classified by experts need to cover all types off possible shapes and combinations. In addition, the system needs to be scalable so that each of the classification levels can be used independently either for full bodies or single body parts as well as for all fashion product categories with their different requirements on regarded measurements and shapes. A simple extensibility is essential for transferability on up-to-date data pool coming from survey with body-scanning in different countries and on diverse ethnic groups. An industrial consultancy project for the US market proved this assumption: three additional classification levels had been added to reflect differences within the market segment "Plus Sizes".

The consideration of the 3-dimensional human body needs provision for length, width, girth, depth measurements and body ankles. The additional regard to visual aspects of the 3D-shapes cannot be simply derived from body measurements but needs to regard various correlations and appropriate proportions. The morphological classification system has to be validated on a statistically representative data base. Various approaches have proven that a maximum of three different characteristics within one classification level can be differentiated by the experts to ensure unique expert classifications [1].

The starting point for the development of the classification scheme was a detailed analysis of the variance of different key measurements in correlation with the most important primary dimensions. If the variance is not covered by the actual range between the sizes an additional morphotype classification is needed. Based on literature and scan review, a morphological classification scheme was developed. It comprises the following 10 levels or features and their according ordinal scales.

Classification Full Body: `Basic shape, relation shoulder and hip, Waist shape, Posture, Waist position`

Classification Upper Body: `Arm shape, Shoulder slope, Relation bust and back`

Classification Lower Body: `Leg Length, Leg shape, Buttock shape`

Thorough analysis revealed that an ordinal, categorical scale fits best. This means that 3 (for one level only 2) possible values are defined and the values have a natural order. This is an important point because otherwise operations like variance and mean value would not be feasible and the CBR-model would be more difficult and less accurate. The following figure shows an example.
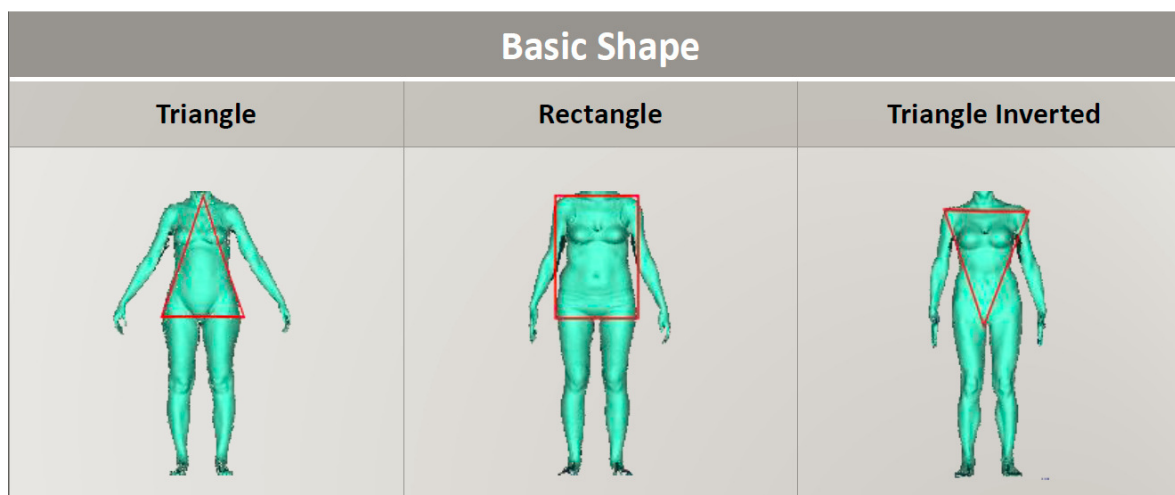

Fig. 2. Classification level "Basic Shape" and the ordinal scale with 3 possible values.

3. Classification by Experts

The heart of the CBR-approach is the case base, i.e. a number of solved problems. Solved means here a scan successfully classified by experts. The problem description are the various images of the scan, the solution is the morphological classification. In order to get good and consistent results from the experts, the following procedure was deployed:

1. **Classification Round 1: 4 experts classify ~200 scans**
2. **Statistical analysis and feasibility check of the results**
3. **Re-classification of the most critical 20 scans**
4. **Classification Round 2: 4 experts classify ~ new 200 scans and 10 already classified scans in order to test consistency**
5. **The resulting case base consists of app. 420 approved cases**

A web-based application allows remote classification for selected experts. They all had to choose the most suitable value without seeing others' results, no in-betweens are allowed. The following figure shows a screenshot of the application in German.
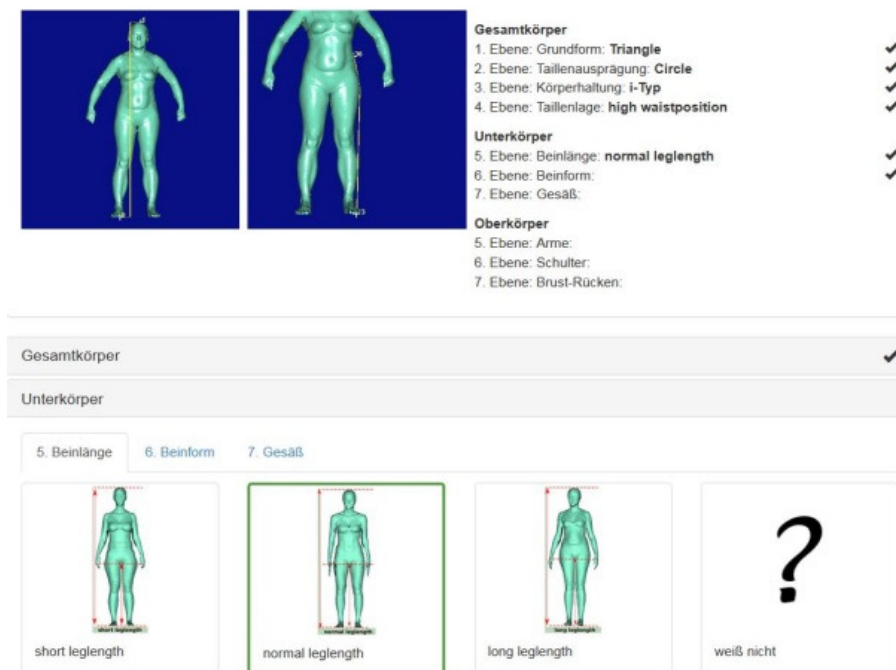

Fig. 3. Screenshot of the web-based classification tool.

In case of real uncertainty experts could select "Cannot classify". These scans had then been discussed in the expert team, sometimes there was a technical problem with the scan not being accurate.

The result of the experts' classification is a classification vector for each scan and each classification level. In case of three experts voting for "Triangle", one for "Rectangle" and none for "Inverted Triangle", the resulting vector looks like this:

Scan ID: 123 – Level: Basic Shape – Classification: (3; 1; 0)

Apparently this means that the experts more or less agree. If they would not, the vector would be e.g. (2; 0; 2). Such a vector has a high variance, while the first one (3; 1; 0) has a small one. The statistical analysis was needed to check a number of aspects:

- Experts: The analysis showed that all of they had the same bias, i.e. all experts had more or less the same average classification and no tendency to classify in a different direction than the others.
- Variance within a scan: if all levels of a scan showed a high variance then the scan was marked as critical and reviewed manually.
- Variance within a level: if some levels showed a much higher variance than others, the meaning of the level values were reviewed and discussed with the experts (e.g. what means "Triangle" and where the difference to "Rectangle" is). The average variance of a single classification is more or less equivalent to a (3, 1, 0) classification, this means that in average 3 out of 4 experts choose the same value and one a neighboring value.
- Correlation: apparently some levels are not fully independent. A correlation analysis revealed correlation factors ranging from 0 (no correlation, independency) to 0.4 in absolute numbers (low – medium correlation). Therefore all the levels are necessary; none are fully dependent on others.

The resulting case base of 420 approved classified scans is the starting point for the automatic classification by the CBR-system.

4. The CBR-Cycle in practice

The philosophy of CBR follows the human reasoning: instead of formulating complex rules, humans tend to remember similar situations and to derive decisions from that situations [8]. This happens when a doctor makes a diagnosis based on similar previous diagnoses with similar symptoms. The same holds for bankers and their credit decisions. The challenge of the morphological classification lies in the plurality of the experts' opinion. If the classifications are substituted by the numerical weighted average, then the classification vector (1; 2; 1) would have the same value as the vector (0; 4; 0) and as the vector (2; 0; 2). This loss of information is prevented by maintaining the vector throughout the whole CBR-process. The approach follows here the well-known, established CBR-Cycle [8].
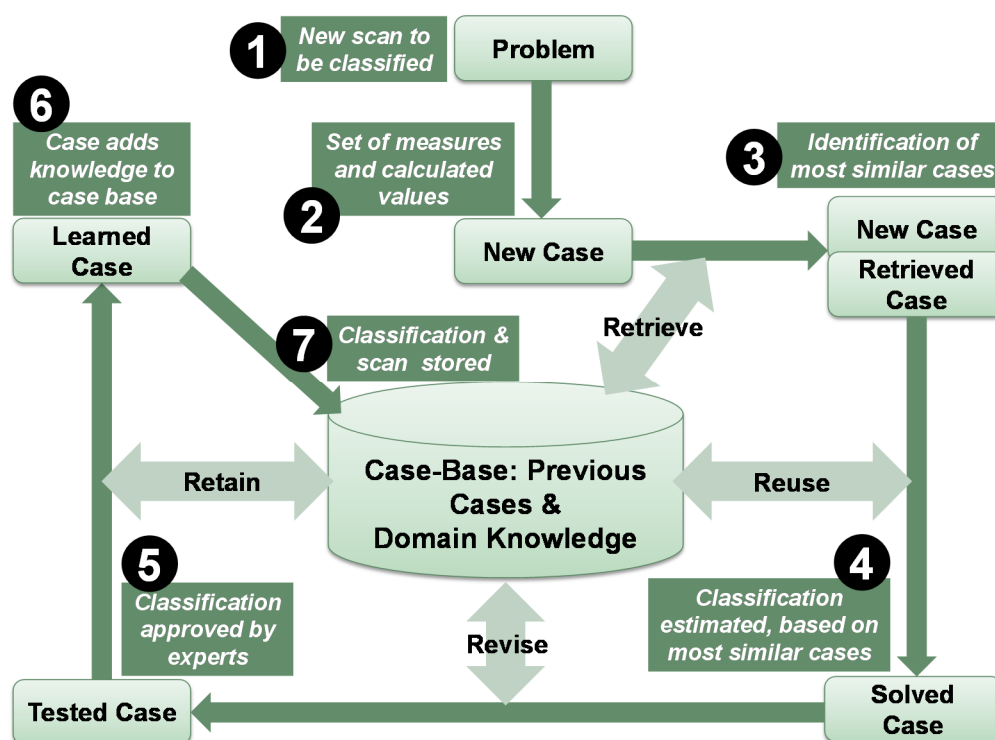

Fig. 4. The CBR-Cycle for morphological classification, adapted from [8]

A case represents a complete episode and comprises problem and solution. The problem is the scan which consists of number of measurement data from the body scanner. The solution to the problem is the morphological classification according to the scheme described above. The CBR system follows the layered approach: each morphotype level is treated separately. This is reasonable because morphotypes are to some extend independent of each other. The correlation analysis confirmed this assumption. This means for example that for the classification of the arms scans with similar measures related to the arms are retrieved, independent of the waist, legs or shoulder because scans differing in weight or height can still have very similar arms. The similarity between two scans is therefore never calculated as a whole but always with respect to a given feature or classification level, answering questions like: How similar are two scans to each other with respect to the feature "Arm Shape?"
Not all measurements are relevant for all levels. Thus, each layer has its own similarity model. These models take into account not only direct measurements but also calculated values. Therefore the similarity is based on distances, girths, angles, differences and quotients. Here is an example: the similarity function of the feature "Relationship Shoulder – Hip" comprises the following data:

Measures: Shoulder Length, Neck Base Diameter, Trunk Length, Hip Girth, Weight, Height, Shoulder Breadth, Hip Breadth

Calculated Value: 2 *Shoulder Length + Neck Base Diameter) / Hip Girth, Shoulder Breadth / Hip Breadth

Not all measures are equally important for the similarity of a level. Therefore, the feature model assigns weights to the features. All other classification features have analogue models. The number of data varies between 5 and 25, where some measures like Hip Girth are part of more than one feature model.

The definition of the models is one of the most important knowledge containers of the approach [8]. The choice of measures, the definition of calculated values such as quotients and the assignment of weights is crucial for the success of the approach. Therefore, a lot of emphasis was put on the development and fine-tuning of the models.

The similarity between two scans according to the feature-specific models is calculated as the weighted sum of the local similarities between two single measures or calculated data [9]. Only numerical, no categorical attributes are used. The similarity as weighted sum thus looks like the following.

```
Similarity Shoulder Hip (Case 1, Case 2) =
weight (Shoulder Length) * sim (Shoulder Length (Case 1), Shoulder Length (Case
2))
+ weight (Neck Base Diameter) * sim (Neck Base Diameter (Case 1), Neck Base
Diameter (Case 2))
+ weight (Trunk Length) * …
```

The local similarity can be calculated with a number of different predefined functions available in the CBR-Toolbox MyCBR [10]. For numerical values symmetric polynomial functions seem to work best and are deployed in the system. The following figure shows an example for such a local similarity between two numerical values.
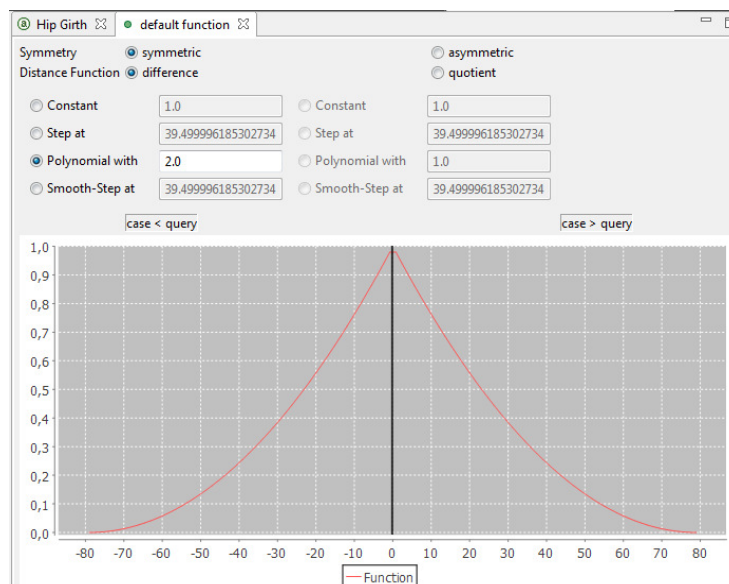

Fig. 5. Local similarity function

The similarity functions are the most important knowledge containers of a CBR-System [8, 9]. Therefore the functions were optimized in a dedicated process:

1. **Experts agreed on the values and calculated values to be deployed**
   → set-up of the basic similarity model
2. **Manual tests and expert knowledge was used to estimate the weights of the similarity function, i.e. which measure is more important than the others?**
   → initialization of the similarity model
3. **Definition of test scenario auto-classification: classification of the already manually classified scans and comparison of the classification by the experts with the classification by the system**
   → definition of a quality function
4. **Definition of an algorithm that improves the result of the auto-classification**
   → optimization of the similarity functions

This process ensured that the similarity functions are tuned to classify in a way very close to the human experts. The test runs show a rather good reproduction of the expert classification. Depending on the given level, the reproduction rate is higher than 85%. Next, the CBR-process could be started.

The RETRIVAL phase of the CBR-Cycle makes use of the described similarity functions and identifies the most similar cases, always with respect to one feature. Based on a number of experiments, the number of retrieved cases was limited to 10 – as long as they have a similarity above 90%.

The REUSE phase estimates the classification of a new query by using these most similar cases. As the result needs to be a plurality vector again, the system calculates the mean value of the classification vectors of the similar cases. This means that the classification of the system may be a vector like (1.2; 2.1; 0.7), while the experts' classification vectors have only integer values.

The following figure shows an example where the system identifies 4 similar cases with similarity > 90%. The "Leg Shape" of these scans was classified (4; 0; 0), (3; 1; 0), (3; 1; 0), and (4; 0; 0). The resulting classification of the new scan is therefore (3.5; 0.5; 0).

New problem:	010-140116				
feature:	Leg shape				
Scans out of case base		Scan1	Scan2	Scan3	Scan4
Similarity (based on measures and relations)		97%	95%	95%	94%
Identified morphotypes					
middle thigh, middle calf		4	3	3	4
strong thigh, middle calf		0	1	1	0
strong thigh, strong calf		0	0	0	0
Amount:	sum	%			
middle thigh, middle calf	14	83,3%	3.5		
strong thigh, middle calf	2	16,6%	0,5		
strong thigh, strong calf	0	0%	0		
result:		regarding to the classification level „leg shape" the analyzed person correlates to 83,3% to the type "middle thigh, middle calf" and 16,6% to the type "strong thigh, middle calf".			

Fig. 6. REUSE-phase, example with 4 similar cases

The REVISE phase picks up the estimation of the CBR-System and applies it to the problem (query) at hand. This phase is currently in progress. Almost 6000 scans have been classified; a revision algorithm that automatically checks for inconsistency is currently developed. First results show that the automatic classification is reasonable, the results below base on the REUSE procedure described above.

The RETAIN phase is the last step of the CBR-process. New cases (approved and tested) will be added to the case base if they bring an added value and make the case base learn. This phase is currently in progress.

5. Results

The described approach has been proved to be valid and comprehensive. It is flexible and extendable because each classification feature has separate similarity and retrieval functions with vast expert knowledge embodied and can be linked to various case-bases. Time consuming individual morphotype classification can be replaced by CBR technology and supports fashion product development as well as size recommendation in online retail.

The most relevant question for the industry is the frequency of morphotypes in the different classification levels. Therefore diverse kinds of analysis with increasing complexity are possible. One opportunity is the evaluation of the frequencies and market shares regarding one single feature. In the following diagram the market shares for the different basic shapes regarding the SizeGERMANY market for women is presented.
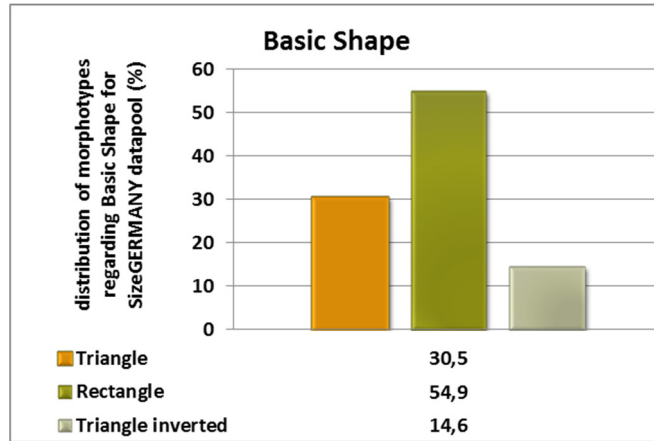
Fig. 7. Distribution and market shares of different Basic Shapes

Charts like shown in Fig. 7 give information regarding the most frequent shape in the market and show on which shapes should be focused on to have a high coverage of the market be guaranteed. An analysis like this can be done for each of the classification levels.

Furthermore the analysis regarding the combinations out of two different classification levels gives information regarding correlations of body shapes. As the example in Fig. 8 shows a correlation between the basic shape and the leg shape could be identified. Triangle basic shapes often come in combination with strong leg shapes like "strong thigh, strong calf". Women that have a triangle inverted basic shape tend to have slimmer legs that correspond to the morphotype "middle thigh, middle calf".
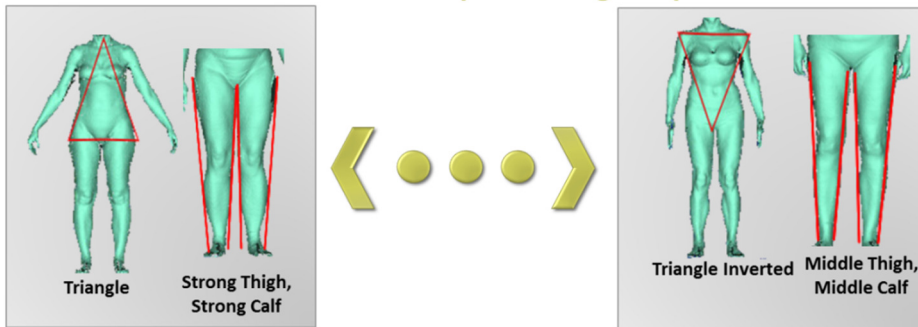

Fig. 8. Example for identified correlations between the Basic Shape and the Leg Shape

Other classification levels can also be fully independent from each other classification levels. The morphotypes for classifying the different shoulder slopes for example do not show any correlation to the other classification levels. So the calculated correlations between the different combinations of classification levels show very different values.

To describe the body shape in detail even three different classification levels can be combined with each other and be analyzed. Fig. 9 shows the frequency of the morphotype combinations regarding the basic shape, the waist shape and the leg shape. The different colors mark areas with higher and lower market shares. The resulting matrix contains eighteen different combinations of body shapes.

SizeGERMANY, DOB 18 – 75 years
Total: 5555

Basic Shape	Waist Shape			Leg Shape			
	Waist Shape	Basic Shape	Waist Shape	middle thigh, middle calf	strong thigh, middle calf	strong thigh, strong calf	Total
Triangle	Hourglass	30,51%	66,17%	8,11%	7,17%	5,75%	21,04%
	Circle	30,51%	33,83%	3,61%	3,05%	2,82%	9,48%
Rectangle	Hourglass	54,91%	66,17%	17,78%	10,62%	7,09%	35,49%
	Circle	54,91%	33,83%	10,13%	5,27%	4,02%	19,42%
Triangle Inverted	Hourglass	14,57%	66,17%	5,52%	2,58%	1,56%	9,66%
	Circle	14,57%	33,83%	2,97%	1,20%	0,75%	4,91%
Total				48,12%	29,89%	21,99%	100,00%

Fig. 9. Distribution of morphotype combinations regarding basic shape, waist shape and leg shape

Rectangle is the most frequent basic shape with a market share of 54.9%. More than 66% of the analyzed women tend to an hourglass waist shape. The distribution of the different waist shapes stays constant within the basic shapes. The leg shape "middle thigh, middle calf" has the highest frequency regarding the available leg shapes although the other shapes also have big impact on the market as the market shares show. The most frequent morphotype combination regarding the analyzed features is the combination out of rectangle, hourglass, middle thigh and middle calf. Here the market share is nearly 18%. The calculation of market shares regarding three features can be done with any of the defined classification levels.

By additional differentiation of age group the changes in waist shape distribution form younger to elderly women can easily be seen. Comparisons of frequency distribution between different sizes show the shift in morphotypes from smaller to larger sizes.

Fig. 10 shows an example of how the waist shape changes within one size row. For the small German sizes 32 to 42 most of the women have an hourglass shaped waist. For the bigger German sizes this distribution shifts significantly. For size 44 to 60 the morphotypes hourglass and circle have nearly identical market shares.
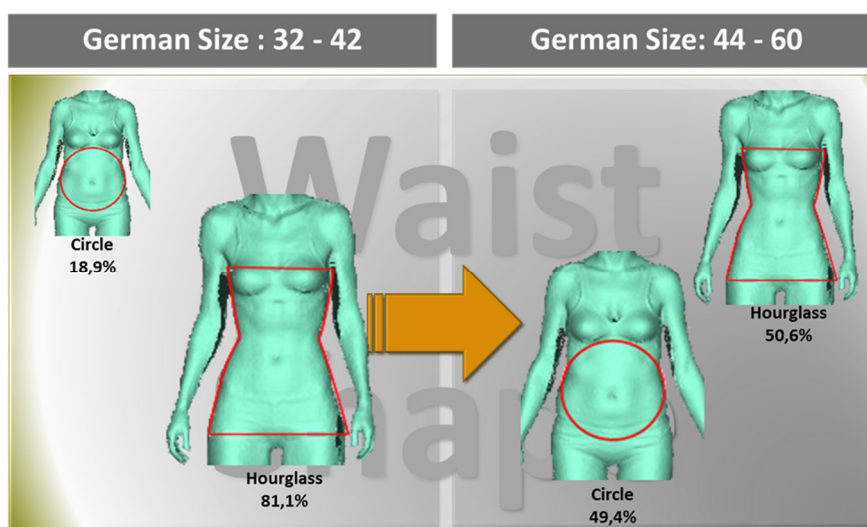

Fig. 10. Comparison of different morphotype frequencies for smaller and bigger sizes.

In the meantime the scientific approach has been evaluated within an industry project for Plus Sizes in der US market with the following results: Different populations with additional shape characteristics need additional classification parameters that can be effectively and easily added to the general classification scheme. Significant changes in morphotypes as well as in single classification levels or classifications can be established between different age groups or between different sizes by calculating the frequency distribution in case of market shares. Finally the combined frequency distribution shows the most important morphotypes as a combination of different classifications.

The results of the morphotype classification lead to the following findings: For obtaining higher market shares with a constant number of sizes morphotypes have to be regarded (size system development based on morphotypes, differentiated into upper and lower body). The size range and the composition of sizes between the different height types have to consider the different frequency distributions per morphotype to maximize the market shares. If companies split their collection into different height types (petite, regular, tall) the grading needs to be differentiated between the height rows (e.g. different thigh girth grading between petite, regular and tall size rows). Some morphotype classifications change within one size range or height type (petite, regular, tall) so that a non-linear grading is indispensable. Another issue is the analysis of different morphotype frequencies within the age groups. Due to different frequency distributions of morphotypes and morphotype combinations in the different age groups basic block concepts have to be adjusted with respect to the individual target group of the company.

In the future size system development can be based on morphotype definition and selection so that special basic blocks and fits can be developed with the respect to the most market relevant body shapes. Body measurement portals will allow statistical analysis and size table optimization with respect to single morphotype features or for a target group characterized by a combination of different classification features (Fig.11).
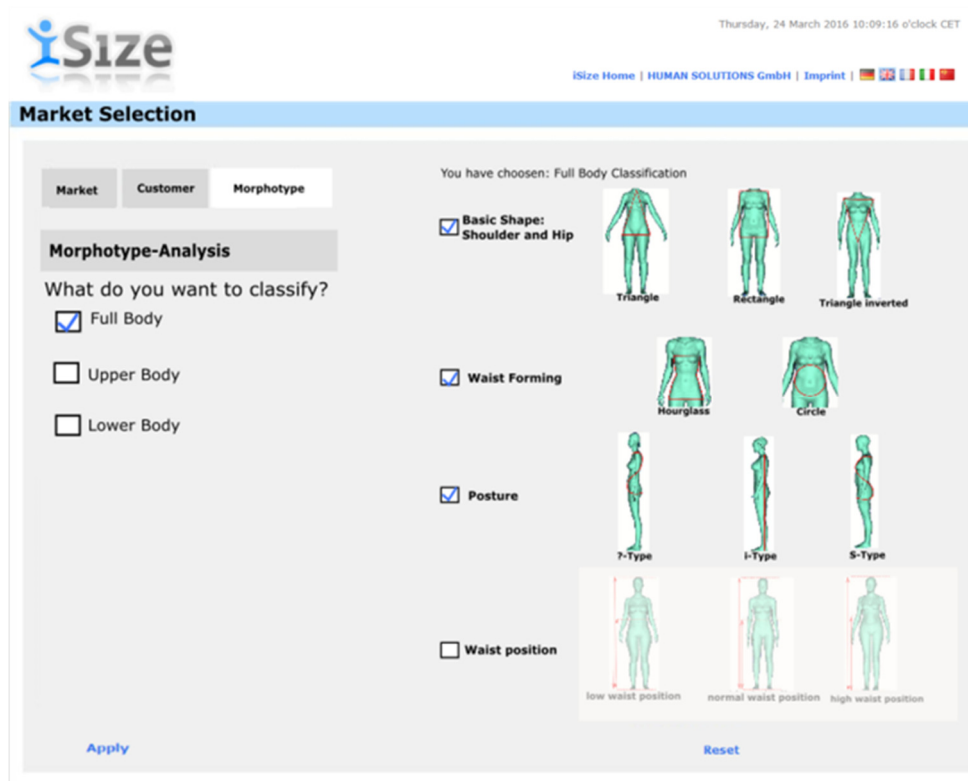
Fig. 11. Example: Definition of target customer profile with respect to morphotypes in the iSize body measurement portal from Human Solutions

Fashion web shops like for example the one in the figure below will use morphotype features to characterize and classify their target customers and to recommend especially those products that will due to their fit characteristics optimally match with the body shape if the individual customer. This will effectively reduce returns rates in fashion online business and help Fashion retailers to achieve a better fit, to enlarge their market potential and at the same time increase customer satisfaction.
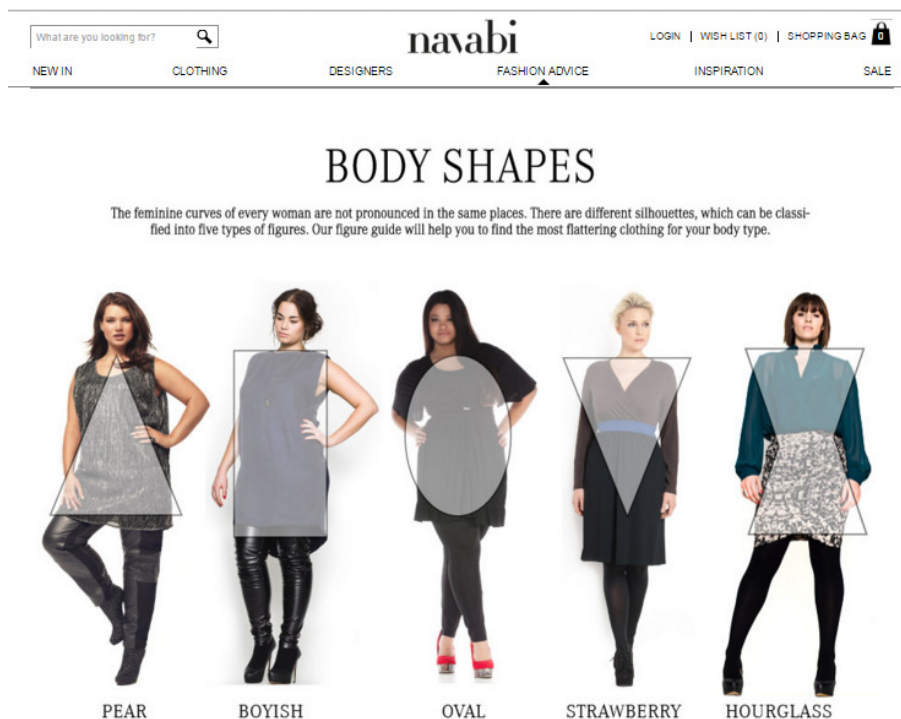

Fig. 12. Morphotype descriptions from navabi used for finding the optimal clothes for the customers´ shape [11]

6. Summary and Outlook

The novel approach presented in this paper allows the morphological classification of body scans by making use of a case base with scans classified by experts. This CBR approach has proven to be valid and useful for the garment making industry. Due to the increasing diversity of human bodies, there is a need for morphotype analysis being part of standard analysis for size surveys. Differences regarding various populations like different ethnic groups, age groups or regarding other socio demographic criteria can be analyzed with relatively small effort. The morphological classification has thus the potential to revolutionize size system development and pattern construction for the fashion industry. Individual customer advertisement by morphotype will be possible. As online shopping is getting more and more important an optimized size recommendation with respect to frequent morphotype combinations of a target group should be next possible steps out of the project. The goals of this achievement are decreased number of returns because of satisfied customers and therefore an increased customer loyalty. All in all morphotype classification contains a huge potential for scientific approaches with lots of benefits for apparel industry. With respect to the advancing customer demands it will be indispensable to for the apparel industry to focus on morphotype specific product development.

Acknowledgement

References

[1]   T. V. Fischer et al., "Towards case-based morphological classification for fashion product development." presentation IT4Fashion Conference, Florence, 2016 (to be published

[2]   J.-M. Surville, S. Herichi, V. Segalen, "HOAXY Body Shapes and Fashion Formula", in *International Conference on 3D Body Scanning Technologies*, Lugano, Switzerland, 19-20 October 2010, pp. 245-258, doi:10.15221/10.249.

[3]   A. Rissiek and R. Trieb, "iSize-Implementation of international anthropometric survey results for worldwide sizing and fit optimization in the apparel industry", in *International Conference on 3D Body Scanning Technologies, Lugano, Switzerland, 19-20 October 2010,* pp. 269-281, doi:10.15221/10.269.

[4]   S. F. Bastos, F. G. Sabrá Senai Cetiqt, "The Body Shape of Brazilian Women", in *5th International Conference on 3D Body Scanning Technologies, Lugano, Switzerland, 21-22 October 2014*, pp. 155-166.

[5]   S. Morlock S, E. Kirchdörfer, M. Rupp, E. Wendt, S. Krzywinski, H. Rödel, *Grundsatzuntersuchung zur Konstruktion passformgerechter Bekleidung für Frauen mit starken Figuren, Abschlussbericht zu AIF-Forschungsvorhaben*, Dresden 2009, pp. 48-53.

[6]   M. Duffy, *HOAX Fashion formula, Dress the Body Type You Have to Look Like the Body You Want*, THE BODY PRESS, Tucson, Arizona 85704, 1987, pp. 20-63.

[7]   J. M. Webster, J. Cornolo, Y. Kelkel, "Comparison of Female shape Analysis Methods for the Development of a New Sizing System", in *3rd International Conference on 3D Body Scanning Technologies*, Lugano, Switzerland, 16-17 October 2012, pp. 280-287, doi:10.15221/12.280.

[8]   A. Aamodt and E. Plaza. "Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches", in *AI Communications* 7(1) 1994, pp. 39-59.

[9]   M. Lenz, H.-D. Burkhard, B. Bartsch-Spörl, S. Wess, (editors*), Case-Based Reasoning Technology. From Foundation to Application*, Berlin; New York: Springer; 1998

[10] MyCBR. [Internet], http://mycbr-project.net/., accessed 2016 March 24

[11] Navabi, *Body shapes* [Internet], https://www.navabi.co.uk/fashion-advice/#figurberatungCont., accessed 2016 March 24