A new method for visual descriptor evaluation

Horst Eidenberger*
Vienna University of Technology, Institute of Software Technology and Interactive Systems,
Favoritenstrasse 9-11, 1040 Vienna, Austria

ABSTRACT

Evaluation in visual information retrieval is usually performed by executing test queries and calculating recall and precision based on predefined media collections and ground truth information. This process is complex and time consuming. For the evaluation of feature transformations (transformation of visual media objects to feature vectors) it would be desirable to have simpler methods available. In this paper we introduce an evaluation procedure for features that is based on statistical data analysis. The new idea is that we make use of the existing visual MPEG-7 descriptors to judge the characteristics of novel feature transformations. The proposed procedure is divided into four steps: (1) feature extraction, (2) merging with MPEG-7 data and normalisation, (3) statistical data analysis and (4) visualisation and interpretation. Three types of statistical methods are used for evaluation: (1) description (moments, etc.), (2) identification of similarities (e.g. cluster analysis) and (3) identification of dependencies (e.g. factor analysis). From statistical analysis several benefits can be drawn for feature redesign. Application of the evaluation procedure suggested and advantages of the approach are shown in several examples.

Keywords: Evaluation, Benchmarking, Statistical Data Analysis, Feature Design, Visual Information Retrieval, Content-based Image Retrieval, Content-based Video Retrieval, MPEG-7

1. INTRODUCTION

Visual information retrieval[4, 18] (VIR) is the art (or alchemy?) of deriving semantically meaningful (high-level) information from binary (low-level) pixel data. The goal may vary: Often it is to find media objects in a database that are (semantically) similar to given examples. VIR is generally based on two principles: the vector space model and visual feature transformations. The vector space model is a product of text information retrieval science. Documents are interpreted as vectors in a high-dimensional space (one existence/non-existence dimension per term). This space (called feature space) is assumed to have a certain geometry (usually Euclidean but generally, any Riemannian geometry can be thought of). In consequence, it is possible to measure (dis-)similarity of documents as vector distance in feature space.

Adopting the vector space model for VIR requires defining visual feature transformations, i.e. a projection from a visual object (an image = a matrix of colour values, or a video = a spatio-temporal cube of colour values) to a numerical vector. In recent years a considerable number of visual feature transformations has been proposed. Among the most prominent are the visual MPEG-7 descriptors[5]. Generally, often used techniques include histograms (e.g. colour histograms), gradient methods (e.g. directionality of textures, camera motion in videos) and segmentation methods (e.g. edge detection and calculation of moments for shape description). A recent comprehensive survey on visual feature transformations has been conducted for the IST SCHEMA project of the European Union[16].

The reason why so many feature transformations have been proposed is that, so far, none of them (nor combinations) have accomplished to imitate human similarity perception properly. This is partially due to the fact that the quality of feature transformations cannot be evaluated directly: in VIR quality indicators are derived from query results (see Subsection 2.1 for details). Therefore, it is not possible to identify how well a certain feature performs but only how well a VIR system performs that makes use of this feature. Feature designers feel that being inadequate.

In this paper we introduce an evaluation procedure intended for *direct* evaluation of features. The method is based on statistical data analysis. It makes use of statistical transformations and visualisation techniques to compare a new feature transformation to existing ones on a common visual data basis. Therefore, in terms of information retrieval theory the proposed evaluation technique is a systematic measure[10]. It looks at feature quality from the system viewpoint (in contrast to the user viewpoint adopted in example queries) and can be applied to answer a wide range of questions on

* eidenberger@ims.tuwien.ac.at; phone 43 1 58801-18853; fax 43 1 58801-18898; www.ims.tuwien.ac.at

feature quality (see Subsection 3.2 for details).

The paper is organised as follows. Section 2 gives background information on evaluation in Visual Information Retrieval and the statistical methods for data analysis used in the proposed evaluation procedure. Section 3 sketches the basic idea and possible applications. Section 4 clarifies technical details including the common data basis and reference features. Finally, Section 5 shows the usefulness of the method in several examples.

2. BACKGROUND

2.1 Evaluation in visual information retrieval

Traditionally, VIR evaluation follows a four-step process[4, 18]: (1) Construction of an evaluation dataset and ground truth information, (2) construction of a test environment, (3) execution of a sufficiently large number of test queries and (4) judgement of method quality as retrieval quality by indicators. Even though this seems to be a straightforward procedure, discussion will show that it contains a lot of open questions and hidden problems.

The construction of evaluation datasets is the easiest step of the procedure. Several activities exist to provide researchers with free media test data (e.g. Benchathlon[1]) and existing media collections from related research areas (e.g. Computer Vision[6]) can also be used. Most existing collections consist of images but in recent years more and more free video collections are becoming available (e.g. the TREC datasets[17]). Building a ground truth can vary from defining groups of similar objects within the data to pair-wise similarity judgements saying that some object *A* is more similar to a certain reference object *R* than a third object *C*. Similarity can be judged on a scale from a low syntactical point of view (similar colours, etc.) to a highly semantic point of view (similar context). Ideally, the ground truth should be globally acceptable and invariant against cultural, social and other dependencies. Already this brief description shows that defining an acceptable ground truth is a non-trivial task.

If a new feature has to be evaluated by retrieval indicators, an environment for testing is needed. This includes methods for media access and transformation (e.g. resolution reduction) and a query engine that applies the feature on test collections. Plainly, the existence of a querying environment influences the performance of the feature. If it is not suitable for the querying paradigm or the implemented querying method performs generally bad, the retrieval quality of the feature is biased. Besides, integrating a feature in an existing querying environment can easily become a complex and time-consuming task.

The execution of test queries should ideally be done automatically. Most querying paradigms support non-interactive querying based on examples. To guarantee statistical correctness a sufficiently large number of test queries has to be executed and query examples should be chosen randomly. The result of each query is a set of retrieved media objects. From the result sets quality indicators can be derived by using the ground truth information. Usually, recall and precision (and derivates like ANMRR for MPEG-7[15]) are used as quality indicators:

$$recall = \frac{|retrieved \cap relevant|}{|relevant\ objects|}, precision = \frac{|retrieved \cap relevant|}{|retrieved\ objects|} \quad (1)$$

Recall and precision are dependent measures and have to be used in combination (e.g. in a recall precision graph). Since both indicators can be optimised if the other is not considered the value of one indicator alone is meaningless. Unfortunately, recall and precision are often not used properly. Sometimes, the quality of a feature is only illustrated by a handful of visual query examples.

2.2 Statistical analysis of high-dimensional data vectors

Various methods exist for the statistical evaluation of the population of high-dimensional vector spaces representing the properties of real-world and abstract objects. If data is given in form of normalised vectors (e.g. as a data matrix of cases and properties) three fields of analysis are mainly relevant in the context of visual information retrieval: (1) description of cases, (2) similarity-based clustering of cases and (3) identification of dependencies of cases.

Cases (e.g. the feature vectors of media objects) can easily be described by moments of first (mean, distance) and second

order (variance). Additionally, distributions can be computed to estimate the likeliness of the occurrence of certain property values. These indicators can be used to test hypotheses on the characteristics of the process that was used to create the data vectors (e.g. the used feature transform).

Cluster analysis[12] aims at inducing a human-understandable structure in a population (often also used for visualisation). Generally, similarity based clustering can be performed in any space. If the original feature space is metric it can also be interpreted as a clustering of the data vectors. On the other hand, relevant for analysis are only those cluster structures that can be interpreted by humans: 1D, 2D, at most 3D. Consequently, the most popular cluster algorithms generate 1D and 2D cluster structures. One-dimensional cluster analysis methods include top-down and bottom-up (agglomerative) approaches and can be visualised as tree-like graphs (so-called dendrograms). Two-dimensional methods generate maps. Famous examples are the Self-Organizing Map[13] and Sammon mapping[7]. The first is intended for data analysis while the second is mainly used for visualisation.

Identifying dependencies is useful to judge the information quality of data vectors (redundancies, etc.). For high-dimensional vectors the task is not as easy to fulfil as it is in the two-dimensional case. The application of simple methods like linear regression is not possible because of the computational complexity. Heuristic methods (e.g. factor analysis[14]) are used instead. One example is the principal component analysis that makes use of the Eigenvalues of the data matrix to identify linear relationships (dependencies) of data vectors. Since these methods make several assumptions on the given data (e.g. Gaussian distribution) they are able to identify only certain types of dependencies. Still, factor analysis can give valuable feedback on the quality of visual information retrieval methods.

3. IDEA & DISCUSSION

Below, the new ideas in the proposed evaluation method for feature transformations are described. All technical details, including workflow, reference data, etc. are described in Section 4. Possible applications are derived in Subsection 3.2. Example applications are sketched in Section 5.

3.1 Motivation and idea

In Subsection 2.1 we outlined the evaluation scheme traditionally used in (visual) information retrieval. Essentially, recall and precision indicators are calculated from well-known media collections for sample queries. As we pointed out this process has its pitfalls and leaves the feature designer with a handful of open problems. The most significant one is defining a ground truth that reflects human similarity judgement independent of cultural aspects and all other human peculiarities.

Two further, practically relevant problems of the traditional approach are the evaluation of feature transformations for non-retrieval applications and the complexity of the evaluation process. Feature transformations are nothing retrieval-specific. They are also used in other application domains including browsing, pattern recognition and computer vision. As an example, the content-based features defined in the visual part of the MPEG-7 standard[5] are intended for retrieval *and* browsing applications. One feature transformation (Texture Browsing) was even designed exclusively for browsing applications. Still, in lack of alternative measures, these features were evaluated using a retrieval rank measure. In consequence, their quality is assured for retrieval applications but not for browsing applications. Using some of the methods proposed in this paper the author could show that the visual MPEG-7 features contain significant weaknesses in terms of data quality[7]. In conclusion, an application-independent evaluation procedure would be desirable.

The traditional evaluation procedure is a heavyweight process. In order to analyse a feature it is necessary to embed it in a querying framework and run hundreds of queries to calculate statistically valid quality indicators. This process has to be repeated after any change in the feature transformation. It is very time-consuming and – in case of undesired results – can be very annoying for the feature designer, because it does not provide any hints on problems in the feature transformation. With recall and precision it is possible to say whether a feature performs well but not *why* it shows the observed behaviour.

In contrast, the proposed evaluation procedure is a lightweight process. We make use of statistical data analysis methods to evaluate the quality of feature transformations. As for the traditional approach the feature transformation is computed for a predefined media collection. The resulting feature vectors are investigated for characteristic properties (e.g.

variance), common properties and similarities (based on the methods listed in Subsection 2.2). A querying framework and ground truth information are not needed. Quality indicators are derived directly from the feature data and allow conclusions on the quality of the extraction process.

Using statistical data analysis on high-dimensional vectors is not a new idea: it has been done many times before. The new element of the proposed evaluation process is that the feature in question is not just compared to itself but also to *reference data*. This reference data is provided by the MPEG-7 standard. Visual MPEG-7 features are calculated for the predefined media collections. By comparing the feature vectors of the evaluated feature transformation to the reference data it is possible to gain additional insights on the characteristics of a new feature (see Subsection 3.2 for details).

This evaluation procedure has several advantages: Firstly, measurement is done in a systematic way: one system (feature) is compared to another. Since the process is independent of the user (no user input required) the results are objective. Secondly, results are application-independent. General data quality is measured instead of retrieval quality. If a feature shows characteristics making it independent of the reference features then it is obviously valuable. Every feature transformation that adds non-redundant information to feature space is worth being utilised for any type of examination of feature space (browsing, retrieval, etc.). Thirdly, as pointed out above, this evaluation procedure gives information on possible weaknesses in the feature transformation process. If a feature is highly redundant by itself then, apparently, the transformation is not able to provide enough discriminance. Additionally, since well-documented reference data are used, the evaluation procedure gives information on the context of a feature. For example, if a feature is highly similar to existing texture features then it is obviously sensitive for texture information (the distribution of light), even though it may have been designed to be a colour feature (the existence of light). Finally, no querying framework is needed to apply this evaluation method. All necessary steps can be fulfilled with mathematical/statistical standard software (e.g. SPSS, Matlab).

3.2 Applications

The proposed evaluation method is not intended to replace recall- and precision-based evaluation but to supplement the existing workbench for evaluation in visual information retrieval. Using statistical evaluation a variety of questions including the following can be answered:

- What is the type of the new feature? With respect to the MPEG-7 norm, is it a colour, texture, shape or motion feature or does it define entirely new criterions for visual media?

- How robust is the new feature against rotation, scaling and other visual media transformations? Are the feature vector elements still similar after transformation? If not, do transformations change the characteristics of the feature transform?

- How robust is the new feature against noise? Do the characteristics of the feature vectors change if the media objects are noisy?

- Is the feature mapping surjective? If the feature transformation is applied to two collections of similar media objects, are the corresponding feature vector elements similar?

Apparently, these questions are much easier and clearer to answer with statistical methods than by recall and precision. In Section 5 we are giving some examples on how the proposed evaluation procedure can be applied.

4. EVALUATION PROCEDURE

The following subsection describes the workflow in the proposed evaluation procedure. It is a four step process that requires the definition of media collections for assessment and the computation of MPEG-7 descriptors. Subsection 4.2 describes the data basis used for the examples in Section 5. Finally, Subsection 4.3 gives remarks on software tools used for feature extraction and statistical evaluation.

4.1 Workflow

Figure 1 depicts the flow of work in the statistical evaluation procedure. First, the new feature transformation is applied

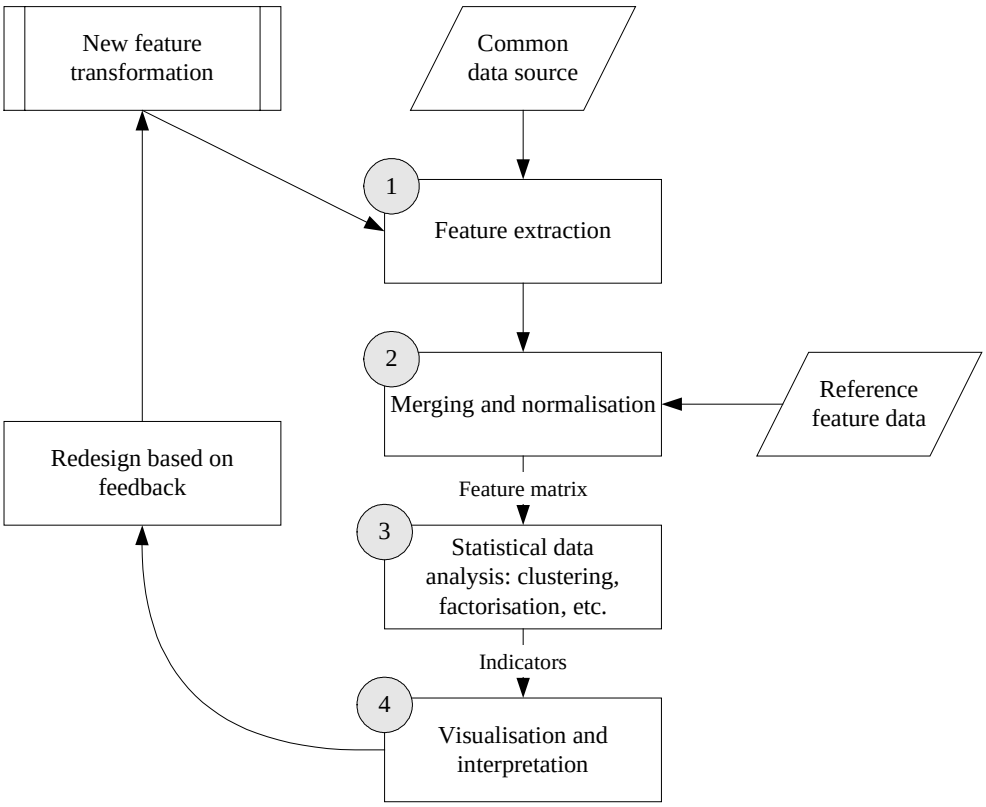

Figure 1: Workflow in evaluation procedure.

to the predefined media collections to extract numerical feature vectors. In the second step these feature vectors are merged with the existing MPEG-7 descriptors for the same collections. After merging, they are normalised to the same interval. This is mandatory to guarantee the correctness of the later applied statistical operations. We propose to use a min-max-normalisation since it preserves mean, variance and distribution of the feature vector elements:

$$\bar{x}_i = \frac{x_i - min_i}{max_i - min_i} \tag{2}$$

min_i, max_i are the minimum resp. maximum value of the *i*-th element of feature vector (x_i). After normalisation, all elements are defined on [0, 1]. All feature vectors together are addressed as feature matrix (elements in columns, cases in rows). On the feature matrix statistical operations are applied and indicators are derived. In the fourth step, these indicators are visualised and interpreted. Based on the interpretation the proposed feature transformation can be iteratively refined.

Various statistical methods exist that can be used for evaluation. Principally, the three main areas relevant for visual information retrieval are (as described in Subsection 2.2) description, detection of similarities and detection of dependencies of/in the feature matrix. In past experiments we found the following methods to be very useful:

- Extraction of moments of first and second order of elements for description as well as computation of a discrete distribution of values for each element. The distribution shows how often each value (down-sampled to a few bits) occurs and allows conclusions on the utilisation of the available scale by a feature.

- One- and two-dimensional cluster analysis of *elements* (not cases!) for similarity assessment. *k*-means clustering and dendrograms for visualisation have proven to be sufficient in the one-dimensional case. Unfortunately, dendrograms become soon unreadable for larger numbers of elements. Therefore, we found two-dimensional clustering with Self-Organizing Maps[13] to be more satisfying. In our opinion Self-Organizing Maps are by far the best available 2D

Figure 2: Common data basis. Left: Brodatz dataset, middle: Corel dataset, right: coats of arms dataset.

clustering technique. In consequence, they will mainly be used in the experiments in Section 5.

- Detection of dependencies of feature vector elements by factor analysis. Eigenvalues extracted from a data matrix by a principal component analysis can be interpreted as hidden factors that have a linear influence on the data values. Elements (media properties) that are significantly influenced by the same factors (expressed by a factor loadings matrix) are obviously dependent on each other.

Additionally, techniques purely intended for visualisation can be used to evaluate the feature matrix. For example, Sammon mapping can be used to visualise high-dimensional vectors on a 2D plane[7]. Unfortunately, if no statistical data reduction is involved, it is not always clear why a particular mapping was computed. Therefore, visualisation techniques are not used in the experiments below.

4.2 Common data basis and reference data

Principally, any media collection can be used for statistical analysis. The definition of a ground truth is not required. Still, it helps the interpretation process if the used media collections have an inherent context. For example, it is a good idea to give outdoor photos in one collection and not to mix them with, for example, trademark images. For the experiments shown below, we are using three media collections: the Brodatz dataset[4] with 112 greyscale images, a subset of 266 images of the Corel dataset[20] (photos) and a set of 446 coats-of-arms images[3] (artificial colour images with few colour shades). Figure 2 shows samples. With these collections it is possible to evaluate all types of image features. Video features are not considered because, currently, no free stable implementation of the MPEG-7 motion descriptors (needed as reference) does exist.

Of course, other collections do exist that could be used as well: the Benchathlon network holds a large number of image collections[1], TREC maintains a large archive of video information[17] and several collections are available for computer vision purposes that can be used for visual retrieval as well[6]. Still, we prefer to use the proposed media collections because they are carefully selected and for statistical evaluations only a small number of media objects can be used (principal component analysis for a data matrix of 350 elements by 446 cases is a computationally expensive task!).

The visual MPEG-7 features are applied to the predefined media collections to obtain the reference data. We are using all colour descriptors[15] (Color Layout, Color Structure, Dominant Color, Scalable Color), two texture descriptors[15] (Edge Histogram, Homogeneous Texture) and one shape descriptor[2] (Region-based Shape). The remaining texture descriptor (Texture Browsing) is not used, because no stable implementation exists and the elements of the output feature vector do not measure on interval scale (prerequisite of most statistical operations used). Contour-based Shape is not used because it does not generate feature vectors of fixed length.

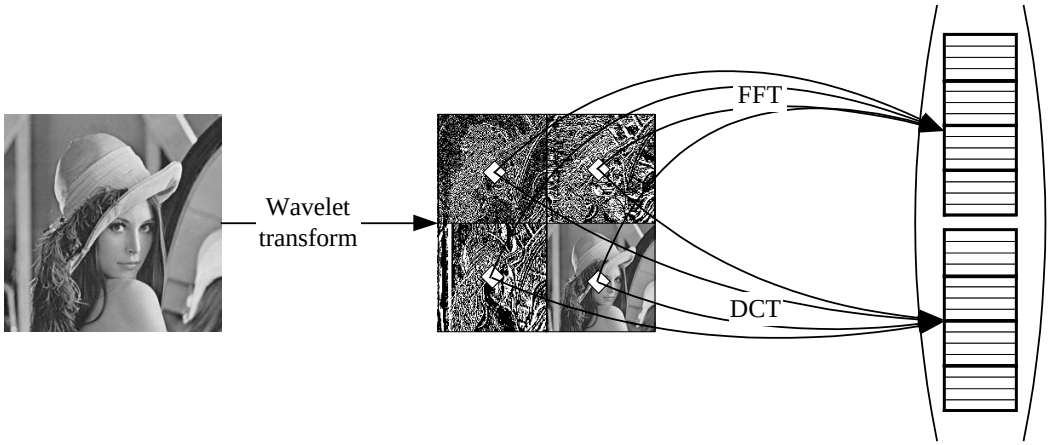

Figure 3: Yet another texture feature extraction process.

4.3 Technical remarks

For the extraction of the MPEG-7 descriptors we made use of the MPEG-7 experimental model[19] (XM) in version 5.6 (provided by the TU Munich). The same parameter settings were used as in earlier experiments[9]. Statistical analysis was performed with standard software like SPSS and Matlab. For the calculation of Self-Organizing Maps the Matlab SOM toolbox[11] was used because it is significantly faster than the standalone version and is shipped with more visualisation tools. All other image and data manipulations were performed in Matlab.

5. EXAMPLES

Below, three examples are given for evaluation of features by statistical methods. Since extraction of description parameters is straightforward we will mainly concentrate on detection of similarities and dependencies. Where cluster analysis is used, we will focus on two-dimensional methods because they are easier to visualise (especially, on paper). In Subsection 5.1 a new descriptor is evaluated and in subsections 5.2 and 5.3 the robustness of existing MPEG-7 descriptors is evaluated.

5.1 Example 1: Evaluating a new descriptor

For the example we defined a straightforward texture feature for images (called Yet Another Texture Feature, YATF) that works on the four (greyscale) sub-images computed by one iteration of a 2D wavelet transformation (using a Daubechies mother wavelet). For each sub-image the first four coefficients of 2D Fourier and cosine transformations are taken as feature elements. Therefore, the resulting feature has 32 elements. Figure 3 shows the extraction process. Using the statistical evaluation procedure proposed above, we would like to explore whether this feature makes sense (e.g. if it is independent of existing MPEG-7 texture features).

Firstly, we apply a principal component analysis (PCA) to find dependencies. All Eigenvalues greater than "1" are considered as factors. The application of a data matrix of YATF and the MPEG-7 reference features for all example images (338 elements, 824 cases) yields to the extraction of 55 factors that explain 78,5% of the totally existing variance. The two first factors alone explain 12% and 10%. Exploring the factor loadings on variables (transformed by Varimax rotation[14]), only a few significant (loading of 90% or higher) dependencies can be identified: the AC DFT coefficients are highly correlated with the corresponding AC DCT coefficients. If YATF elements alone are investigated by PCA, eleven factors are needed to explain 70,5% variance of 32 variables. The first factor explains 8,5%. If only YATF is considered all DFT and corresponding DCT coefficients are highly correlated.

These results are not very surprising. More insights on similarities of YATF to existing features can be found out by cluster analysis. Below, we are always using Self-Organizing Map (SOM) clustering with hexagonal output maps of six by twelve elements. For training a Gaussian neighbourhood kernel[13] is used and clustering quality is measured by quantisation error[13] (average distortion of data vectors to cluster means). All SOMs used below have a quantisation error

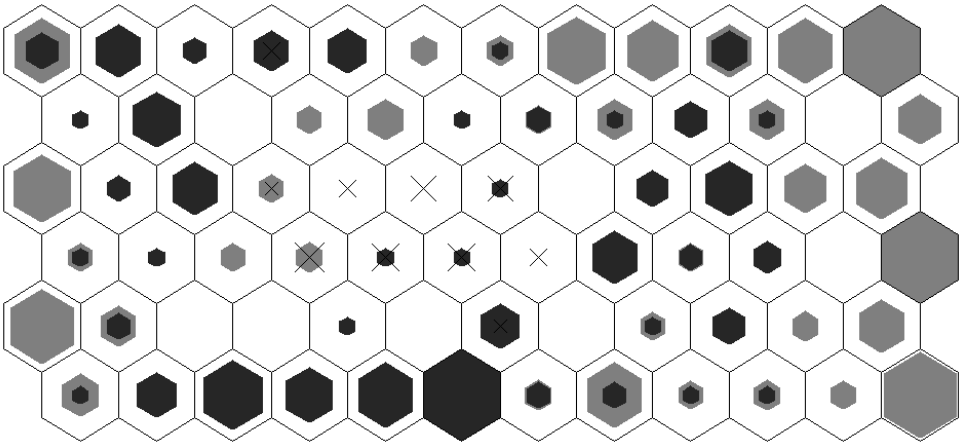

Figure 4: Distribution of YATF ("X"), MPEG-7 colour (light grey), texture and shape feature elements (dark grey).

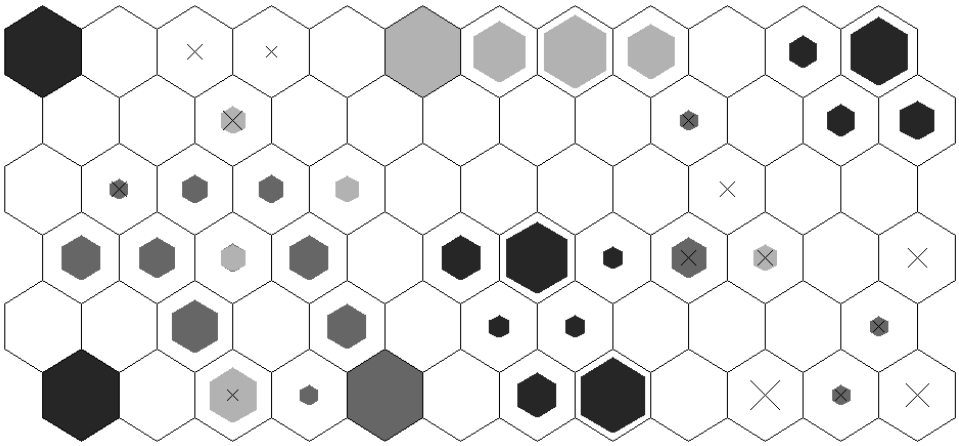

Figure 5: Distribution of YATF ("X"), MPEG-7 texture and shape feature elements on Brodatz dataset. Dark grey: Edge Histogram, medium grey: Homogeneous Texture, light grey: Region-based Shape.

between 0,5 and 1,3 (very small).

Figure 4 shows the general relationship of YATF elements to MPEG-7 features. Clusters with texture and shape features are shown in the same shade, because in earlier work the author could show that Region-based Shape has characteristics similar to a texture feature[9]. The size of markers and coloured hexagons represents the number of elements in a cluster. From the figure it can be seen that YATF does not show significant similarities to colour or texture features. It lies between these two groups showing similarities to a small number of texture as well as colour elements. Additionally, it can be seen that YATF is remarkably self-similar. All elements are clustered in relative proximity.

Figure 5 shows YATF applied to the Brodatz dataset in comparison to the MPEG-7 texture and shape features. In this ideal case for texture features YATF is still mostly independent of other features. Only a small overlapping with Region-based Shape can be identified. Figure 6 shows YATF and all MPEG-7 features on Corel and coats of arms dataset. Then, YATF shows similarities to Region-based Shape and, to a lesser extent, to the colour features. We can draw the conclusion that for all three tested types of content YATF is highly independent of most MPEG-7 descriptors showing only a minor similarity to the Region-based Shape descriptor.

If we compare YATF to itself (Figure 7) we can see for content with sharp edges (Brodatz, coats of arms dataset) that DFT and DCT elements are highly similar while for content with less strong edges (Corel) such similarities exist only to a smaller extent. These evaluations are just examples. Many more statistical techniques could be applied in the same

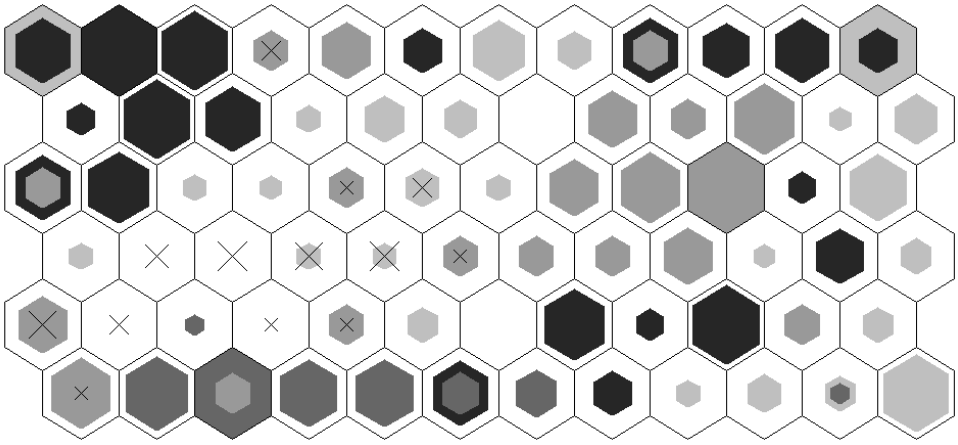

Figure 6: YATF ("X") and MPEG-7 descriptor elements on Corel and coats of arms collection. Dark grey: Edge Histogram, medium grey: Homogeneous Texture, light grey: Region-based shape, very light grey: colour features.

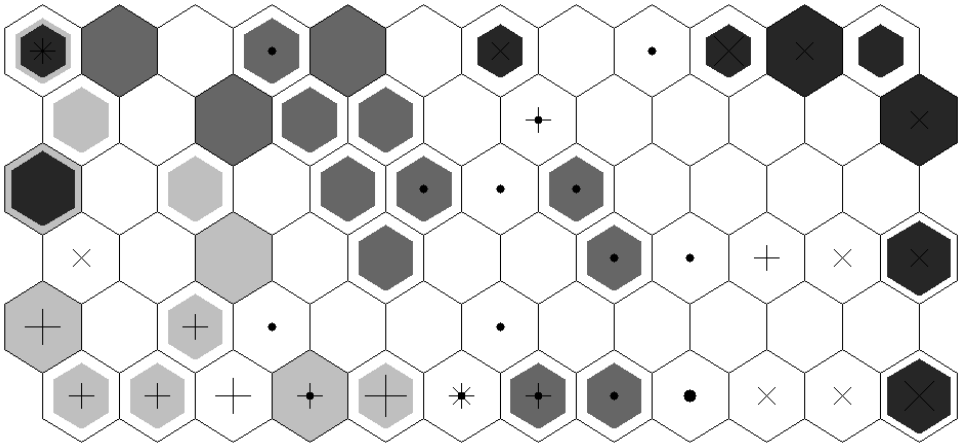

Figure 7: Similarity of YATF Fourier (shades) and Cosine (markers) elements. Dark grey/"X": Brodatz dataset, medium grey/".": Corel dataset, light grey/"+": coats of arms dataset.

manner. Still, we can already see from these evaluations that YATF has qualities that make it a reasonable supplement for the existing MPEG-7 texture and shape descriptors.

5.2 Example 2: Evaluating the robustness of MPEG-7 texture and shape descriptors

In this example we would like to measure robustness by comparing test images to their edge maps. Well-defined texture measures should produce similar feature data for such input. For the experiment the Sobel operator and local thresholding are used to construct edge maps (see Figure 8 for an example). Since it is difficult to identify "hard" dependencies we will again concentrate on identifying similarities from topological cluster structures.

Figure 9 shows the cluster structure for the Brodatz and coats of arms dataset (few long edges). Markers and corresponding colour shades should overlap if the features were able to identify image and edge map pairs. Since this is not the case, the MPEG-7 texture and shape features are apparently not robust against (non-)existence of fine-grained structures in the texture. Figure 10 shows the same descriptors for the Corel dataset (many short edges). Again, the MPEG-7 features describe differently for images and their edge maps. In conclusion, MPEG-7 texture and shape features are highly dependent on the existence of fine-grained texture structures. Coarse structures are – even for Edge Histogram – not sufficient.

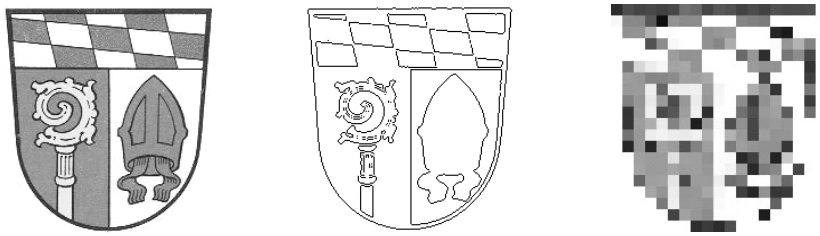

Figure 8: Image transformations. Left: original image, middle: edge map, right: coarse representation.

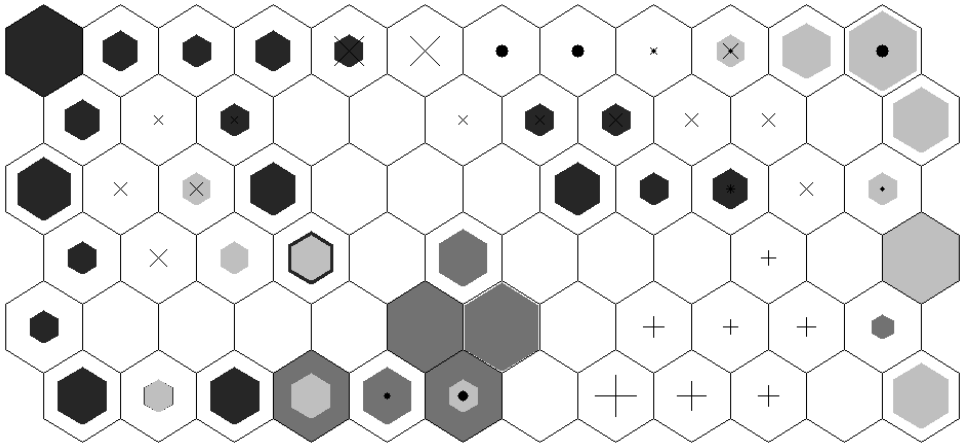

Figure 9: MPEG-7 texture and shape descriptors on original (shades) and edge images (markers; Brodatz and coats of arms dataset). Dark grey/"X": Edge Histogram, medium grey/"+": Homogeneous Texture, light grey/".": Region-based Shape.

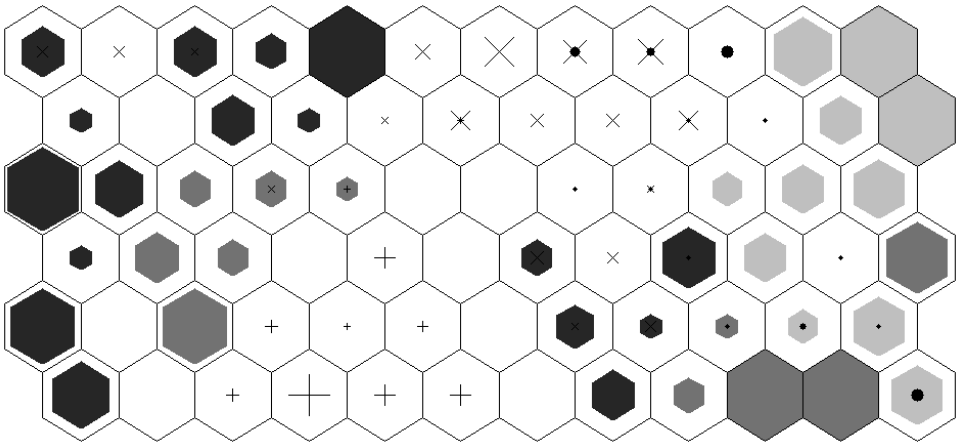

Figure 10: MPEG-7 texture and shape descriptors on original (shades) and edge images (markers; Corel dataset). Dark grey/"X": Edge Histogram, medium grey/"+": Homogeneous Texture, light grey/".": Region-based Shape.

5.3 Example 3: Evaluating sensitivity of visual MPEG-7 descriptors on rotation and coarse representation

In the last example we would like to test the robustness of MPEG-7 descriptors against rotation and coarse representation (computed by shrinking images to 10% and reconstructing them to original size by bilinear interpolation). See Figure 8 for an example. Again, we are relying on topological cluster structures. Figure 11 shows the behaviour of texture and shape descriptors. We can see that Edge Histogram and Region-based Shape are rotation-invariant (markers and corresponding shades in the same clusters). Homogeneous Texture is only partially rotation-invariant. Colour features are not shown, because they are absolutely invariant against rotation.

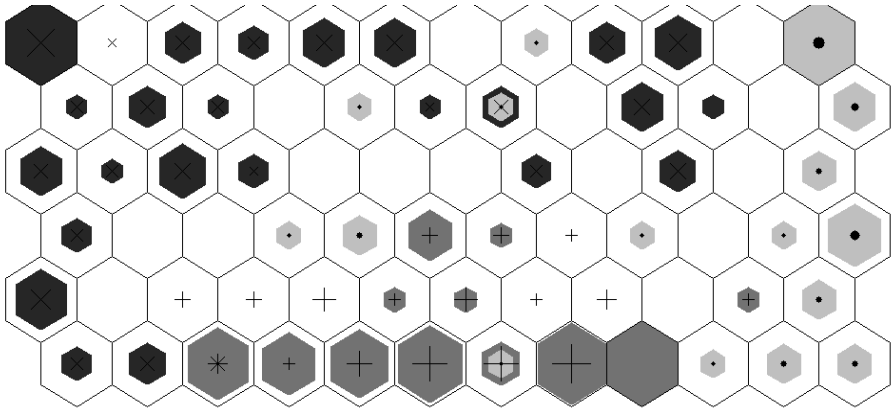

Figure 11: MPEG-7 texture and shape descriptors on original (shades) and rotated images (markers). Dark grey/"X": Edge Histogram, medium grey/"+": Homogeneous Texture, light grey/".": Region-based Shape.

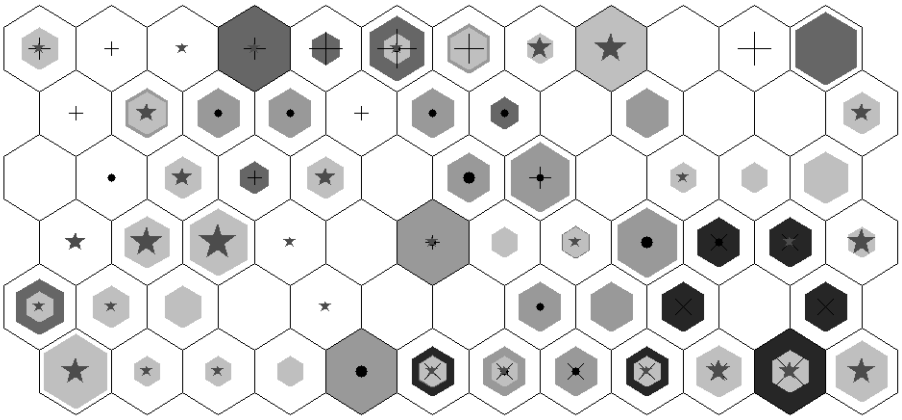

Figure 12: MPEG-7 colour descriptors on original (shades) and coarse images (markers). Dark grey/"X": Color Layout, medium grey/"+": Color Structure, light grey/".": Dominant Color, very light grey/star: Scalable Color.

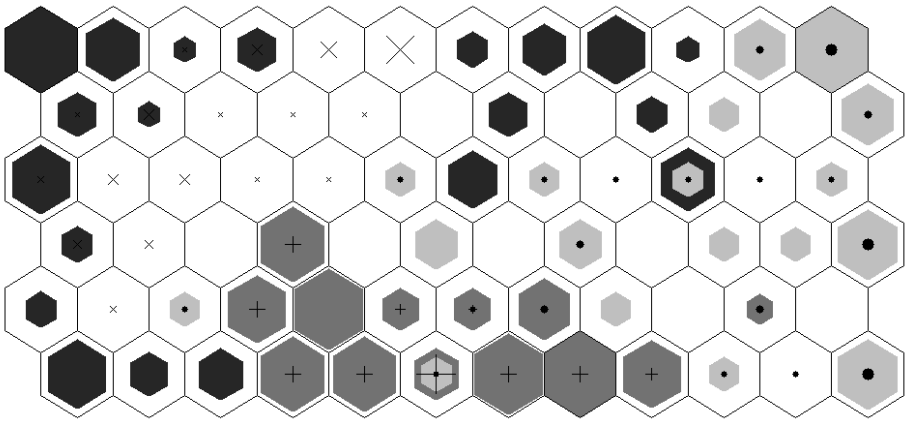

Figure 13: MPEG-7 texture and shape descriptors on original (shades) and coarse images (markers). Dark grey/"X": Edge Histogram, medium grey/"+": Homogeneous Texture, light grey/".": Region-based Shape.

Figure 12 shows the MPEG-7 colour features for coarsely represented images. Dominant Color is absolutely invariant; Color Structure and Scalable Colour are mostly invariant. The most sensible colour feature is Color Layout. Only about 50% of corresponding elements match. Figure 13 shows the texture and shape descriptors for the same datasets. While

Homogeneous Texture is highly invariant against coarseness and Region-based shape contains at least 50% invariant elements, Edge Histogram is highly dependent on the scale of an image. Still, in conclusion, the MPEG-7 features are mostly highly robust against transformation and quality changes.

6. CONCLUSION

In this paper we propose an evaluation procedure for feature transformations in visual information retrieval that is based on statistical data analysis. The essential innovation in comparison to past approaches is the integration of visual MPEG-7 features as reference data. The process is divided into four steps. Professional software tools exist that support the execution of each step. The reasonability of the approach is shown in three examples in Section 5.

The proposed evaluation method has several advantages. Statistical analysis of feature transformations is a lightweight process in comparison to evaluation based on ground truth and test queries. Additionally, statistical results allow conclusions on *how* feature data is extracted and can be used to refine feature transformations. Finally, using MPEG-7 as reference it is possible to determine the characteristics (type) of a feature transformation statistically.

ACKNOWLEDGEMENTS

The author would like to thank Christian Breiteneder for his valuable comments and suggestions for improvement. This work is part of the VizIR research project[8] supported by the Austrian Scientific Research Fund (FWF) under grant no. P16111-N05.

REFERENCES

1. Benchathlon network website, http://www.benchathlon.net/ (last visited 2003-09-24).

2. M. Bober, "MPEG-7 Visual shape descriptors", *Special Issue on MPEG-7, IEEE Transactions on Circuits and Systems for Video Technology*, **11/6**, 716-719, 2001.

3. C. Breiteneder, H. Eidenberger, "Content-based Image Retrieval of Coats of Arms", *Proceedings IEEE International Workshop on Multimedia Signal Processing*, 91-96, IEEE, Helsingör, 1999.

4. A. Del Bimbo, *Visual Information Retrieval*, Morgan Kaufmann Publishers, San Francisco, 1999.

5. S.F. Chang, T. Sikora, A. Puri, "Overview of the MPEG-7 standard", *Special Issue on MPEG-7, IEEE Transactions on Circuits and Systems for Video Technology*, **11/6**, 688-695, 2001.

6. Computer vision image library, http://www-2.cs.cmu.edu/~cil/v-images.html (last visited 2003-09-24).

7. J.D. Edwards, K.J. Riley, J.P. Eakins, "A technique for mapping irregular sized vectors applied to image collections", *Proceedings SPIE Visual Communications and Image Processing Conference*, vol. 5150, 467-475, SPIE, Lugano, 2003.

8. H. Eidenberger, C. Breiteneder, "A Framework for Visual Information Retrieval", *Proceedings Visual Information Systems Conference*, 105-116, Springer Verlag, HSinChu, 2002.

9. H. Eidenberger, "How good are the visual MPEG-7 features?", *Proceedings SPIE Visual Communications and Image Processing Conference*, vol. 5150, 476-488, SPIE, Lugano, 2003.

10. N. Fuhr, "Information Retrieval Methods for Multimedia Objects", *State-of-the-Art in Content-Based Image and Video Retrieval*, R.C. Veltkamp, H. Burkhardt, H.P. Kriegel, 191-212, Kluwer, Boston, 2001.

11. Helsinki University of Technology, SOM toolbox for Matlab website, http://www.cis.hut.fi/projects/somtoolbox/ (last visited: 2003-09-24).

12. A.K. Jain, M.N. Murty, P.J. Flynn, "Data clustering: a review", *ACM Computing Surveys*, **31/3**, 264-323, 1999.

13. T. Kohonen, "The Self-Organizing Map", *Proceedings of IEEE*, **78/9**, 1464-1480, 1990.

14. J.C. Loehlin, *Latent variable models: An introduction to factor, path, and structural analysis* (3rd edition), Lawrence Erlbaum Assoc, Mahwah NJ, 2001.

15. B.S. Manjunath, J.R. Ohm, V.V. Vasudevan, A. Yamada, "Color and texture descriptors", *Special Issue on MPEG-7, IEEE Transactions on Circuits and Systems for Video Technology*, **11/6**, 703-715, 2001.

16. SCHEMA EU project website, Delivery on visual information retrieval techniques, available from http://www.iti.gr/SCHEMA/preview.html?file_id=67/ (last visited 2003-09-24).

17. A.F. Smeaton, P. Over, "The TREC-2002 video track report", *NIST Special Publication*, SP 500-251, available from http://trec.nist.gov/pubs/trec11/papers/ VIDEO.OVER.pdf (last visited: 2003-09-24), 2003.

18. A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, "Content-based image retrieval at the end of the early years", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22/12**, 1349-1380, 2000.

19. TU Munich, MPEG-7 experimentation model website, http://www.lis.e-technik.tu-muenchen.de/research/bv/topics/mmdb/e_mpeg7.html (last visited: 2003-09-24).

20. University of California Berkeley, Corel dataset website, http://elib.cs.berkeley.edu/photos/corel/ (last visited 2003-09-24).