Enriching Texture Analysis with Semantic Data

Tim Matthews, Mark S. Nixon and Mahesan Niranjan
Communications, Signal Processing and Control Group
School of Electronics and Computer Science
University of Southampton
{tm1e10,msn,mn}@soton.ac.uk

Abstract

We argue for the importance of explicit semantic modelling in human-centred texture analysis tasks such as retrieval, annotation, synthesis, and zero-shot learning.

To this end, low-level attributes are selected and used to define a semantic space for texture. 319 texture classes varying in illumination and rotation are positioned within this semantic space using a pairwise relative comparison procedure. Low-level visual features used by existing texture descriptors are then assessed in terms of their correspondence to the semantic space. Textures with strong presence of attributes connoting randomness and complexity are shown to be poorly modelled by existing descriptors.

In a retrieval experiment semantic descriptors are shown to outperform visual descriptors. Semantic modelling of texture is thus shown to provide considerable value in both feature selection and in analysis tasks.

1. Introduction

Visual texture is an important cue in numerous processes of human cognition. It is known to be used in the separation of 'figure' from 'ground' [13], as a prompt in object recognition [22], to infer shape and pose [5], as well as in many other aspects of scene understanding. Over eons of human existence this importance has led to the development of a rich lexicon suitable for concise description of texture. We may speak of *fractured* earth, or of a *rippling* lake, and in doing so are able to convey considerable information about the surface and appearance of these objects.

Although computational texture analysis has achieved fine results over recent decades, there still remains a disparity between the visual and semantic spaces of texture – the so-called *semantic gap*. Computational approaches usually operate on the basis of *a priori* notions of texture not necessarily tied to human experience. This means they are often unsuitable for applications requiring closer or more intuitive human interaction, such as content-based image retrieval, texture synthesis and description, or zero-shot learning, where a classification system is taught new categories without having to observe them.

Our work seeks to bridge this semantic gap for texture, and acts to unify separate research efforts into structuring the semantic [1] and visual [6, 8, 23] texture spaces, and into robustly identifying correspondences between semantic and visual data [21]. Separate semantic modelling has been shown to improve retrieval of natural scenes [29] and gait signatures [26], and indoor-outdoor classification of photographs [27]. In this paper we outline a semantic modelling of texture, allowing it to be described, synthesised, and retrieved using fine-grained high-level semantic constructs rather than solely using low-level visual properties. As well as the clear benefits this has for human-computer interaction, the semantic data collected is a rich source of information in its own right. Humans are capable of analysing texture in a way resistant to noise, and invariant to illumination, rotation, and scale [10, 30]. It is of tremendous advantage to learn – either from human-provided labels, or from investigation of the underlying biological mechanisms – methods of image analysis that are similarly robust.

Texture in particular provides interesting challenges of its own in part because it has historically proven so difficult to define. We sidestep this thorny issue by adopting a subjective definition of texture embedded in human experience. Because our task involves tying some visual texture space to a semantic space borne from human interpretation of that visual space, it is fitting to adopt a definition of texture derived from human perception. In this sense texture is anything describable by constructions from our semantic space and emerges as a natural consequence of our eventual definition of that space. This semantic characterisation of texture allows us to address the problem of feature selection in a principled way, due to hundreds of thousands of years of embodiment within a world of diverse and abundant texture resulting in an evolved language which provides a natural balancing between expressiveness and efficiency.

The main contributions of this paper are:

- A publicly-available[1] labelling of over 300 classes of texture from the Outex dataset. The design and format of this labelling is described in Section 2.

- An assessment of how well a selection of visual texture descriptors are able to capture this semantic data. We show how textures may be ranked according to both their semantic labels and their visual features in Section 3, and then describe the experiment used to compare these rankings in Section 4.

- A demonstration of the benefits of explicit semantic modelling for texture retrieval, given in Section 5.

We finish with discussion of our results in Section 6.

2. Semantic space of texture

We choose to construct our semantic space using *attributes*, low-level visual qualities – often adjectives – shared across objects. In this section we create an expressive but efficient lexicon of attributes to make up our semantic space, select a dataset of texture from which we will derive our visual space, and finally obtain 'ground truth' labellings from human subjects so that we may discover correspondences between these two spaces.

Attributes have received much attention in recent literature in computer vision, particularly within object recognition. Their use permits a shift in perspective from the traditional approach of object recognition in which object classes themselves are atomic units of recognition. They allow the association of visual data with shared low-level qualities, facilitating efficient class-level learning and generalisation [3, 14], and they provide a means for intuitive and fine-grained description, such as when describing unusual features of an object, or in stating the ways in which one object is similar to another. Farhadi *et al*. [4] state that attributes allow us to "*shift the goal of recognition from naming to describing*".

Attributes have been found to be particularly appropriate for domains in which key features exist along continua, such as in biometrics [12, 24, 26] and scene classification [20, 25]. We see texture as being ill-suited to strict categorisation: key properties in which texture has been stated to vary include its *coarseness*, *linearity*, and *regularity* [15, 28], all of which may be expected to vary continuously. Texture is particularly suitable for description with attributes as they may be readily sourced from the rich lexicon that has evolved in order to describe it.

Numerous elegant insights into the nature of the English-language texture lexicon were made by Bhushan *et al*. [1], who asked subjects to cluster 98 texture adjectives according to similarity, without access to visual data. From the responses, they were able to make two important insights into the structure of the semantic space of texture:

Cluster interpretation	Sample words
Linear orientation	*furrowed*, *lined*, *pleated*
Circular orientation	*coiled*, *flowing*, *spiralled*
Woven structure	*knitted*, *meshed*, *woven*
Well-ordered	*regular*, *repetitive*, *uniform*
Disordered	*jumbled*, *random*, *scrambled*
Disordered linear primitives	*cracked*, *crinkled*, *wrinkled*
Disordered circular primitives	*dotted*, *speckled*, *spotted*
Disordered circ. 3D primitives	*bubbly*, *bumpy*, *pitted*
Disordered woven structure	*frilly*, *gauzy*, *webbed*
Disordered, circular blurring	*blemished*, *blotchy*, *smudged*
Disordered, linear blurring	*marbled*, *scaly*, *veiny*

Table 1: Interpretations of the eleven texture word clusters identified in [1].

- A principal components analysis revealed that just three dimensions were sufficient to account for 82% of the variability in the similarity responses. Examination of these three dimensions revealed them to correspond approximately to notions of *linearity* (linear vs. circular texture orientation), *repetition* (disordered vs. structured texture), and *complexity* (simple vs. intricate texture).

- Hierarchical clustering of the data identified eleven major clusters of texture adjectives. These clusters are shown in Table 1, along with interpretations. To illustrate the spread of these clusters within the 3D space described above, a representative word from each cluster is plotted in Figure 1.

By selecting a single word from each of the clusters identified by Bhushan *et al*. we are able to create a new attribute lexicon of manageable size which adequately covers the semantic space of texture. The words chosen are the same as those displayed in Figure 1: `blemished`, `bumpy`, `lined`, `marbled`, `random`, `repetitive`, `speckled`, `spiralled`, `webbed`, `woven`, and `wrinkled`. These words were chosen above others from the same cluster for their frequency of use and perceived generality, so as to aid subject understanding when obtaining labels.

2.1. Dataset

The Outex dataset [18] was adopted for our analysis. We use the 319 texture classes included in test suite `Outex_TC_00016` captured with three different illuminants (`horizon`, `inca`, `tl84`), and at four different rotation angles (0°, 30°, 60°, 90°), giving twelve samples per texture class, and a total of 3,828 samples all together.

[1] `www.ecs.soton.ac.uk/~tm1e10/texture.html`

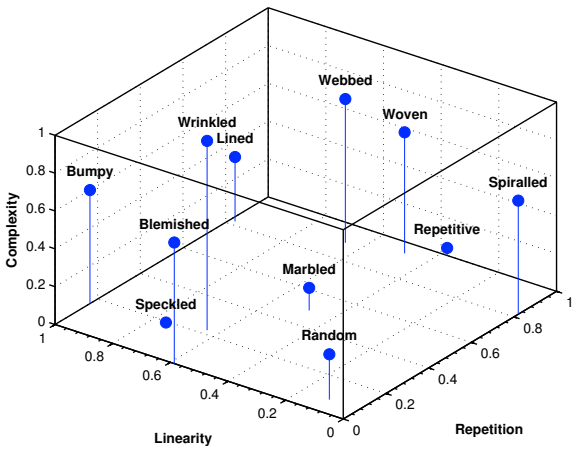


Figure 1: Representative attributes from each of the clusters in Table 1, with approximate locations (normalised between 0 and 1) across the three texture dimensions identified in [1].

Colour is taken to be a separate visual cue, and so all texture samples are converted to grayscale before experimentation.

2.2. Obtaining labels

Human-provided labels are required for each of the eleven attributes selected so that we have some means of assessing algorithmic performance. It is clear that normal binary classification is insufficient for the attributes at hand, as they correspond specifically to the *degree* of expression of visual features rather than to simply their presence or absence. We therefore require some labelling mechanism allowing the Outex textures to be placed along a continuum according to how strongly they evince each attribute. This can be done by having subjects directly rate the perceived strength of attributes within each texture along a bounded rating scale [23]. However, this is unintuitive when the assumption of an underlying bounded continuum is inappropriate: it is not obvious, for example, what form a *maximally* marbled texture would take.

This issue may be overcome within the framework of *pairwise comparison*, a psychometric procedure in which a subject is shown two stimuli simultaneously and prompted to choose the stimulus exhibiting more of some quality. In this way we can work with so-called *relative attributes*, which have been shown to outperform ordinary categorical attributes in retrieval tasks [21, 24] and to provide a more intuitive and natural user experience [11]. This technique has previously been used specifically for texture by Tamura *et al*. [28] and it has been hypothesised that the textural processes of the human cognitive system operate using such a comparison mechanism [8].

Our methodology is as follows: a subject is shown two textures side-by-side along with a single attribute. The subject is prompted to select the texture that exhibits a greater

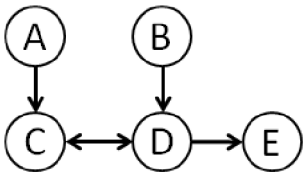


Figure 2: Comparison graph for an attribute, where a directed edge represents a dominance relation, and a double-directed edge represents a similarity relation. `A` and `B` would be the next comparison pair presented to subjects as no directed path exists between them, and so it is not possible to infer which one is dominant for the attribute.

level of expression of the attribute in question, or to rate them as similar. Subjects are also given the option of stating that the attribute is completely absent from both textures, so as to avoid confusion when textures are shown for which a particular attribute is perceived to not apply. Absence comparisons are equivalent to a similarity judgement in all subsequent analysis. The attribute shown is the one involved in the fewest comparisons at that point. Representing each attribute's comparisons in terms of a directed graph (where vertices are textures and edges are comparisons — see Figure 2), textures are selected by randomly choosing texture pairs with no path between them within that attribute's comparison graph. This is to increase the probability of useful, discriminative comparisons being generated so that fewer total comparisons are needed to infer a complete ordering.

Initially, 7,689 comparisons were obtained from ten subjects unfamiliar with the work being performed, along with the paper authors. This is only a tiny proportion of the $\binom{3828}{2}$ comparisons possible. We increase the coverage by assuming comparisons to apply equally to all 12 samples within each texture class, owing to the natural human visual robustness to illumination and rotation when describing surface texture. Because of this assumption only textures with rotation of $0°$ and illumination of `horizon` are displayed to users. After duplicating each comparison for all 144 possible combinations of between-class samples we have 1,107,216 comparisons. At around 100,000 comparisons per attribute, this is still far fewer than the complete $\binom{3828}{2}$ case, but it has been indicated [21] that relatively few comparisons – in the order of the number of items being compared – are needed to achieve results comparable to that of the complete case. In the next section we describe how these comparisons are used to infer rankings.

3. Ranking textures

In order to bridge the semantic gap we require a way of measuring the level of expression of each attribute within a texture, a quality we will refer to as an attribute's *strength*. This section addresses how we may obtain this measure of

perceptual strength both directly from the comparison graph and from low-level visual data. In what follows, $\mathcal{D}$ is a set of dominance comparisons where each ordered pair $(a, b) \in \mathcal{D}$ indicates that a subject considered image a to possess a higher strength of some attribute than image b, and $\mathcal{S}$ is a set of similarity comparisons where each pair $(a, b) \in \mathcal{S}$ indicates that a subject considered images a and b to possess similar levels of some attribute.

3.1. Ranking from visual features

When a new texture is provided we may wish to determine the strength of an attribute within it based only on its visual features. To do this we derive a *ranking function* capable of mapping a visual descriptor to real value measures of attribute strength. Using $\boldsymbol{w}$ to represent the coefficients of a linear ranking function for some attribute, and $\boldsymbol{x}_i$ to represent the location in feature space of the i^{th} texture in the dataset, the perceived strength of the attribute within that texture can be given as $\boldsymbol{w} \cdot \boldsymbol{x}_i$. A soft-margin Ranking SVM [9] is used to derive $\boldsymbol{w}$ where ξ_{ij} is the misclassification error between items i and j, and C is the trade-off between maximising the margin and minimising the misclassification error. Support for similarity constraints is gained using the formulation of [21]:

$$\begin{aligned} \underset{\boldsymbol{w}}{\text{minimise}} \quad & \frac{1}{2}||\boldsymbol{w}||^2 + C\sum \xi_{ij}^2 \\ \text{subject to} \quad & \boldsymbol{w} \cdot (\boldsymbol{x}_i - \boldsymbol{x}_j) \geq 1 - \xi_{ij},\ (i,j) \in \mathcal{D} \\ & \boldsymbol{w} \cdot (|\boldsymbol{x}_i - \boldsymbol{x}_j|) \leq \xi_{ij}, (i,j) \in \mathcal{S} \\ & \xi_{ij} \geq 0 \end{aligned} \tag{1}$$

This is similar to a traditional soft-margin SVM in which the data being separated are differences between feature vectors, as opposed to the vectors themselves.

3.2. Ranking from comparison graphs

It is desirable to plot the textures along a continuum reflecting as fully as possible the orderings obtained through the pairwise comparison methodology. This can be done using the comparison graph directly, and may act as a 'ground truth' measure of each attribute's perceived strength within each texture. Taking $\boldsymbol{r}$ to be a vector of numeric ratings for each of the n textures, we wish to find some $\boldsymbol{r}$ such that for each dominance relation $(a, b) \in \mathcal{D}, r_a > r_b$, and for each similarity relation $(a, b) \in \mathcal{S}, |r_a - r_b| = 0$. By introducing error variables to allow for cases when these constraints cannot be satisfied, we can in fact express these goals in a form very similar to Equation 1:

$$\begin{aligned} \underset{\boldsymbol{r}}{\text{minimise}} \quad & \frac{1}{2}||\boldsymbol{r}||^2 + C\sum \xi_{ij}^2 \\ \text{subject to} \quad & r_i - r_j \geq 1 - \xi_{ij},\ (i,j) \in \mathcal{D} \\ & |r_i - r_j| \leq \xi_{ij}, (i,j) \in \mathcal{S} \\ & \xi_{ij} \geq 0 \end{aligned} \tag{2}$$

Equation 1 is equivalent to the above when $\boldsymbol{x}$ is an $n \times n$ identity matrix. At this point, we illustrate the attributes chosen in the previous section by displaying in Figure 3 the highest-rating texture in $\boldsymbol{r}$ for each attribute when $C = 1$.

In the next section we show how the rankings produced using these two different methods can be compared in order to ascertain which low-level features best reflect the high-level semantic attributes at our disposal.

4. Semantic correspondence of visual descriptors

In this section we appraise each of a number of existing texture descriptors in terms of how well they reflect the structure of the semantic comparison graph for each of the eleven attributes. These results allow us to identify regions within the semantic space of texture which are poorly modelled by current techniques, as well as the visual features which correspond best with human perception and which will provide the basis for our semantically-enriched descriptors.

4.1. Visual descriptors

The five different texture descriptors to undergo assessment are:

- Co-occurrence matrices [7] are calculated for points situated along the perimeters of circles of radii 1, 2, 4, 8, and 16 pixels. Each of these five matrices are summarised in terms of their contrast, homogeneity, uniformity, entropy, variance, and correlation, resulting in a 30-element feature vector. (`CoM`)

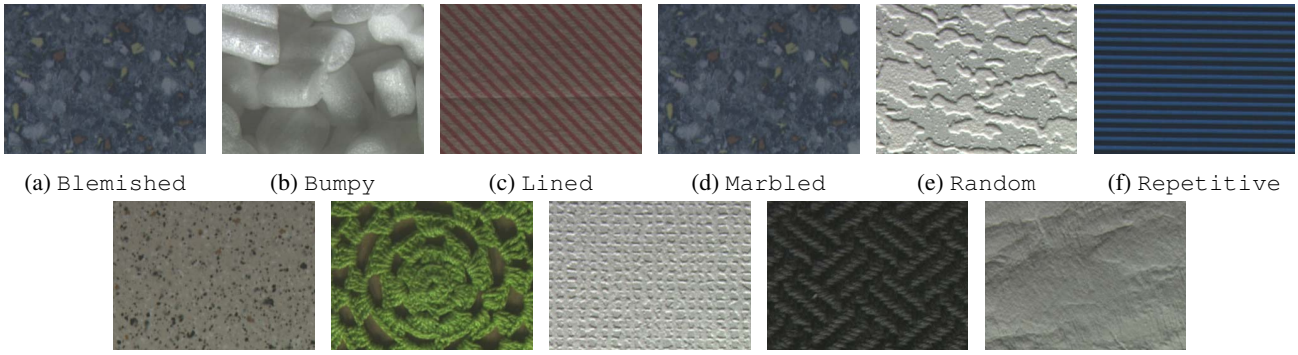
- The mean and standard deviation of the Gabor wavelet responses of 24 orientation and scale combinations given in [17], yielding a feature vector of 48 elements. (`Gab`)

- The 8 optimal Liu noise-resistant features of the Fourier transform [16]. (`Liu`)

- 16-dimensional feature vector derived from the Statistical Geometrical Features procedure [2] being performed at 31 regularly-spaced threshold levels. (`SGF`)

- Uniform (local) binary patterns [19] calculated for eight points around circles of three different radii: 2, 4, and 8. This gives a total concatenated feature vector of dimension 30. (`UBP`)

In the next section we describe the methodology used to assess these five descriptors.



(a) Blemished (b) Bumpy (c) Lined (d) Marbled (e) Random (f) Repetitive

(g) Speckled (h) Spiralled (i) Webbed (j) Woven (k) Wrinkled

Figure 3: Illustrative Outex textures for each of the eleven attributes. The texture shown is that with the highest value in the ratings calculated directly from the comparison graph for each attribute using Equation 2.

4.2. Methodology

The semantic correspondence of each of the visual descriptors described above is evaluated using a 4-fold cross-validation procedure. For each attribute the optimal ranking function $\boldsymbol{w}$ is learned from the training images using the Ranking SVM formulation shown in Equation 1. The free parameter in the Ranking SVM equation, C, is allowed to vary between 21 logarithmically spaced values $(4^{-10}, 4^{-9}, \ldots, 4^{9}, 4^{10})$, the optimal value of which is selected through the cross-validation procedure.

A per-attribute ranking of all 3,828 textures is then derived from $\boldsymbol{w}$. The misclassification rate is calculated over all dominance comparisons involving at least one of the textures in the hold-out set by simply comparing the rankings of the respective textures.

We also measure the correspondence between each learned ranking and the 'ideal' ranking inferred directly from the semantic comparison graph with the procedure in Section 3.2. The Spearman's rank correlation coefficient, $-1 \leq \rho \leq 1$, is calculated for this purpose, where $\rho = 1$ indicates a perfect monotonic relationship between the two rankings (that is, the visual features) and $\rho = -1$ indicates a perfect *inverse* monotonic relationship.

4.3. Analysis

Misclassification rates and Spearman's rank correlation coefficients for each combination of descriptor and attribute are shown in Tables 2 and 3.

It is apparent that the uniform binary patterns descriptor is the most suitable of those tested for capturing the structure of the semantic comparison graph. In particular, it performs well for those attributes relating to disordered placement of small-scale primitives – `blemished`, `bumpy`, and `speckled` – as well as for another attribute associated with disorder, `marbled`. The uniform binary pattern descriptor is calculated as a histogram of local intensity patterns and so it is unsurprising that it performs relatively well for spatially-localised primitives such as speckles and bumps. By its nature the histogram makes no regard

Attribute	CoM	Gab	Liu	SGF	UBP
Blemished	0.28	0.29	0.31	0.34	**0.27**
Bumpy	0.26	0.34	0.35	0.29	**0.24**
Lined	0.33	0.36	**0.22**	0.34	0.24
Marbled	0.23	0.25	0.24	0.31	**0.17**
Random	0.30	0.31	**0.26**	0.33	0.27
Repetitive	0.29	0.35	0.29	0.33	**0.28**
Speckled	0.27	0.30	0.25	0.26	**0.21**
Spiralled	**0.18**	0.27	0.36	0.24	0.22
Webbed	0.24	0.29	0.28	0.28	**0.21**
Woven	0.23	0.26	**0.20**	0.23	0.22
Wrinkled	**0.28**	0.33	0.38	0.36	0.33

Table 2: Misclassification rates for each combination of descriptor and attribute. Boldface denotes the strongest scoring descriptor for an attribute.

Attribute	CoM	Gab	Liu	SGF	UBP
Blemished	0.39	0.34	0.38	0.30	**0.47**
Bumpy	0.43	0.37	0.29	0.40	**0.46**
Lined	0.25	0.25	**0.53**	0.26	0.47
Marbled	0.40	0.41	0.39	0.32	**0.48**
Random	0.57	0.57	**0.70**	0.52	0.67
Repetitive	0.37	0.28	0.39	0.35	**0.42**
Speckled	0.35	0.35	0.40	0.38	**0.44**
Spiralled	**0.26**	0.19	0.15	0.24	0.23
Webbed	**0.26**	0.18	0.24	0.23	0.21
Woven	0.30	0.21	**0.34**	0.20	0.28
Wrinkled	**0.36**	0.31	0.24	0.30	0.31

Table 3: Rank correlation coefficients for each combination of descriptor and attribute. Boldface denotes the strongest scoring descriptor for an attribute.

for placement rules making it better suited to capturing aspects of disorder. However, its structure is not amenable to deeper understanding as the local intensity patterns it detects have no immediately intuitive definition.

Another descriptor based upon small-scale intensity patterns is that comprising statistics of the co-occurrence matrix. The misclassification rate for the `spiralled` attribute was unexpectedly low, at just 0.18. Inspection of $\boldsymbol{w}$ reveals that this is due to variations in the co-occurrence matrix energy at different radii and, to a lesser extent, the correlation. This suggests that the curved lines of the structures perceived by subjects as being `spiralled` are of similar scale to the large shifts in pixel uniformity occurring between 8 and 16 pixels from each reference pixel. A similar effect is observed for `lined` and `woven`, but due to these having associations of strong global texture orientation, and the co-occurrence matrix being based on spatially-localised patterns, they did not result in similarly low misclassification rates scores as for `spiralled`.

The Liu descriptor – comprising frequency measures based on the Fourier transform – performs well for the two attributes involving regular placement of linear texture primitives: `lined` and `woven`. Inspection of $\boldsymbol{w}$ reveals that a high moment of inertia and low proportion of energy for the first quadrant of the normalised Fourier transform are the pertinent features for these two attributes. The Liu descriptor is also amongst the best performers for the polar notions of `random` and `repetitive`: here, the inertia and energy of the first quadrant is again decisive. Low inertia and high energy indicates `random` texture while the opposite aligns more closely with `repetitive` texture.

The two remaining descriptors, Gabor and statistical geometrical features, achieved good correspondence for certain attributes, but were invariably eclipsed by one of the other three descriptors.

Overall, the results indicate that there is considerable opportunity for improvement in the identification of visual features corresponding closely to human perception, especially for attributes exhibiting aspects of complexity or disorder such as `spiralled`, `webbed`, and `wrinkled`. These deficiencies are hardly unexpected, and tally with our knowledge of the workings of visual texture descriptors, but it is notable too that even correspondence with strongly regular attributes such as `lined` and `repetitive` is only average. Even despite the lack of correspondence between these human and machine interpretations of texture, semantic data may still be used to improve performance in tasks involving texture analysis. In the next section we demonstrate that semantic texture description results in considerable performance gains over a purely visual approach.

5. Retrieval

In this section we demonstrate the practical benefit of semantic data in a retrieval experiment.

5.1. Methodology

Each of the 3,828 samples in the dataset is used in turn as a query texture against the remaining 3,827 textures in the target set, of which only 11 are relevant to the query (each texture class has 12 samples due to variation in rotation and illumination). All textures in the target set are then sorted by the Euclidean distance of their descriptors from the query texture's descriptor yielding a ranking $\boldsymbol{r}$ where $r_i = 1$ if the member of the target set at rank i is relevant to the query, and 0 otherwise. This is done for all five descriptors introduced in Section 4.1.

Next, for each descriptor eleven ranking functions are learned, one for each attribute. The learning process operates only with the textures in the target set and uses the cross-validation procedure described in Section 4.2. These eleven ranking functions are then used to create a new eleven-dimensional *semantic descriptor* for each texture sample. Lastly, we create a concatenated descriptor from all five visual descriptors which is in turn allows another semantic descriptor to be learned from the most discriminative features across all descriptors. Again, the distances between the target set samples and the query sample are calculated for these concatenated and semantic descriptors, and a ranking derived.

From the relevance indicators of the n closest textures $(r_1, \ldots, r_n)$ for each query image we are able to calculate *precision* and *recall* measures, where precision is the proportion of the retrieved samples that are relevant, and recall is the proportion of the relevant samples that are retrieved:

$$\text{precision}(n) = \frac{\sum_{i=1}^{n} r_i}{n} \qquad \text{recall}(n) = \frac{\sum_{i=1}^{n} r_i}{11}$$

Precision and recall are then calculated as n is allowed to vary from 1 to 3,827. We also calculate two summary measures of the ranked data:

- Mean average precision (MAP). The average precision (AP) for a given query is the average of the precision at each rank at which a relevant item is located:

  $$\text{AP} = \frac{\sum_{i=1}^{3827} r_i \, \text{precision}(i)}{11}$$

  This quantity is in turn averaged over all 3,828 queries.

- Equal error rate (EER). denoting the error rate at the point where the true positive rate (the recall) equals the false positive rate. It is equivalent to the point on an ROC curve which intersects the diagonal connecting 100% on the X and Y axes.

Descriptor	MAP		EER	
	Visual	Semantic	Visual	Semantic
CoM	42.6%	**52.3%**	8.4%	**5.9%**
Gab	21.4%	**38.7%**	20.3%	**10.8%**
Liu	18.9%	**24.0%**	22.7%	**12.6%**
SGF	27.6%	**29.2%**	14.1%	**13.2%**
UBP	**76.9%**	61.0%	5.6%	**4.6%**
Concatenated	49.6%	**63.3%**	6.5%	**2.5%**

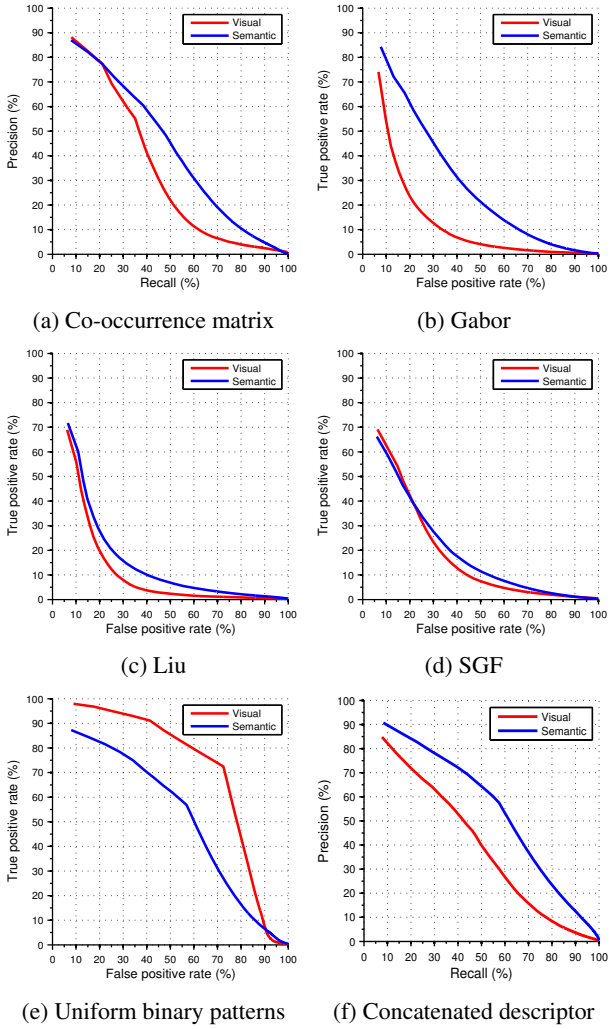
Table 4: Mean average precision (MAP) and equal error rates (EER) for each descriptor across all 3,828 texture queries. Boldface denotes the highest scorer of each visual and semantic descriptor pair.

5.2. Analysis

Precision-recall curves for both the visual and semantic version of each descriptor are shown in Figure 4. MAP and EER scores are shown in Table 4.

In all but one of the curves – that for uniform binary patterns – it is immediately evident that the semantic descriptor gives higher retrieval performance than for the corresponding low-level visual descriptor. This benefit is especially pronounced for higher rates of recall, where the semantic descriptor often retrieves relevant textures with a considerably higher rate of precision. This improved precision at higher recall values could be interpreted as being indicative of greater robustness in the semantic descriptor: whereas the visual descriptors appear to struggle to recall all variations of rotation and illumination for a given query, the semantic descriptor is imbued with the invariant qualities that come from learning from the semantic comparison graph, and so generally is able to better recall variations of the same texture. This initial impression from inspecting the curves is reinforced upon viewing the summary values in Table 4: the semantic descriptor achieves higher MAP and EER scores in all cases but one. However, although the semantic form of the concatenated descriptor is the best overall descriptor in terms of EER, the visual form of the UBP descriptor is the best in terms of MAP.

The inferior MAP of the semantic UBP descriptor against its visual counterpart is possibly due to the relatively compact 11-dimensional semantic representation failing to capture as much variability as the 30-dimensional UBP descriptor. Further experimentation is required to investigate the advantage a more expressive semantic space (in the form of more attributes) has for precision and recall. Again, however, the semantic descriptor improves upon the visual one for higher recall rates. This effect is more obvious in the ROC curve for this descriptor, reflected by the fact the EER for the semantic UBP descriptor is lower.



(a) Co-occurrence matrix (b) Gabor

(c) Liu (d) SGF

(e) Uniform binary patterns (f) Concatenated descriptor

Figure 4: Average precision-recall curves for each descriptor across all 3,828 texture queries. For each query there are 11 relevant and 3,816 irrelevant samples.

6. Discussion

An explicit semantic modelling step provides numerous benefits when describing textures. As well as allowing for more user-friendly interaction due to the bridging of the semantic gap, we demonstrated an improvement in retrieval rate for all but one of the descriptors tested. Furthermore, the use of attributes introduces a natural efficiency and robustness in the design of feature vectors, owing to the evolution of human language and the invariant qualities of human visual perception.

The introduction of the dataset enables new semantic performance metrics to be used when assessing texture descriptors. It is important that the deficiencies encountered in our appraisal of visual descriptors are addressed so as to properly bridge the semantic gap for texture and to pave the

way for closer correspondence to human perception and expectations in user-centred visual applications.

In future work we aim to build on the work of [1] – whose methodology was only performed on the limited Brodatz dataset and without modern facilities such as crowdsourcing – so that a more principled and refined texture lexicon is available to vision researchers. We also aim to further explore the visual space of texture and to describe novel texture features that align particularly closely with human perception.

Acknowledgments

Tim Matthews was supported by a EPSRC Doctoral Training Grant.

References

[1] N. Bhushan, A. R. Rao, and G. L. Lohse. The texture lexicon: Understanding the categorization of visual texture terms and their relationship to texture images. *Cognitive Science*, 21(2):219–246, 1997. 1, 2, 3, 8

[2] Y. Q. Chen, M. S. Nixon, and D. W. Thomas. Statistical geometrical features for texture classification. *Pattern Recognition*, 28(4):537–552, Apr. 1995. 4

[3] A. Farhadi, I. Endres, and D. Hoiem. Attribute-centric recognition for cross-category generalization. In *Proc. IEEE Conf. on CVPR*, pages 2352–2359, 2010. 2

[4] A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth. Describing objects by their attributes. In *Proc. IEEE Conf. on CVPR*, pages 1778–1785, 2009. 2

[5] J. J. Gibson. *The perception of the visual world*, volume xii. Houghton Mifflin, Oxford, England, 1950. 1

[6] R. Gurnsey and D. J. Fleet. Texture space. *Vision Research*, 41(6):745–757, Mar. 2001. 1

[7] R. Haralick. Statistical and structural approaches to texture. *Proc. IEEE*, 67(5):786–804, 1979. 4

[8] L. O. Harvey Jr. and M. J. Gervais. Internal representation of visual texture as the basis for the judgment of similarity. *Journal of Experimental Psychology: Human Perception and Performance*, 7(4):741–753, Aug. 1981. 1, 3

[9] T. Joachims. Optimizing search engines using clickthrough data. In *Proc. 8th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, KDD '02, pages 133–142, 2002. 4

[10] F. A. Kingdom and D. R. Keeble. On the mechanism for scale invariance in orientation-defined textures. *Vision Research*, 39(8):1477–1489, Apr. 1999. 1

[11] A. Kovashka, D. Parikh, and K. Grauman. WhittleSearch: image search with relative attribute feedback. In *Proc. IEEE Conf. on CVPR*, pages 2973–2980, 2012. 3

[12] N. Kumar, A. Berg, P. Belhumeur, and S. Nayar. Attribute and simile classifiers for face verification. In *Proc. IEEE 12th Int. Conf. on Computer Vision*, pages 365–372, 2009. 2

[13] V. A. Lamme. The neurophysiology of figure-ground segregation in primary visual cortex. *Journal of Neuroscience*, 15(2):1605–1615, Jan. 1995. 1

[14] C. H. Lampert, H. Nickisch, and S. Harmeling. Learning to detect unseen object classes by between-class attribute transfer. In *Proc. IEEE Conf. on CVPR*, pages 951–958, June 2009. 2

[15] K. I. Laws. *Textured Image Segmentation*. Ph.D. thesis, University of Southern California, 1980. 2

[16] S.-s. Liu and M. Jernigan. Texture analysis and discrimination in additive noise. *Computer Vision, Graphics, and Image Processing*, 49(1):52–67, Jan. 1990. 4

[17] B. Manjunath and W. Ma. Texture features for browsing and retrieval of image data. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(8):837 –842, Aug. 1996. 4

[18] T. Ojala, T. Maenpaa, M. Pietikainen, J. Viertola, J. Kyllonen, and S. Huovinen. Outex - new framework for empirical evaluation of texture analysis algorithms. In *Proc. 16th Int. Conf. on Pattern Recognition*, volume 1, pages 701–706, 2002. 2

[19] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002. 4

[20] A. Oliva, A. B. Torralba, A. Gurin-Dugu, and J. Hrault. Global semantic classification of scenes using power spectrum templates. In *Proc. Int. Conf. on Challenge of Image Retrieval*, Electronic Workshops in Computing series, 1999. 2

[21] D. Parikh and K. Grauman. Relative attributes. In *IEEE Int. Conf. on Computer Vision*, pages 503–510, 2011. 1, 3, 4

[22] C. J. Price and G. W. Humphreys. The effects of surface detail on object categorization and naming. *The Quarterly Journal of Experimental Psychology*, 41(4):797–828, 1989. 1

[23] A. Rao and G. Lohse. Towards a texture naming system: Identifying relevant dimensions of texture. In *Proc. IEEE Conf. on Visualization*, pages 220–227, 1993. 1, 3

[24] D. A. Reid and M. Nixon. Using comparative human descriptions for soft biometrics. In *Proc. Int. Joint Conference on Biometrics*, pages 1–6, Oct. 2011. 2, 3

[25] B. E. Rogowitz, T. Frese, J. Smith, C. A. Bouman, and E. Kalin. Perceptual image similarity experiments. In *Proc. SPIE Conf. on Human Vision and Electronic Imaging*, pages 576–590, 1998. 2

[26] S. Samangooei, B. Guo, and M. Nixon. The use of semantic human description as a soft biometric. In *Proc. 2nd IEEE Int. Conf. on Biometrics: Theory, Applications and Systems*, pages 1–7, Oct. 2008. 1, 2

[27] N. Serrano, A. E. Savakis, and J. Luo. Improved scene classification using efficient low-level features and semantic cues. *Pattern Recognition*, 37(9):1773–1784, Sept. 2004. 1

[28] H. Tamura, S. Mori, and T. Yamawaki. Textural features corresponding to visual perception. *IEEE Trans. on Systems, Man and Cybernetics*, 8(6):460–473, June 1978. 2, 3

[29] J. Vogel and B. Schiele. Semantic modeling of natural scenes for content-based image retrieval. *International Journal of Computer Vision*, 72(2):133–157, Apr. 2007. 1

[30] J. Zhang and T. Tan. Brief review of invariant texture analysis methods. *Pattern Recognition*, 35(3):735–747, Mar. 2002. 1