Anthropocentric Biocybernetic Computing for Analysing the Architectural Design of House Façades and Cityscapes

Stephan K. Chalup and Michael J. Ostwald
Anthropocentric Biocybernetic Computing for Analysing the Architectural Design of House Façades and Cityscapes
Stephan K. Chalup, The University of Newcastle, NSW, Australia
Michael J. Ostwald, The University of Newcastle, NSW, Australia

Abstract: Research in artificial intelligence and autonomous agents envisions that future robots will accompany humans in their daily lives. The aim is to provide support not only for routine, challenging, or dangerous tasks, but also to improve quality of life through personal assistance and coaching. In order to allow artificial agents to communicate sensibly and to participate in human society, it is important to equip them with the ability to perceive and appreciate aesthetic features of design in a human-like manner. The present study investigates how methods from anthropocentric biocybernetic computing (ABC) can be assembled in an intelligent control module for architectural design evaluation. Central to the system is an abstract model of aesthetic experience, which is established through statistical learning. For the experiments, a database of images of house façades is employed. The learning algorithm extracts line distributions, which characterise façade design, and represents them abstractly in the form of a non-linear manifold. Each point on the manifold corresponds to one façade. The proposed module includes two additional affective perceptual pathways, which are implemented using paradigms that are believed to reflect responses of the human emotional system. One paradigm involves concepts of facial expression recognition, and the other is based on calculating the fractal dimension of the skyline of cityscapes. Future applicability of the proposed system for design evaluation will rely on suitable data preparation and calibration of the associated algorithms using test subjects. The article describes characteristic details of the system’s architecture and discusses whether it would be able to acquire the level of sophistication required to provide aesthetic judgment that is convincing for humans.

Keywords: Affective Computing, Architecture, Artificial Intelligence, Autonomous, Agents, Cityscapes, Façade Design, Face Recognition, Fractal Dimension, Manifold Learning, Robotics, Skyline, Statistical, Learning

Introduction

For an understanding of the nature of design, it is important to investigate which principles and concepts are involved when humans process sensory signals associated with design. One possible approach is to implement relevant affect computing mechanisms in an artificial system and to analyse how the system performs when evaluating design. This approach is associated with several relatively new research areas in artificial intelligence (AI). These include affective computing (Picard, 1997), artificial life (Bedau, 2003; Johnston, 2008), bionics (Passino, 2005), computational models of emotion (Palensky and Barnard, 2009), human robot-interaction (Dautenhahn, 2007), and intelligent autonomous agent design (Albus and Meystel, 2001; Siegwart and Nourbakhsh, 2004; Choset...
et al., 2005; Thrun et al., 2005; Chalup et al., 2007c; Siciliano and Khatib, 2008; Dietrich et al., 2009).

The aim of anthropocentric biocybernetic computing (ABC) is to gain a better understanding of the mechanisms underlying human information processing, including aspects of language processing, face recognition, and perception of aesthetics. Crucial in this context are concepts of machine learning, which allow us to refine and tune a system iteratively until high skill levels can be obtained (Mitchell, 1997; Bishop, 2006). The name cybernetics was introduced by Norbert Wiener in the 1940s and became popular as part of the title of his book, *Cybernetics, or Control and Communication in the Animal and Machine* (Wiener, 1948). A central characteristic of cybernetics is to investigate the relations between different features of a complex system in order to gain a global understanding of the concepts governing the entire system. Biocybernetics is a part of systems biology and focuses on living systems and their environment. The approach is, as in cybernetics, holistic, and integrates a broad scale of information processing, from sub-cellular levels to ecosystems. The aim is to gain a new and deeper understanding of how complex organisms function by analysing the whole organism and its interaction with the environment. The purpose of the anthropocentric view in ABC is not to describe or explain the world from a human-centered perspective, but to gain new insights about how specifically humans develop, function, and survive. ABC regards humans as complex information processing systems and investigates information processing on different levels, including the cell, body, language and interaction with the environment. This includes modelling and analysing what impact processes such as design perception have on social factors, health, and emotional development.

There are different opinions about the role of emotions in research. Traditionally, emotions were regarded as a diffuse topic that was not well-suited as an object of scientific investigation. Mainstream AI almost exclusively focused on emotion-free tasks, such as chess or theorem proving (Russell and Norvig, 2003; Luger, 2009). In science, there is still an ongoing debate about the fundamental nature and understanding of emotions (Damasio, 1994; Panksepp, 1998; Barrett, 2006). Nonetheless, newer directions of AI, increasingly, address the processing of artificial emotions, that is, emotions in models and simulations (Palensky and Barnard, 2009).

The present project’s approach is motivated by two physiologically grounded properties of human visual perception that involve the emotional system. The first property is based on the hypothesis that the fractal dimension (Mandelbrot, 1983; Barnsley, 1988) of an image has physiologically measurable effects on the human body (Taylor et al., 2005; Taylor, 2006; Hagerhall et al., 2008). The second property is that the recognition of facial expressions is one of the most highly developed perceptual skills in humans for which designated brain regions have evolved or developed (Kanwisher et al., 1997; Farah and Aguirre, 1999; George et al., 2001; Pelphrey et al., 2003; Engell and Haxby, 2007).

A central aim of the present study is to utilise the latter two perceptual properties in a software system for architectural image analysis. This system could be interpreted as part of a “design evaluation module”, which could contribute to a more general “emotion module” for controlling future autonomous companion robots. Questions that arise in this context are: How do humans perceive aesthetics of the environment? What happens in the head of an architect? How can the underlying mechanisms be implemented in a software system? Robots and computers equipped with this kind of ability should be able to interact with humans in a much more natural and efficient way.
A number of previous studies help to pave a path towards this goal. Braitenberg (1984), for example, showed, with his thought experiments on perception of vehicles, that a simple system can already trigger unexpected emotional responses. This motivated us to use a relatively simple system design in order to obtain agent behaviour that may, in many cases, sufficiently convince the user, and would potentially be a useful tool for architects. Brooks (1999), meanwhile, proposed the subsumption architecture for behaviour-based robotics. It assumes that intelligent robotic behaviour is a property that emerges from a layered agent architecture through interaction with the environment. It is not, therefore, an encapsulated property of an algorithm or a software system (Brooks, 1990). Brook’s ideas motivate us to employ machine learning to enable the system to learn from the environment, in order to achieve high levels of skill. Doya (2002) presented a biologically inspired proposal, which explained how several neuromodulatory systems of the human brain may interact, and proposed relations to algorithmic concepts in machine learning. These concepts, the underlying circuitry and the associated chemical functionality, showcase how, in principle, an artificial emotion module could be designed and how it closely interacts with the other modules of the brain. Doya’s study indicates the level of complexity and finesse required for design of such a system. Breazeal (2002)’s work is one of the examples where concepts of “empathy” and a “theory of mind” have been implemented in systems of human robot interaction.

Beyond integration of the above-mentioned perceptual properties in a software system (Chalup et al., 2008a, 2008b), another central goal was to establish a world-model for aesthetic visual experience. For application in architectural image analysis, we had developed the concept of streetmanifolds (Chalup et al. 2007a, 2007b). Streetmanifolds use manifold learning (Lee and Verleysen, 2007) to extract a low-dimensional, non-linear representation of the essential content of a large set of images of house façades.

The remainder of this article consists of three parts. The first two address the above-mentioned properties of fractal dimension and facial expression perception in relation to architectural image analysis. The third part describes how to utilise streetmanifolds as a world-model to represent architectural visual experience.

Cityscape Analysis Based on Fractal Dimension Calculation of Skylines

The expression “fractal” can be used to describe point sets, for example, curves in space, which are highly non-smooth (Mandelbrot, 1983). There are many different ways to define the term fractal dimension (Edgar, 1993). One possibility is based on the box-dimension,

\[ D(S) = \lim_{\varepsilon \to 0} \frac{\log(N_{\varepsilon}(S))}{\log \varepsilon} \]

where \( S \) is a given set of points and \( N_{\varepsilon}(S) \) is the number of boxes in an overlayed lattice of boxes with edge length \( \varepsilon \) which intersect with \( S \) (Bouligand, 1928). Fractal methods have been employed in image analysis (Soille and Rivest, 1996) and for the analysis of the urban environments (Ostwald, 2001; Ostwald et al., 2009). A discrete approximation of the box-dimension can be achieved via box-counting, a method that has become common in architectural image analysis (Bovill, 1996; Ostwald and Tucker, 2007; Ostwald et al., 2008).

The use of the fractal dimension in our study is motivated by recent physiological studies that employed skin-conductance measurements (Taylor et al., 2005; Taylor, 2006) and quantitative electroencephalography (qEEG) (Hagerhall et al., 2008) to investigate what influences the perception of visual patterns of different fractal dimensions have on the human body. The results of these studies indicate that perception of patterns of mid-range fractal
dimension causes the least stress for the observer. These results support the hypothesis that
the fractal dimension is an expressive measure that can contribute to determine how visual
properties of the built environment impact on human-well being.

Using psychological eye-tracking experiments, it has been demonstrated that contours
with high-intensity gradients attract subjects’ attention (Rayner and Pollatsek, 1992). The
skyline is the contour of the sky segment in an image. Its important role for the aesthetic
assessment of distant urban views was emphasised by Heath et al. (2000). The fractal dimen-
sion of skylines can, therefore, be regarded as an important feature that may be used to
characterise cityscapes and natural scenes (Keller et al., 1987; Hagerhall et al., 2004). Fractal
analysis of urban skylines has previously been conducted by Cooper (2000, 2003) and Oku
(1990). The theory of contextual fractal fit implies that cityscapes look better if the fractal
dimension of their skyline matches the fractal dimension of the environment (Bovill, 1996;
Stamps, 2002).

Part of the present project was the development of a method for fractal analysis of a city-
scape’s skyline (Chalup et al., 2008a). The basic approach can extract the skyline from
standard images if the sky is a homogeneous, connected component which is adjacent to the
upper edge of the image and if the sky region is of higher brightness than the other parts of
the image. Technical details about how to exclude narrow objects such as cranes, powerlines,
and poles which intersect the skyline can be found in (Chalup et al., 2008a). Future research
may address images of cityscapes that display extreme lighting conditions, distinctive cloud
structure, extreme weather conditions, and other non-standard situations.

Pilot experiments showed that calculating a sensible skyline approximation can be an
unstable process that depends on various characteristics of the input image. It is a challenge
to decide which skyline among many possible approximations of the sky contour is the best.
We proposed a method that controls the skyline approximation locally via intensity cut-off
values (co). The smoothest approximation is determined by a suitable local minimum of the
fractal dimension when plotted as a function in dependency of the intensity cut-off values.
The process is visualised in Figures 1 and 2. The arrow in Figure 2 highlights the locally
best approximation at a local minimum for co=146. For lower cut-off values, the skyline
starts to invade façades. For higher cut-off values, the skyline starts to jump gaps between
buildings or to take off into the sky region. The graph in Figure 2 suggests that the fractal
dimension is locally stable for cut-off values where the smoothest approximation that appro-
priately separates the sky region from the cityscape is reached. Locally, therefore, the selection
of the skyline can be automated by using the local minima, but, globally, the process, cur-
cently, is only semi-automated and requires user input to select the most appropriate local
minimum.
Figure 1: Top: Original; 2\textsuperscript{nd} Row: $co = 95$; 3\textsuperscript{rd} Row: $co = 146$; 4\textsuperscript{th} Row: $co = 155$; 5\textsuperscript{th} Row: $co = 180$. The Best Skyline Approximation was Obtained for Cut-off $co = 146$ (Image in Third Row). Associated Fractal Dimensions are Displayed in Figure 2.
The described method was further employed in several experiments (Chalup et al., 2008a), which support two hypotheses:

1. Intersecting trees typically increase a skyline’s fractal dimension.
2. Fractal dimension of cityscapes’ skylines can be used to distinguish different types of cityscapes, for example, those consisting of historic houses vs. cityscapes consisting of modern apartment and office buildings.

In summary, the described fractal dimension of skylines provides an easy-to-calculate feature of the visual environment that is associated with aesthetic characteristics of cityscapes, and has physiological impact on the human emotional system.

**Façade Evaluation Using Facial Expression Classification**

Research in cognitive and brain science has found that large areas of the brain are associated with face processing and that communication via facial expressions involves the emotional centres of the brain (Engell and Haxby, 2007). The human visual system tends to associate abstract face-like patterns, such as smileys, with emotions of corresponding human facial expressions. Design elements, as they occur, for example, in house façades or frontal views of cars, can form abstract face-like patterns. The underlying hypothesis for this project is that abstract face-expression patterns can, similarly as smileys, trigger emotional responses of observers (Chalup et al., 2008b).

In order to utilise this property of the human visual system for automated architectural design analysis, a two-stage approach was implemented in software. In the first stage, a test image of a house façade was scanned to detect face-like patterns. If a face was detected, a box was put around it. In the second stage the detected patterns were classified into classes of different facial-expressions and their associated emotions. Both stages employed v-support vector machines (v-SVMs) with radial basis function kernel for classification (Schölkopf et al., 2000).

The face detection stage has five steps and is followed by facial expression classification in stage 2:

**Stage 1: Face detection**

- Preprocessing (includes, for example, histogram equalisation and edge detection)
• Select a random point c within the image.
• Select a random box size.
• Crop the image to extract the interior of the box generated with the random point c as center. Re-scale the interior of the box to a $20 \times 20$ pixel resolution.
• Apply a one-class ν-SVM classifier to decide if the box contains a face.

Stage 2: Facial expression classification using 8-class SVMs

Apply an eight-class ν-SVM classifier to determine to which of eight different facial expression classes the face in the box should be assigned to.

Figure 3: Application of Face Detection and Emotion Classification to the Image of Lenna; Small (violet) boxes = “surprised”; White boxes = “happy”; Larger (green) boxes = “disgusted” (Note Lenna’s Concave Shoulder Edge)

Support vector machines (SVMs) were previously employed for face detection, for example, by Heisele et al. (2001). The present study used one-class SVM classification for face detection where the output of the classifier delivers a decision value that indicated the probability that the tested box contains a face. Based on the decision values, a cloud of candidate solutions (highlighted yellow in Figures 3 and 4), which consisted of the center points of boxes with the highest decision values, was generated.

Once a box that contained a face-like pattern was detected, stage 2 of the procedure was applied to decide to which facial expression class the face pattern belongs. The field of affect recognition is growing and a recent overview can be found in the work of Zeng et al. (2008). Excellent results have been obtained in multi-modal approaches that use visual and acoustic data (Wang and Guan, 2008). Due to the planned application in architectural image analysis the present study can only use visual data. For training of the multiclass ν-SVM in stage 2, a labelled data set of 1828 images, which consisted of eight emotion classes, was selected. The classes correspond to Ekman et al. (2002)’s facial expression classification system’s (FACS) eight emotional states: surprise, fear, happiness, sadness, anger, disgust, contempt or neutral. Facial expression classes were colour coded via the face-enclosing boxes, which were detected in the first stage. The following colours were assigned to the eight emotional states:

• sad = blue;
• angry = red;
• surprised = violet;
• in fear = black;
• disgusted = green;
• contempt = orange;
• happy = white/yellow; and
• neutral = grey.

An example of how the face detection and facial expression classification system was applied is shown in Figure 3, which uses the well-known image processing test image of “Lenna”. Boxes that most tightly fit Lenna’s face were classified as “surprised” (violet) or “happy” (white). Some of the other boxes with high decision values were larger and were classified as “disgusted” (green).

![Figure 3: Image of Lenna](image)

Figure 3: Image of Lenna

After the system was tuned and trained on an image database of human faces and abstract smileys, it was applied to images of selected house façades. It was tested to see if the trained system would recognise, as humans would, face-like patterns in house façades similar to the one shown in Figure 4. As in this characteristic example, the system typically detected several faces associated with different emotions within the same house façade. The large (grey) boxes in Figure 4 indicate that the section of the façade included in these boxes corresponds to a “neutral” face. The middle-sized blue boxes propose that the enclosed face pattern belongs to the category “sad” and the small, white boxes, which contain a smaller fraction of the garage door, as “mouth”, contain the “happy” faces. In this example, all of the displayed boxes had a similarly high decision value.

**Streetmanifolds**

Horizontal and vertical lines play an important role in the architectural evaluation of house façades and streetscapes. Our hypothesis is that the visual experience gained by an architect through perception of thousands of line distributions can be represented by a non-linear geometric structure which we previously had introduced and called a “streetmanifold” (Chalup et al., 2007a, 2007b).

The human visual cortex is organised in such a way that local signals received at the retina are processed through a hierarchy of neuron layers into higher level representations which are multimodal, abstract, and holistic (Grill-Spector and Malach, 2004). Retinal ganglion cells are able to report the direction of movement (Barlow et al., 1964; Barlow and Levick, 1965; Masland, 2004). An important discovery by Hubel and Wiesel (1962, 1968) was that so-called simple cells in the striate cortex (V1) of cats and monkeys respond to oriented bars/edges in their receptive fields. Later it was found that these simple cells are not only...
sensitive to the orientation and location of bars, but also to the spatial frequency of visual stimuli (De Valois and De Valois, 1988). Engel (2005) suggested the presence of orientation-selective adaptation in human V1. These results were supported by Larsson et al. (2005), who also found that second-order stimulus orientation can emerge through continued processing in multiple visual areas, including V1 and V2. The perception and processing of line directions is, therefore, a fundamental concept of the human brain’s visual information processing system and it is regarded as essential for form recognition.

Many discoveries of modern neuroscience were predicted by Gestalt-theorists around 60-80 years ago (Westheimer, 1999). Gestalt psychology (Wertheimer, 1923; Koffka, 1935) argues, for example, that human visual perception is holistic. It proposes that the quality of a perceptual configuration depends on factors such as coherence, regularity, smoothness, continuity, unity, and simplicity. It was assumed that visual perception prefers continuous, over broken, transitions and that there should be neurons sensitive to collinearity (Spillmann and Ehrenstein, 2004). The visual system was predicted to complement missing parts of a structure and to be able to filter noise. Similarity-based clustering, driven by features such as colour, texture, size, and form, can occur (Kanizsa, 1979; Palmer and Rock, 1994).

Geometrically, a line is a set of points \( x = (x_1, x_2) \in \mathbb{R}^2 \) which can formally be described by \( x \in \mathbb{R}^2 : [\cos \phi, \sin \phi] \cdot [x - b] = 0 \), where \( \phi \in [0, 360^\circ] \) controls the slope of the line’s normal vector and \( b \in \mathbb{R} \) is its perpendicular distance from the origin. The Hough transform (HT) for lines takes a global view at an image to determine edge directions or lines, including interrupted or virtual lines (Hough, 1962; Illingworth and Kittler, 1988; Shapiro and Stockman, 2001; Gonzalez and Woods, 2002; Forsyth and Ponce, 2003). HT associates each image with an array of parameters \( (\phi, b) \in [0, 360^\circ] \times \mathbb{R} \) where each point corresponds to a line in the image. This array is often called accumulator array or Hough array. An example of a Hough array showing the line distribution extracted from a digital image of a house façade is shown in Figure 5.

![Figure 5: Left: Lines extracted by Hough transform overlayed to image of a house façade. Right: Corresponding Hough array. The height of the displayed peaks is the intensity of the corresponding line in the image above. Dominant lines were detected at angles 90°, 180°, 270°, and 360°](image-url)
The concept of manifolds has developed in mathematics over more than 100 years and is now central to geometry and topology (e.g., Spivak, 1979; Ito, 1987; James, 1999). Manifolds can represent high-dimensional, non-linear data concepts in a general and precise way. The mathematical definition and properties of topological, differentiable or Riemannian manifolds appear to reflect some of the terminology, for example, continuity, smoothness, coherence or connectivity, which was used by Gestalt theorists in the context of vision processing.

Manifold learning describes a set of methods for non-linear dimensionality reduction (NLDR). Burges (2005), and Lee and Verleysen (2007), give an overview of NLDR. The aim of NLDR is to take a set of high-dimensional data as input and return a low-dimensional representation of the essential geometric structure, which could be a low-dimensional manifold that was “hidden in” or “intrinsic” to the high-dimensional input data space. In the present study, the technique of isometric feature mapping (Isomap), by Tenenbaum et al. (2000), was applied. Isomap can be regarded as a generalisation of classical multi-dimensional scaling (Cox and Cox, 2001) where the Euclidean distance between input data points is replaced by a graph approximation of their geodesic distances on the intrinsic manifold.

The concept of streetmanifolds (Chalup et al., 2007a, 2007b) was introduced to provide an holistic geometrical representation of the visual experience that can be gained through evaluation of a large set of house façades. Navigation within a streetmanifold corresponds to continuous morphing or interpolating between façade designs represented by the data set. Streetmanifolds are based on the calculation of pairwise distances between digital images of house façades. The procedure applied in this study was as follows:

1. Take a large set of images of house façades.
2. Calculate their Hough arrays.
3. Calculate a distance matrix consisting of pairwise distances of the Hough arrays.
4. Apply isomap (Tenenbaum et al., 2000).

An example streetmanifold is shown in Figure 6. The displayed manifold is three-dimensional and was calculated with isomap following the procedure outlined above. The third dimension is encoded in colour. Details of the calculation can be found in Chalup et al. (2007a). On the streetmanifold, similar houses were assembled in clusters. Figure 6 shows four example clusters. Cluster A consists of narrow, two-storey houses. Cluster B contains wide, flat houses. Clusters C and D are more centrally located on the manifold and show houses with a relatively homogeneous distribution of horizontal and vertical lines. Cluster C is characterised by the fact that a tree is standing in front of all the houses. More results, including how to obtain proto-plans for new houses by interpolating between points on the streetmanifold, were discussed by Chalup et al. (2007b).
Discussion

Human aesthetic judgement is based on many factors. In its current state of development, the software system proposed in the present article employs only three perceptual concepts, namely fractal dimension calculation, facial expression recognition, and line detection. A system that integrates all three concepts and represents the extracted information in the form of an emotion manifold, similar to the streetmanifolds described above, would, therefore, still be far from being able to emulate human performance when judging design. Machine learning can help to improve performance of the system, but it cannot replace missing concepts.

However, the thought experiments of Braitenberg (1984) suggest, that despite its restricted scope and speculative nature, the described system design may already be sufficient to equip a companion robot with a basic emotion module. In the long run, these types of emotion modules may help to improve human-robot interaction and may also serve as tools to assist architects on selected tasks of architectural image analysis in areas associated with fractal analysis or face recognition. A robot, equipped with this type of control system, would be able to evaluate its environment and to comment on the architecture of surrounding buildings.
and streetscapes. An implementation of an exact biological model of emotion processing is not realistic. But it is promising to include biologically inspired mechanisms into the system design as long as these mechanisms can extract features from the environment that are relevant for the aesthetics of design.

Biological evidence that local features are integrated in the brain into global shapes through processing in multiple visual pathways comes from fMRI studies in humans and monkeys (Kourtzi et al., 2003). In a similar way, the local features of line distributions, the fractal characteristics of skylines, and facial expression associations of the visual environment described in the present study, together with some additional features that may be part of future investigation, may lead to a sophisticated internal world-model of architectural visual experience.

Acknowledgement

This study received support from the ARC discovery project grant DP0770106 “Shaping social and cultural spaces: the application of computer visualisation and machine learning techniques to the design of architectural and urban spaces”. We are grateful to all our research assistants and previous co-authors for their contributions to software and experiments in the underlying series of technical papers (Chalup et al., 2007a, 2007b, 2008a, 2008b).

References


79

About the Authors

Dr. Stephan K. Chalup
Dr. Stephan K. Chalup is director of the Newcastle Robotics Laboratory and a senior lecturer in computer science and software engineering at the University of Newcastle, Australia. He has a PhD in Computer Science/Machine Learning from Queensland University of Technology in Brisbane, Australia, and a Diplom in mathematics with biology from the University of Heidelberg in Germany. In his research he investigated applications of artificial neural networks, evolutionary algorithms, kernel machines, and techniques for dimensionality reduction in areas such as image processing, intelligent system design, language processing, and robotics.

Prof. Michael J. Ostwald
Dr. Michael J. Ostwald is Professor and Dean of Architecture at the University of Newcastle, Australia. He is a Visiting Fellow at SIAL and a Professorial Research Fellow at Victoria University Wellington. He has a PhD in architectural philosophy and a higher doctorate (DSc) in the mathematics of design. He is co-editor of the journal Architectural Design Research and on the editorial boards of Architectural Theory Review and the Nexus Network Journal. His recent books include The Architecture of the New Baroque (2006), Homo Faber: Modelling Design (2007) and Residue: Architecture as a Condition of Loss (2007).