NOVA

University of Newcastle Research Online

nova.newcastle.edu.au

Hong, Kenny; Chalup, Stephan K.; King, Robert A. R. "Affective visual perception using machine pareidolia of facial expressions". Originally published in IEEE Transactions on Affective Computing Vol. 5, Issue 4, p. 352-363 (2014)

Available from: <http://dx.doi.org/10.1109/TAFFC.2014.2347960>

© 2014 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works

Accessed from: <http://hdl.handle.net/1959.13/1060047>

Affective Visual Perception using Machine Pareidolia of Facial Expressions

Kenny Hong, *Member, IEEE*, Stephan K. Chalup, *Senior Member, IEEE*, and Robert A.R. King

Abstract—This article presents a computer vision approach that can detect and classify abstract face-like patterns, including subliminal faces within a scene. This can be regarded as a way of simulating the phenomenon of pareidolia, that is, the tendency of humans to ‘see faces’ in random structures such as clouds or rocks. The paper describes the system consisting of a component-based face detector and an expression classifier. The face detector creates a number of component images from the original image at different resolutions. A component image is a binary edge image where the edges are segmented into components using a labelling method with a border-following technique. The component images are then overlaid to produce a component height map where large and notable components across all resolutions have high values, while specular and noisy components have low values. The method retains three-shape components, representing two eyes and a mouth, that have height map values that are larger than the noise cut-off value. Support vector machines using scale-invariant feature vectors are applied for ranking these three-shape components by their geometry and size, and their shape semblance to human faces in the training data. The outcome is a facial expression analysis system that uses face components, with the potential to estimate an emotional expression value for a scene by producing an array of emotion scores corresponding to Ekman’s seven Universal Facial Expressions of Emotion. An advantage of this technique, when compared to a holistic method, is that the face components are explicitly isolated. This supports a process of abstraction that can facilitate the detection of distorted and minimal face-like patterns.

Index Terms—Affect, design perception, machine pareidolia, faces, facial expressions, support vector machines, heuristics

1 INTRODUCTION

CAN a machine describe human emotional responses to a visual scene using methods of face perception and facial expression analysis? One associated process that is involved in how humans perceive a scene is pareidolia – a phenomenon where the brain tries to draw significance from random and vague stimuli [7]. The most common examples are when an observer sees faces where actual faces do not exist. The reason why faces and facial expressions play a central role in producing emotional responses can be explained by the interconnected nature of specific brain regions that are active for processing visual scenes (the parahippocampal gyrus [11]), faces (the fusiform gyrus [23]) and emotions (the amygdala [31]). The present article builds on some of our previous work and its neurophysiological inspirations [6, 7, 19]. Among these it should be highlighted that the brain responds to faces already very early, before cognitive processes and awareness set in [14], and that the fusiform gyrus is believed to be a brain area dedicated to process faces holistically [39]. For a simulation of facial pareidolia, however, a holistic system can have some limitation that only faces with components that bear similar geometry (i.e. size and shape of mouth and

eyes) to that of the training data would be found. For distorted face components to be detected (i.e. cartoon and caricature faces), an alternative direction beyond a holistic path has to be explored.

It has been reported that the inferior occipital gyrus structurally processes face components in a hierarchical order, giving preference to the eyes followed by the mouth and then the nose [2]. There are some studies that argued that also the fusiform gyrus is sensitive to face components [2, 21]. Figure 1 illustrates where the parahippocampal gyrus, amygdala, inferior occipital and fusiform gyri are located in the human brain.

Inspired by the existence of a specific brain region known to structurally process faces by components, we investigated literature from the machine learning and computer vision community for suitable component-based face detectors. A requirement in our investigation was to look for face detectors that can detect at least the eyes and the mouth because we assumed that these would be the minimal set of components required to elicit a pareidolia-like perception of faces. This assumption is supported by studies that used schematic face-like patterns in experiments with newborns and suggested that these babies were capable of perceiving such patterns [22]; also implying that the human brain may be hard-wired to detect faces with increased priority from birth. Another assumption of our investigation was that it is unnecessary to use skin colour in the detection mechanism because we wanted to highlight the structural aspects of the input patterns and it seemed inappropriate to limit the planned face processing approach to specific skin colours.

• Kenny Hong is with the School of Electrical Engineering and Computer Science at the University of Newcastle, Australia. Email: Kenny.Hong@uon.edu.au

• Stephan K. Chalup is with the School of Electrical Engineering and Computer Science at the University of Newcastle, Australia. Email: Stephan.Chalup@newcastle.edu.au

• Robert A.R. King is with the School of Mathematical and Physical Sciences at the University of Newcastle, Australia. Email: Robert.King@newcastle.edu.au

Under the above assumptions we found only a few component-based face detection studies that were relevant for the planned system. These can be categorised as following either a top-down or a bottom-up approach. Top-down approaches are methods that use found faces, and then isolate the face components using facial feature detectors – this can, for example, be a Viola & Jones Haar-like face and facial feature detector for finding a face and then its mouth, nose, eyes and ears using expected regions of interest for the search [4] (the Viola & Jones Haar-like feature detector detects objects using the Integral Image and has gained popularity because of its computational speed for finding faces [36]). Among the studies that assume a face is already found, a common technique for finding face components is using manually inserted feature points to train (or construct) a statistical model, and then searching for or aligning each feature point in the model over the found face. Several studies share this procedure – for example, an Active Shape Model (ASM) to look for eyes, brows, nose, mouth and jawline [20], a Bayesian model with a consensus of exemplars to look for brows, eyes, nose, mouth and chin [3], a Conditional Regression Forest model to look for eyes, nose and mouth [9], and also a multiple of Support Vector Machines (SVMs) and statistical Point Distribution Models (PDMs) to look for brows, eyes, nose and mouth [29].

Bottom-up approaches are methods that find the face components first and then assesses their configuration using statistical models – among these is, for example, a boosted cascade of classifiers to look for eyes, nose and mouth, followed by an assessment of their topology using a PDM of face component centres [10], and similarly, an AdaBoost trained classifier cascade of Haar-like features to look for eyes, nose and mouth followed by an assessment of their topology using graph matching techniques [13]. Lastly, [25] uses a weighted mask function and matching rules to look for faces in the form of isosceles or right-angled triangles where the corners correspond to face components, and the face components are extracted using a binarised greyscale image.

The methods currently available in the literature for component based face detection do not fully meet the requirements of the present project. Firstly, the available methods, with the exception of [25], do not aim at detecting minimal face-like patterns but instead detect individual face components such as noses, brows, jawlines and chins. Secondly, the methods available in the literature are focused on images of real human faces. And finally, all cases except [25] uses negative data. The use of negative data is not desirable for the present project because, as previously mentioned, the background literature in neuroscience suggests that our brains are hard-wired to look for faces from birth [22]. The present study consequently assumes that the system's (and the brain's) encoded knowledge would primarily be influenced by face data (i.e. a single positive class) rather than a second negative class encoding non-face data (and accordingly,

one-class SVMs were chosen for this study [33]). Given that [25]'s study do use only positive data, however, the detection depends on two threshold values – one at the initial step for creating a binarised greyscale image and the other at the decision making, which can be cumbersome. For a more comprehensive literature review we refer to [15].

Our proposed method is based on the idea that the raw features of an image allow to reveal the faces within. We achieve this by using a connected components labelling method to segment and label the edges of an image into components. Then by analysing grouped components, distorted abstract faces can be perceived, as it is desired for simulating the pareidolia phenomenon. Using this approach creates a very generalisable component-based face detector – it allows the eyes and mouth to be captured and explicitly isolated. The one-class SVMs are used for estimating the semblance of the detected face components to the faces in the training data via scale-invariant feature vectors, and the multi-class *pa*-SVMs are used for classifying the expressions [17]. The idea of our method appears to have some similarity to [25], which used triangles of face components to detect human faces; however, our approach offers several advantages – which we will discuss in Section 8.

A system overview is presented in the next section, followed by a description of the dataset in Section 3. Section 4 presents the classification method. The new component face detector and the scale-invariant feature vectors used for the semblance classifiers are described in sections 5 and 6. An example walkthrough of the component face detection process is provided in Section 7. The experimental results are discussed in Section 8, where also extensions of our component face detector are presented. Summary and Conclusion follow in Section 9 also addressing possible roads of future research emerging from this study.

Fig. 1: The regions of the brain that are active for faces and facial expressions and scenes: The inferior occipital and fusiform gyri are responsible for face perception, the amygdala is responsible for processing emotions, and the parahippocampal gyrus is responsible for scene perception. This illustration is based on images from [5, 26].

2 SYSTEM OVERVIEW

The proposed system consists of several steps as illustrated in Figure 2 for the face detection procedure, and Figure 3 for the classification process. Beginning with Figure 2, the first step downsamples an image into a number of low resolution images. A binary edge image is created using a Canny edge operator for each downsampled image. Then a component image is created using a connected components labelling method and the most relevant components are then selected, representing respectively two eyes and a mouth, and are ranked by geometry, size and shape semblance classifiers trained using the natural human faces of the training data. Section 5 discusses this in detail.

The process for training the semblance and expression classifiers is shown in Figure 3. The first step of Figure 3 was to align the face dataset using manually inserted feature points (purely for training). Then, using the feature points we created cartoon faces (described in the next section) and extracted the face components for the semblance classifiers (to assemble the scale-invariant feature vectors reflecting the geometry, size and shape semblance of the faces in the training data) and the expression classifiers (by normalising the input of the face component classifiers between the values of -1 and +1 for training). Next, for both processes the dataset is partitioned into classes (seven classes for the expression classifiers and one-class for the face detectors), and divided into training and validation sets using a 10-fold cross-validation algorithm.

3 DATASET AND IMAGE PREPROCESSING

The face dataset was generated using the JACFEE and JACNEUF datasets by the Paul Ekman Group LLC [27]. We specifically chose these datasets because they have discrete expression classes (the seven universal facial expressions of emotion as well as neutral faces). The generated dataset consists of 280 face images, including 140 neutral faces and 20 images for each of the seven facial expressions: happy, sad, surprised, fearful, disgusted, angry and contemptuous. We consider the neutral face separate from the seven universal facial expressions of emotion, and because of this, it was not used for expression classification; however, we used it for creating the semblance classifiers.

The face model defined by the manually inserted feature points was proposed in our previous publications [16, 18] which showed a component approach performing better than a holistic approach for classifying discrete facial expressions. An illustration of our face model is shown in Figure 4, where a total of 60 feature points were manually inserted for each face image along the brows, eyes and mouth; the 2 feature points of the inner eye corners (a.k.a. lacrimal caruncle) were used for aligning the faces in the dataset.

The cartoon faces for each image are created by drawing lines between the manually inserted feature points,

Fig. 2: Overview of the component face detection process showing the flow of the stages and states.

and the face components (eyes, brows and mouth) are cropped using these feature points. Because the sizes of the cropped face components are different from one another, we found for each set of face components the maximum width and height of the cropped image, and then padded the images to that maximum size. Figure 5 shows sample images of the seven universal face expressions of emotion in cartoon and cropped face components.

4 CLASSIFICATION METHODS

Following our previous publication [19] *pa*-SVMs were chosen for creating the expression classifiers [17]. Specifically, we evaluated *C*-SVMs and ν -SVMs with a Gaussian kernel and a polynomial kernel of degree 2.

4.1 The SVM and kernel parameters to optimise

The parameters that had to be tuned included (C, γ) for the *C*-SVMs and (ν, γ) for the ν -SVMs – where *C* is a parameter for *C*-SVMs, ν is a parameter for ν -SVMs, and γ denotes the parameters for the Gaussian (1) and the polynomial (2) kernels.

$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2} \quad \gamma > 0 \quad (1)$$

Fig. 5: Sample images of the seven universal expression of emotion classes in cartoons (first row) and cropped face components (second row). Note that the width and the height of the cropped face components between expressions are the same. This is because they have been padded to the maximum values found within that set. The dataset was generated from the JACFEE and JACNEUF databases [27].

Fig. 3: Overview of the classification process.

Fig. 4: Face model with manually inserted feature points [16, 18]. The feature points shown in red and labelled '1' are used for aligning the faces in the training dataset.

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}_i \cdot \mathbf{x}_j + 1)^2 \quad \gamma > 0 \quad (2)$$

where $\mathbf{x}_i, \mathbf{x}_j$ are the samples.

4.2 SVM and kernel search parameters

We use the same (C, γ) search parameters as in our multiclass SVM experiments [17]:

$$C \in \{2^{-5}, 2^{-4}, \dots, 2^{14}\} \text{ and } \gamma \in \{2^{-17}, 2^{-16}, \dots, 2^2\}$$

And likewise, for the sequences of (ν, γ) we grow ν by a linear step of 0.05 up to 1.0, while γ remains the same, i.e. $\nu \in \{0.05, 0.10, \dots, 0.95, 1.00\}$. Note that although C and ν are closely related ν can only vary between 0 and 1 while $C > 0$ is not restricted [8]. For the one-class SVM experiments, which use only ν and γ , we use the same (ν, γ) values as described for the multi-class SVM experiments.

5 COMPONENT FACE DETECTOR

This section describes details of the component face detection process that was presented in the system overview in Figure 2. For the initial step of downsampling, we have chosen 10 resolutions with a height and width of 15, 19, 24, 30, 38, 47, 59, 73, 92 and 114 pixels (a scaling up of 1.25). Then a Canny edge operator was

applied to compute the binary edges. The edge operator had several parameters that impact the final edge image and the selection was by visual evaluation for ideal face edges. The chosen parameters were: aperture size of 3, lower threshold of 127 and upper threshold of 255.

5.1 Computing the Component Height Map

The purpose of the height maps is to provide regions of interest defined by global and local peaks. We consider these peaks to hold relevant candidate face components because they are prominent in all the component images.

We obtained the component images using a connected components labelling algorithm with a border-following technique [35]. We segmented the edges into ‘components’. For each component, we construct a rectangle around those pixels – defined by x -min, x -max, y -min, y -max.

Because the components are described by pixels and rectangles, there are two possible methods for building the height map. One using the pixels (‘pixel-based’), and the other using the rectangles (‘block-based’). In this article we present the ‘block-based’ approach as it has shown to return a better performance than the ‘pixel-based’ approach. This is discussed in detail in [15].

5.2 Computing the Integral Rectangle

To model the sum of the components contributing to each peak we construct a new rectangle, which we call an ‘integral rectangle’. This process is by averaging each of the four corner points over the selected rectangles, and masking the area to be ignored from further processing. To begin computing the integral rectangles we start from the highest point of the height map down to a noise cut-off (which is defined by an average of non-zero height values). For each operating point, we selected all the rectangles (derived from the component images) containing that point. A heuristic is used to reject large selected rectangles by downsampling. This heuristic averages the pixel height value for each component, and if it falls below the noise cut-off value it is rejected. The purpose of this is to allow the computed integral rectangle with the tightest fit to the peak region.

5.3 Merge conditions for overlapping rectangles

The next step is to decide whether to merge overlapping integral rectangles (they usually overlap due to high height values residing next to masked regions). Two rectangles are merged if they satisfy two conditions. The first condition is that the two rectangles have the same aspect ratio. The second condition is that the height values (from the component height map) are within a threshold:

$$\text{threshold} = \frac{\sum_{i=k}^{m-1} (y_{i+1} - y_i)}{m - k} \quad (3)$$

where y_i is the pixel count of height values along the x axis from low to high, m is the number of height values, and k is the height value along the x axis that is above the noise cut-off. The purpose for the threshold is to prevent the merging of components with largely different height values. A walkthrough in Section 7 will show the plots of the height values versus pixel count with the noise cut-off.

5.4 Matching rules to collate three-shape components

Once the merge operation is complete we use a set of matching rules to produce a set of three-shape components, where the shapes are the integral rectangles from the previous step – representing (respectively) two eyes and a mouth. The matching rules are:

- 1) Mouth rectangle center below and between left and right eye rectangles.
- 2) Top of one eye is not below the bottom of the other eye.

The final step is to determine a confidence rating for the candidate faces. We evaluated two metrics for this purpose. The first metric sums the height values of each face component using the component height map; a higher sum reflects a higher confidence rating. This is then compared with the decision values of the one-class semblance classifiers. The decision value is considered a confidence rating in relation to the training data [19], i.e. if the decision value is positive and it is very high, it means that the classifier sees this as a face and the components are very similar to the face components of the training data. Then, the expressions of the face components are determined. This is by classifying the face components after they have been resized, preprocessed and padded to the maximum widths and heights used for training the component expression classifiers (see Section 3). Details of the feature vectors used for the semblance classifiers are described next.

6 SCALE-INVARIANT FEATURE VECTORS FOR SEMBLANCE CLASSIFIERS

This section describes the feature vectors used for the semblance classifiers that were presented in the system overview section in Figure 3. As described earlier, our goal is to create scale-invariant feature vectors that model the geometry, size and shape semblance from the natural faces of the training data. To begin, we construct rectangles around the face components in the training data using the manually inserted feature points – i.e. brows, eyes and mouth. We use angles, area and aspect ratios to describe the geometry, size and shape of the face components. Figure 6 is used to assist with the description.

Fig. 6: Face rectangles for producing scale-invariant feature vectors; the angle for the middle of the mouth (m), and the aspect ratio using the width (x) and height (y) is shown.

6.1 Geometry feature vectors

For *geometry*, we proposed 2 feature vectors:

$$\angle \text{mid eyes} \quad (4)$$

$$\angle \text{mid mouth} \quad (5)$$

where $\angle \text{mid mouth}$ is shown as m in Figure 6 and $\angle \text{mid eyes}$ is the opposite angle. Both angles use radians as input to the SVM.

6.2 Size feature vectors

For *size*, we proposed 2 feature vectors:

$$(L + R)/2M \quad (6)$$

$$(L - R)/M \quad (7)$$

where L , R , and M represent (respectively) the areas of the left eye, right eye and mouth; defined by the width multiplied by the height of the rectangle. The meaning of the above feature vectors relates to:

- Average area of the eyes divided by the area of the mouth (6).
- Eye area asymmetry divided by the area of the mouth (7).

6.3 Shape feature vectors

Lastly, for *shape*, we proposed 3 feature vectors:

$$(L_y/L_x + R_y/R_x)/2 \quad (8)$$

$$\frac{2(L_y/L_x - R_y/R_x)}{(L_y/L_x + R_y/R_x)} \quad (9)$$

$$M_y/M_x \quad (10)$$

where (L_y, L_x) , (R_y, R_x) and (M_y, M_x) are the aspect ratios of the left eyes (L), right eyes (R), and mouth (M), respectively, with an aspect height defined by y and an aspect width defined by x . The meaning of the above feature vectors relates to:

- Average aspect ratios of the eyes (8).
- Scaled eye aspect ratio asymmetry (9).
- Mouth aspect ratio (10).

In our experiments we assessed the cross-validation performance when we joined the feature vectors together and when more data was added from different resolutions. As an extension, the above procedure was repeated

for brows instead of eyes; this is because the three-shape configuration is applicable to both.

7 EXAMPLE WALKTHROUGH OF THE COMPONENT FACE DETECTION PROCESS

We use the image in Figure 7 to provide a walkthrough of the steps of the component face detection and expression classification method that was described in Section 5.

Fig. 7: Example image for describing the steps of the component face detection and expression classification method [Flickr image under a creative commons license, taken by flickr ID: Alexandre Hamada Possi].

The component images of the initial downsampling step using the 10 chosen resolutions are shown in Figure 8.

Fig. 8: Component images of the walkthrough example in ten different resolutions. In the 'block-based' approach a component (in each subimage) is established by the pixels of the same colour and their tightest enclosing rectangle.

The 'block-based' approach for building the component height map is shown in Figure 9, with an equal weight for overlaying the components.

Figure 10 shows the plots of the height values (= block depth) versus the pixel count. As explained in Section 5.2

block-based

Fig. 9: A 'block-based' component height map with an equal weight for overlaying the components. That is, we see here the sum of the component rectangles extracted from all the different resolution images in Figure 8.

the average pixel count is used to determine the noise cut-off, which separates noisy and relevant regions. The height map of the relevant regions (without noise) is shown in the bottom image of the same figure. We use the components from the relevant regions in the bottom image for calculating the threshold used to determine a merge condition, as described in Section 5.3.

relevant regions

Fig. 10: Finding the noise cut-off value and showing the relevant regions of the component height maps.

The top face candidate using the 'block-based' approach is shown in Figure 11. The top face candidate has the highest summed height values from each face component. The colour coded component expressions of

the 'block-based' approach shows that the left eye is 'in fear' (grey) and the right eye and mouth is 'surprised' (violet).

The decision values of the one-class classifiers for the 'block-based' approach is also shown in Figure 11. Examining the decision values, the ν -SVM one-class classifiers return a positive value. The positive value indicates a positive face candidate and the number is considered a general indication of the classifier's confidence rating [19], but is not comparable across kernels. In the same figure the classified expressions are shown below the image. The four multi-class expression classifiers show a unanimous decision for the expression of the left eye (fear), right eye and mouth (both surprise).

Given the above example, it is possible to arrive at conflicting expressions between face components. One question springs to mind: 'Which face component expression is a reliable indicator?' Our results in Section 8.6 show that the mouth expression returns a cross-validation rate almost equivalent to that of the whole face.

'block-based'

one-class decision values	mouth expression
$(\nu, poly) : 1.4e-06$	$(\nu, poly) : surprise$
$(\nu, rbf) : 2.7e-04$	$(\nu, rbf) : surprise$
$(svdd, rbf) : -0.14$	$(svdd, rbf) : surprise$
$(svdd, poly) : -0.20$	$(svdd, poly) : surprise$
left eye expression	right eye expression
$(\nu, poly) : fear$	$(\nu, poly) : surprise$
$(\nu, rbf) : fear$	$(\nu, rbf) : surprise$
$(svdd, rbf) : fear$	$(svdd, rbf) : surprise$
$(svdd, poly) : fear$	$(svdd, poly) : surprise$

Fig. 11: Top candidate face detected and expressions classified. In both situations the one-class classifier using a ν SVM and the polynomial kernel returned a positive value. The positive value indicates a positive face candidate and the number is considered a general indication of the classifier's confidence rating [19], but is not comparable across kernels. The four multiclass expression classifiers show a unanimous decision for the expression of the left eye, right eye and mouth.

8 RESULTS AND DISCUSSION

We begin this section by presenting a number of different face types detected and expressions classified, followed by cases of unsuccessful detection. Then we go into the individual classification results, starting with the best

cross-validation rates of the semblance classifiers using the scale-invariant feature vectors. We then show that the feature vectors can easily distinguish three-shape eye faces from three-shape brow faces (but are inadequate for expression classification). Next, we show that the mouth expression alone is almost equivalent to that of the whole face. Later in this section, we present experiments that show the impact of excluding the contemptuous expression (which is common in the literature), followed by a comparison of our method with the literature.

8.1 Successful detection of different face types

We wanted to show that our system can work for a variety of pareidolia scenarios and that the judgement/selection of the images was not done by us alone. Therefore 66 samples of *pareidolia* images were randomly selected from the flickr ‘found faces’ pool (under the creative commons license) and manually cropped by our experimenter. Using the flickr data represents a kind of ground truth because it is the outcome of a selection of pareidolia pictures from a large number of people who already had labelled these examples as pareidolia images. In order to obtain further objectivity in data selection and labelling a specifically designed crowd sourcing experiment could be designed in future research. Of the 66 samples selected and tested from the pool, Figure 13 shows a subset of 24 examples. The detected faces shown have the highest summed height values (from each face component) of all the face candidates. The colour code for each component expression is as follows: sad (blue), angry (red), surprised (violet), fearful (grey), disgusted (green), contemptuous (orange) and happy (yellow). The SVM type and the kernel of the one-class classifiers that returned a positive decision value are shown directly below the image. As mentioned earlier, for conflicting face component expressions we show in Section 8.6 that the mouth expression returns a cross-validation rate almost equivalent to that of the whole face.

Examining the figure, the first observation is that the expressions of face components are sensitive to the size and location of the bounding rectangle. A larger bounding rectangle would prevent proper alignment of face components to the trained images; for instance, the first image in Figure 13 by ‘joseanavas’ and the yellow image by ‘thentoff’ (fifth row), where the larger bounding rectangle shifts the alignment of the mouth to prevent a surprised expression being properly classified. Similarly, if a bounding rectangle is small enough to clip important edge features, a different expression would be seen. Examples are the earphone image in Figure 13 by ‘thentoff’ (third row) and the image by ‘nottsexminer’, where a smaller bounding rectangle returns a disgusted expression.

The larger bounding rectangles are caused by large components prominent across the number of different resolution images. For example, looking at the ten resolution images in Figure 8 of Section 7, we can see that

some components around the door of the house consume most of the image. In this situation, the size and location of expected ‘integral rectangles’ are shifted. To show this more clearly, we refer to the images in Figure 12 using the ‘pixel-based’ approach. The shifts in location of face components shows that the ‘block-based’ approach is reliable (the reason why it was chosen over the ‘pixel-based’ approach). These shifts can be deduced via the peaks of the component height maps shown in the same figure.

Fig. 12: Example of component height maps shifting location of face components. The reason for the shifts in the ‘pixel-based’ approach can be deduced by examining the cluster of heights in the ‘block-based’ height map, and the highest height values of the ‘pixel-based’ height map.

Because the larger-than-expected components obtained by the border-following connected components labelling method [35] can increase the size of face components, we have rejected face candidates where face components overlap. And because of this, there are several examples in Figure 13 where no face candidates were returned by the ‘pixel-based’ approach.

Lastly, as we were gathering the decision values of the one-class semblance classifiers, we noted that in all cases the $(\nu, poly)$ one-class classifier returned positive decision values, followed by the (ν, rbf) one-class classifier. The $(svdd, poly)$ and $(svdd, rbf)$ one-class classifiers produced mostly negative decision values, which means they did not see the top candidate faces as positive faces.

8.2 Cases of unsuccessful detection of different face types

In Figure 14, we have selected a few examples of *pareidolia* images where no face candidates were produced. We can see that the ‘integral rectangles’ are present for each face component; however, the rectangles are either overlapping each other or they have been merged with other rectangles that cause the face components to overlap, thus preventing them being a face candidate. The images shown are from the ‘block-based’ approach, and only the height values in the height map that are above the noise cut-off are shown.

Next we present the individual classification results of the scale-invariant feature vectors used for the semblance classifiers, and the face components used for the expression classifiers, followed by the impact when the contemptuous expression is excluded in the classification process.

Fig. 13: Different face types detected and the expressions classified: The faces shown have the highest summed height values (from each face component) of all the face candidates. The colour code for each component expression is as follows: sad (blue), angry (red), surprised (violet), fearful (grey), disgusted (green), contemptuous (orange) and happy (yellow). The SVM type and the kernel of the one-class classifiers that returned a positive decision value are shown directly below the image [Flickr images under a creative commons license - the flickr IDs are shown below the first image].

Fig. 14: Example of unsuccessful detection of different face types using the ‘block-based’ approach: We can see that the ‘integral rectangles’ are present for each face component; however, the rectangles are either overlapping each other or they have been merged with other rectangles that cause the face components to overlap (2nd column), thus preventing them being a face candidate. The height maps shown have values that are above the noise cut-off.

8.3 Semblance classifiers using joined feature vectors return best results

We compared the performance of the one-class semblance classifiers over the 4 possible combinations of C or ν -SVMs and RBF or polynomial kernels.

The ν -SVM with the polynomial provided the best performance (100% correct classification rate across all resolutions from 15×15 to 114×114). This applied to both the ‘eyes and mouth’ and ‘brows and mouth’ three-shape configurations.

Other configurations performed well, with most having a correct classification rate above 96%. The worst performer was the C -SVM with the polynomial kernel for some resolutions, which had an 86% correct classification rate. Combining similar resolutions into groups reduced the variance of the correct classification rates.

More information and graphs showing details of the classification results of various techniques tested (including those for the results described in sections 8.4 to 8.7 below) are available in the thesis [15].

8.4 Eye faces are easily differentiated from brow faces using two-class SVMs

We performed two-class cross-validation experiments and found [15] that the feature vectors using eyes and mouth can be easily differentiated from the feature vectors using the brows and mouth. The highest results are when the geometry, size and shape features are joined (i.e. 7 feature vectors); followed by shape (3 feature vectors using aspect ratios), geometry (2 feature vectors using angles) and then size (2 feature vectors using areas).

8.5 Feature vectors are inadequate for expression classification

The feature vectors were derived from rectangles of face components (see Section 7). We performed multi-class

cross-validation experiments and obtained only very low accuracy for the classification of expressions using these feature vectors. These results are expected because the rectangles do not contain sufficient discriminatory information

8.6 Expression analysis of face components shows mouth with the best result

The multi-class cross-validation results of Figure 15 show that the mouth is the best feature for discriminating between expressions. When compared with the holistic results, the mouth appears to be equivalent to the results involving the complete face.

8.7 Excluding the contemptuous expression returns better results

In the literature the contemptuous expression is often not considered. This expression is special because it is the only facial expression with an asymmetrical property. By excluding this expression the experiments return improved results. This is as expected because adding the expression increases the classification difficulty.

8.8 Comparison to the literature

As we have mentioned in the introduction, our method appears to follow a similar idea to [25], which used triangles of face components to detect human faces. Our approach differs in the following ways. Firstly, we use binary edges as the initial step rather than a binarised greyscale – this should be more robust than a manually tuned threshold. Secondly, instead of using weighted masked functions and providing a threshold value for decision making, we construct a height map and use the combined peak heights of each face component to quantify the confidence of a face. This is then complemented by decision values (which reflect confidence) from the one-class classifiers.

9 SUMMARY AND CONCLUSION

In this article we have extended our work of producing an emotional response to a scene from holistically perceived faces, to perceiving face components such as the eyes and mouth, and classifying the expressions of emotions. The idea is to detect abstract faces where the eyes and mouth are distorted beyond that of a natural human face. The significance of this study is to simulate the pareidolia capability of humans to produce an emotional response to a scene using expressions of perceived face components.

The motivation of our system design came partially from neuroscience studies that show that the lateral occipital complex responds to illusory contours and contour integration [1, 34]. Contour integration is “the linking of collinear but disconnected visual elements across space, [which] is an essential facet of object and

Fig. 15: Expression analysis of face components shows that the mouth is the best feature for discriminating between expressions. The left graph shows the classification results (for both SVM and both kernel types) for the three face components (brows, eyes, and mouth) and the right graph shows the results of the holistic expression classifiers (i.e. whole face). We can see that the mouth (top green markers in the left graph) returns about the same accuracy as the whole face classifier in the right graph. The results for the mouth are clearly above the results for eyes and brows and this is the case for all combinations of SVMs and kernels tested.

scene perception [34]". Illusory contours are black and white images of ordered shapes, which are used to show our capability in manifesting missing information to see illusions of other shapes [1] – mostly in the form of partial silhouettes. We considered this technique a component-induced form of pareidolia, processed without our awareness [28]. This concept bears similarity to a scene perception study where simple line drawings suffice to show our capability in categorising natural scenes of beaches, city streets, forests, highways, mountains, and offices [38].

In the pattern recognition and computer vision literature we could find only a few relevant studies that were aimed at detecting faces by face components; however, they were all specific to finding real human faces. For simulating facial pareidolia we therefore developed a new component face detector, which is capable of detecting many different face types. A particular advantage of this technique, when compared to a holistic method, is that the face components are explicitly isolated. This was achieved by using heuristics to define a suitable metric for evaluating a range of candidate faces.

To complement the component face detector we trained several one-class classifiers using scale-invariant feature vectors derived from the face components (i.e. eyes and mouth) to capture the geometry, size and shape of the face components. The feature vectors were then evaluated using four one-class classifiers (a combination of ν and SVDD SVM type with the Gaussian and the polynomial kernel). Our assessment of the output of the classifiers showed that in all cases the $(\nu, poly)$ one-class

classifier performed best, followed by the (ν, rbf) one-class classifier.

Further experiments were conducted to show how the feature vectors (having only seven dimensions) are capable of distinguishing three-shape faces of two eyes with a mouth from those made up of two brows with a mouth; however, these feature vectors were not adequate enough to distinguish expressions. To create the component expression classifiers we sliced the dataset of cartoon faces into face components – brows, eyes and mouth. This was possible because of the feature points from our face model that we manually inserted to human faces in the dataset. Our evaluation of the component expression classifiers shows that the classification of the mouth expression is almost equivalent to that of the whole face, while the eyes and brows were harder to determine across the seven discrete expressions of emotion.

The described approach and the resulting system are planned to be further extended in future research. This includes a more streamlined and automated implementation that can be released to researchers in other fields. It appears that the experience of pareidolia can have a large subjective variance in humans. Therefore training data selection and tuning of the system of our pilot study could be extended by large scale crowd sourcing experiments or the application of neurophysiological tests such as EEG to a group of human subjects. The system may then become applicable for further interdisciplinary research on pareidolia effects in psychology and could potentially be used for professional applications in ar-

chitecture and design.

Future research on this topic will also have to address more fundamental questions about the nature of emotions and affect in humans, how they can be modelled and how they may interact with the perception of form [30, 37]. While the present study focused on patterns associated with facial expressions of emotion and their early unconscious perception, future research may also address gestures and multimodal patterns associated with the communication of emotions as well as their cognitive processing [12, 24, 32].

ACKNOWLEDGMENTS

The authors would like to thank a range of people who supported or contributed to this study. These include flickr contributors for offering some of the displayed images online, the anonymous reviewers for making many useful suggestions, our English language editors for checking the manuscript, the postgraduate team of UoN for supporting the first author's PhD and our collaborators in neuroscience and architecture for valuable discussions and support.

REFERENCES

- [1] Ted S. Altschuler, Sophie Molholm, Natalie N. Russo, Adam C. Snyder, Alice B. Brandwein, Daniella Blanco, and John J. Foxe. Early electrophysiological indices of illusory contour processing within the lateral occipital complex are virtually impervious to manipulations of illusion strength. *NeuroImage*, 59(4):4074–4085, 2012.
- [2] Lindsay R. Arcurio, Jason M. Gold, and Thomas W. James. The response of face-selective cortex with single face parts and part combinations. *Neuropsychologia*, 50(10):2454–2459, 2012.
- [3] P.N. Belhumeur, D.W. Jacobs, D.J. Kriegman, and N. Kumar. Localizing parts of faces using a consensus of exemplars. In *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 545–552, 2011.
- [4] Modesto Castrillón-Santana, Daniel Hernández-Sosa, and Javier Lorenzo-Navarro. Combining face and facial feature detectors for face detection performance improvement. In Luis Alvarez, Marta Mejail, Luis Gomez, and Julio Jacobo, editors, *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, volume 7441 of *Lecture Notes in Computer Science*, pages 82–89. Springer Berlin Heidelberg, 2012.
- [5] Marco Catani and Michel Thiebaut de Schotten. *Atlas of Human Brain Connections*. Oxford University Press, USA, 2012.
- [6] Stephan K. Chalup, Kenny Hong, and Michael J. Ostwald. A face-house paradigm for architectural scene analysis. In Richard Chbeir, Youakim Badr, Ajith Abraham, Dominique Laurent, and Fernando Ferri, editors, *CSTST 2008: Proceedings of The Fifth International Conference on Soft Computing as Transdisciplinary Science and Technology*, pages 397–403. ACM, 2008.
- [7] Stephan K. Chalup, Kenny Hong, and Michael J. Ostwald. Simulating pareidolia of faces for architectural image analysis. *International Journal of Computer Information Systems and Industrial Management Applications (IJCISIM)*, 2:262–278, 2010.
- [8] Chih-Chung Chang and Chih-Jen Lin. Training ν -support vector classifiers: Theory and algorithms. *Neural Computation*, 13:2119–2147, 2001.
- [9] M. Dantone, J. Gall, G. Fanelli, and L. Van Gool. Real-time facial feature detection using conditional regression forests. In *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2578–2585, 2012.
- [10] B.A. Efraty, M. Papadakis, A. Proffit, S. Shah, and I.A. Kakadiaris. Facial component-landmark detection. In *2011 IEEE International Conference on Automatic Face Gesture Recognition and Workshops (FG 2011)*, pages 278–285, 2011.
- [11] R. Epstein, A. Harris, D. Stanley, and N. Kanwisher. The parahippocampal place area: recognition, navigation, or encoding? *Neuron*, 23:115–125, 2000.
- [12] Donald Glowinski, Nele Dael, Antonio Camurri, Gualtiero Volpe, Marcello Mortillaro, and Klaus Scherer. Toward a minimal representation of affective gestures. *IEEE Transactions on Affective Computing*, 2(2):106–118, April–June 2011.
- [13] L. Goldmann, U.J. Monich, and T. Sikora. Components and their topology for robust face detection in the presence of partial occlusions. *IEEE Transactions on Information Forensics and Security*, 2(3):559–569, 2007.
- [14] Nouchine Hadjikhani, Kestutis Kveraga, Paulami Naik, and Seppo P. Ahlfors. Early (n170) activation of face-specific cortex by face-like objects. *Neuroreport*, 20(4):403–407, 2009.
- [15] Kenny Hong. *Scene Perception using Machine Pareidolia of Facial Expressions*. PhD thesis, School of Electrical Engineering and Computer Science, The University of Newcastle, NSW, Australia, 2013.
- [16] Kenny Hong, Stephan K. Chalup, and Robert A.R. King. A component based approach improves classification of discrete facial expressions over a holistic approach. In *2010 International Joint Conference on Neural Networks (IJCNN 2010)*, pages 90–97. IEEE, 2010.
- [17] Kenny Hong, Stephan K. Chalup, and Robert A.R. King. An experimental evaluation of pairwise adaptive support vector machines. In *The 2012 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2012.
- [18] Kenny Hong, Stephan K. Chalup, and Robert A.R. King. A component based approach for classifying the seven universal facial expressions of emotion. In *IEEE Symposium on Computational Intelligence for Creativity and Affective Computing (CICAC)*, Singapore 16–19 April 2013, pages 1–8, 2013.
- [19] Kenny Hong, Stephan K. Chalup, Robert A.R. King, and Michael J. Ostwald. Scene perception using pareidolia of faces and expressions of emotion. In *IEEE Symposium on Computational Intelligence for Creativity and Affective Computing (CICAC)*, Singapore 16–19 April 2013, pages 79–86, 2013.
- [20] Yuchi Huang, Qingshan Liu, and D.N. Metaxas. A component-based framework for generalized face alignment. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 41(1):287–298, 2011.
- [21] Thomas W. James, Eunji Huh, and Sunah Kim. Temporal and spatial integration of face, object, and scene features in occipito-temporal cortex. *Brain and Cognition*, 74(2):112–122, 2010.
- [22] Mark H. Johnson. Subcortical face processing. *Nature Reviews Neuroscience*, 6:766–774, 2005.
- [23] N. Kanwisher, J. McDermott, and M.M. Chun. The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17(11):4302–4311, 1997.
- [24] Andrea Kleinsmith and Nadia Bianchi-Berthouze. Affective body expression perception and recognition: A survey. *IEEE Transactions on Affective Computing*, 4(1):15–33, January–March 2013.
- [25] Chiunhsun Lin and Kuo-Chin Fan. Triangle-based approach to the detection of human face. *Pattern Recognition*, 34(6):1271–1284, 2001.
- [26] John Nolte. *The Human Brain: An Introduction to Its Functional Anatomy 5th edition*. Mosby, 2002.
- [27] Paul Ekman Group LLC. JACFEE and JACNEUF dataset <http://www.humintell.com/for-use-in-research/>.
- [28] Michael A. Pitts, Antígona Martínez, and Steven A. Hillyard. Visual processing of contour patterns under conditions of inattention blindness. *Journal of Cognitive Neuroscience*, 24(2):287–303, 2012.
- [29] V. Rapp, T. Senechal, K. Bailly, and L. Prevost. Multiple kernel learning svm and statistical validation for facial landmark detection. In *2011 IEEE International Conference on Automatic Face Gesture Recognition and Workshops (FG 2011)*, pages 265–271, 2011.
- [30] Rainer Reisenzein, Eva Hudlicka, Mehdi Dastani, Jonathan Gratch, Koen Hindriks, Emiliano Lorini, and John-Jules Ch. Meyer. Computational modeling of emotion: Toward improving the inter- and intradisciplinary exchange. *IEEE Transactions on Affective Computing*, 4(3):246–266, July–September 2013.
- [31] D. Sabatinelli, P.J. Lang, M.M. Bradley, V.D. Costa, and A. Keil. The timing of emotional discrimination in human amygdala and ventral visual cortex. *Journal of Neuroscience*, 29:14864–14868, 2009.
- [32] Klaus R. Scherer and Heiner Ellgring. Multimodal expression of emotion: Affect programs or componential appraisal patterns?

- Emotion*, 7(1):158–171, 2007.
- [33] Bernhard Schölkopf, John C. Platt, John C. Shawe-Taylor, Alex J. Smola, and Robert C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Computing*, 13(7):1443–1471, 2001.
 - [34] Marina Shpaner, Sophie Molholm, Emma-Jane Forde, and John J. Foxe. Disambiguating the roles of area V1 and the lateral occipital complex (loc) in contour integration. *NeuroImage*, 69:146–156, April 2013.
 - [35] Satoshi Suzuki and Keiichi Abe. Topological structural analysis of digitized binary images by border following. *Computer Vision, Graphics, and Image Processing*, 30(1):32–46, 1985.
 - [36] Paul Viola and Michael Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.
 - [37] Peter Walla and Jaak Panksepp. Neuroimaging helps to clarify brain affective processing without necessarily clarifying emotions. In Kostas N. Fountas, editor, *Novel Frontiers of Advanced Neuroimaging*, volume 13, chapter 6, pages 93–118. InTech, 2013. doi.org/10.5772/51761.
 - [38] Dirk B. Walther, Barry Chai, Eamon Caddigan, Diane M. Beck, and Li Fei-Fei. Simple line drawings suffice for functional mri decoding of natural scene categories. *Proceedings of the National Academy of Sciences*, 108(23):9661–9666, 2011.
 - [39] Jiedong Zhang, Xiaobai Li, Yiyang Song, and Jia Liu. The fusiform face area is engaged in holistic, not parts-based, representation of faces. *PLoS ONE*, 7(7):e40390, 2012.

Kenny Hong has completed his PhD in the area of machine learning and image analysis at the University of Newcastle, Australia, in 2013. He is now working for a company that specialises on medical image analysis.

Stephan Chalup is an associate professor in computer science and software engineering at the University of Newcastle where he leads the Newcastle Robotics Laboratory. His research interests include machine learning, dimensionality reduction, timeseries analysis, autonomous robots, computer vision, affective computing and companion robots.

Robert King is a lecturer in statistics in the School of Mathematical and Physical Sciences at the University of Newcastle. His research interests include statistical learning, statistical computing, statistics for computing, the study of distributions and ecological and social statistics.