Recognising Familiar Facial Features In Paintings Belonging to Separate Domains

Wilbert Tabone and Dylan Seychell

St. Martin's Institute of Higher Education
wtabone@stmartins.edu
dseychell@stmartins.edu

Abstract. We present a system[1] that detects faces in various paintings and subsequently recognise and points out any similarities that a certain face in one painting may have to another on a different artwork. The results would be ranked up according to similarity in a bid to produce an output that may assist art researchers to discover new links between different works which pertain to the same or different artist. Through various tests conducted, we have proved that our method was successful in exposing new links of similarity in various scenarios including cases where the human visual system failed to pinpoint any.

1 INTRODUCTION

Face detection and recognition systems have been developed for numerous applications which span across different areas of research. However we have noticed that there is a lack of published research which apply such techniques for such a purpose in the area of fine art. Our system allows a user to input a photograph or scan of a painting and receive a result that ranks other paintings according to a similar face detected within. There are cases where systems try to identify where a portrait has been featured across different mediums. Our research differentiates from this as it aims to find this face exclusively on the painted media, whilst also returning other similar faces. The latter would aid in establishing new links between the human models and the artists. We also aim to target museum displays with this system.

For the purpose of this research, we have chosen paintings from the Baroque period, specifically focusing on works by Francesco Zahra, a Maltese artist of that era. It is noted that there are several occurrences in Zahra's work in which the same or very similar faces are found on different artworks pertaining to the same or different character being portrayed [15]. This rendered the evaluation of such a domain, that is difficult to evaluate, possible. Furthermore, the results obtained were directly compared to the human visual system in order to assess the accuracy and success of the rankings.

[1] This work is based on a report submitted in partial fulfilment of the B.Sc. (Hons.) Creative Computing, offered by the University of London International Programmes.

The first part of this paper will discuss the research background, followed by the methods implemented in order to design an artefact that would aid in assessing the thesis explored. The last part will present the results and subsequently a discussion, followed by the conclusion.

2 RESEARCH

In this section, research areas that were drawn upon in order to develop our system are discussed. Research was spread over various areas such as classifiers, edge detection and face detection and recognition techniques. The first three areas were explored as preparation to the various classifiers and image similarity methods explored in the face detection and recognition section, since various combinations of the three are utilised.

2.1 Classifiers

A classifier can be expressed as a function that takes a value of a feature and subsequently predicts the class that that example belongs to [10]. In preparation for research on classifiers utilised for face detection and recognition, linear classifiers were researched. These are defined by [5] as being classifiers that use feature functions $\mathbf{f}(x) = (f_1(x), \ldots, f_m(x))$ and feature weights $\mathbf{w} = (w_1, \ldots, w_m)$ to assign $x \in X$ to class $c(x) = \text{sign}(\mathbf{w}.\ \mathbf{f}(x))$, where $\text{sign}(y) = +1$ if $y>0$ and -1 if $y<0$. In the preceding notation, $x \in X$ defines an example whilst the classifier is defined by the mapping of the example to a binary class $c(x) \in$ -1, 1.

2.2 Edge Detection

Some of the most prominent edge detection techniques were researched since edges extracted from an image may facilitate in the detection and classification of faces. The general edge detection method is given as a system that smoothens the image and subsequently enhances and thresholds the edges to produce a binary edge map. There exist two categories of detectors: first order and second order, where each refers to the application of first or second order differentiation [9].

The Sobel edge detector calculates the magnitude and direction of the gradients in an image by the utilisation of the Euclidean distance for magnitude and the *arctan* of the angle for the orientation (direction) [13]. On the other hand, the Canny edge detector utilises five steps [6] that enable it to produce clearer results by following the criteria outlined by Canny in [4] which include having a low error rate, having well localised edge points and giving only one response to a single edge. The last method to be explored was the Laplacian second order edge detector. In order to alleviate the problem of its sensitivity to noise, the detector employs Gaussian smoothing beforehand [16]. This detector highlights the zero crossing of the image (the points at which its second derivative equates to zero) in order to produce the edge map [6]. In conclusion it is noted that the

Canny edge detector handles noisy images better than its counterparts discussed here. This is ideal for paintings as they are intrinsically noisy.

2.3 Face Detection

Various image-pre-processing methods such as histogram equalisation were explored as precedents to face detection techniques. It was decided that from the different detector classes, the appearance-based methods would be the most suitable for the system since these probabilistic-methods classify a random variable x as a *face* or *non-face* [19]. This is basically what is wanted in the context of what is to achieved.

The chosen detector from this category was the Viola-Jones Haar classifier [17]; this produces an integral image, which is an array that contains the sum of the pixel intensities of the pixel at location *(x,y)* and its neighbours through which it eliminates false positives [7]. Furthermore, through the use of Haar features (wavelets), rectangular groups of pixels are formed based on the intensity value and hence the facial features are found [8, 12].

2.4 Face Recognition

Although research was conduced on the eigenfaces approach (principal component analysis), fisherfaces approach (linear discriminant analysis), support vector machines (SVM) and Gaussian naive Bayes (GNB), it was later discovered that these were nonoptimal for the purpose of what was to be achieved, due to the limited amount of training examples that paintings provide. Therefore, a different approach was taken and image similarity measurements were researched instead, as similarities between faces can still be assessed through such methods. Even though some of these methods were created to be applied on video data, they were still utilised as internally these are operating on image frames.

The first method to be explored was the Wilkie, Stonham and Aleksander's Recognition Device (WiSARD) which is a collection of RAM-discriminates. In general, the WiSARD classifier has a collection of RAM units (neurons) each of which are trained on a particular pattern [2]. When it receives an input pattern, each neuron outputs a 0 if there is no match or 1 if there is a match in the pattern area assigned to it. The WiSARD sums up the result to produce the total value, which may be expressed as a percentage of recognition [7, 3]. The architecture of such a system is presented in Figure 1.

Two further methods explored were the PSNR (peak signal-to-noise ratio) and SSIM (structural similarity index) measurements which may be used as a crosscheck for the WiSARD. While PSNR measures the change in the signal between two frames, SSIM correlates to how humans visually perceive a scene and therefore it measures similarity on the basis of luminance and contrast amongst others [18].

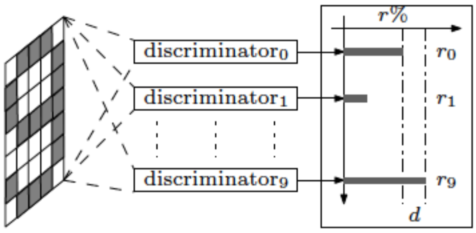


Fig. 1. A representation of a 10 RAM- descriminator WiSARD (Source: [1]).

3 METHODS

We chose OpenCV[2] as the programming framework.

Detection
In order to prepare the paintings for detection, transformation from the 3D to 2D colour space was performed, followed by histogram equalisation. Choosing the detector involved various experiments that first identified the most optimal classifier and secondly enhanced the latter further. For the purposes of this work, a pre-trained classifier was chosen over training.

The two pre-trained cascades provided by OpenCV: LBP (Local Binary Patterns) and Haar (Viola-Jones detector) were tested. Each of these detectors were executed on a test set made up of eleven paintings and one photograph. While the LBP detector was found to be more efficient in execution time, with a difference of 22.92 ms, the Haar classifier had a higher success rate (by 4%) together with 100% precision and was therefore deemed the ideal cascade to utilise for the face detector. In order to find the best possible way to enhance the success rate of the Haar detector, various tests were conducted on the Haar detector with different minimum neighbour parameter values, switching between having histogram equalisation or not and utilising different edge detection techniques in order to enhance the edges (by superimposing the produced edge image onto the original).

From the conducted experiments, it was discovered that the most efficient method is the Haar detector using histogram equalisation, a minimum neighbour parameter value of one (how many detections in a particular area is allowed) and utilising no edge detector whatsoever. It was noted that when the recommended default value of three (for min. neighbour) is utilised, the Precision is excellent (100%), however the Recall rate is not satisfactory enough. Furthermore, it was noted that while histogram equalisation improves the results, edge detection does not. These results are tabulated below in Figure 2.

[2] OpenCV: http://opencv.org/ (programming library for real-time computer vision applications)

Comparison of Results

HistEq	minNeighbour	Edge	Recall	Precision	Failure Rate
Yes	1	N/A	59%	97%	41%
Yes	2	N/A	53%	87%	47%
Yes	3	N/A	51%	100%	49%
No	1	N/A	55%	87%	45%
No	2	N/A	53%	93%	47%
No	3	N/A	51%	96%	49%
Yes	1	Canny	46%	82%	53%
No	1	Canny	49%	92%	51%
Yes	1	Sobel	29%	74%	71%
No	1	Sobel	33%	94%	67%
Yes	1	Laplacian	29%	88%	71%
No	1	Laplacian	37%	86%	63%

Fig. 2. Comparison of Haar experiment results.

Now that the best combination of parameters for the Haar classifier was found, the next problem to be tackled was the detection of faces having a slightly rotated pose in order to aim for a rotation-robust face detection process.

Head pose

One of the various challenges attributed to face detectors is the head pose of the subject [20]. In order to aim for a rotation-robust detection process, an automated rotation system that attempted to detect a face at each ten degree interval was created. This resulted in a large number of false positives being recorded. Therefore, since the focus of the conducted research was on the recognition of faces, a simpler system that utilises basic coordinate geometry was developed.

In this system, a line is first constructed between the two eyes and another straight horizontal line is constructed from the topmost eye to the position where the bottommost eye should be if the face was properly aligned. This is demonstrated in Figure 3.

Both gradients, m_1 and m_2 are calculated and subsequently, the angle θ between the two lines is found. The image is rotated by θ following complimentary angle laws. This process greatly improved the number of detections and TPs in paintings containing tilted faces, as evidenced by Figure 4.

Recognition:

As the commodity of having several images of the same face at different angles is not provided by paintings, it was not possible to train any classifiers and therefore there was no need for the vector space model to be utilised. Image similarity measurements were employed instead. The WiSARD classifier, which is applied on bi-level images was utilised as the main discriminator. There was also a need for a crosscheck since the system needed to be automated. Therefore, the list of

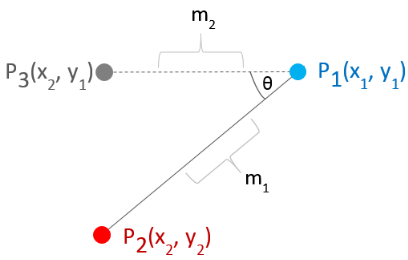


Fig. 3. Mathematical principle behind the head pose rectification solution.

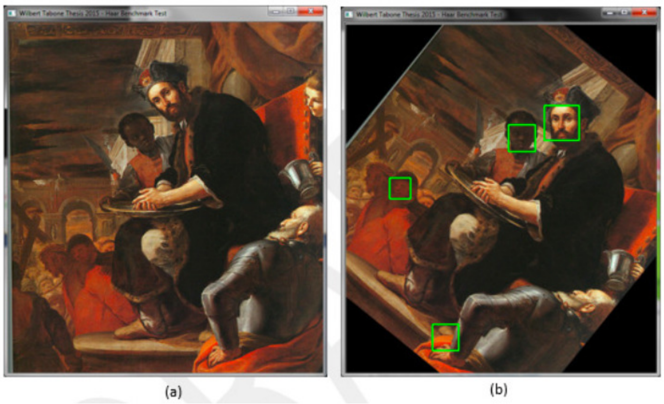


Fig. 4. Improvement of results once head is correctly aligned. (a): Detection without alignment; (b): detection after alignment.

outputted similar faces would have to be filtered in order to only allow the most accurate results to be displayed. Both SSIM and PSNR was tested, together with the WiSARD classifier on a second test set containing similar photographs and paintings. Since SSIM is based on the human visual system, it gave the most positive results when applied to both the paintings and photographs.

Therefore, SSIM was selected as the crosscheck for the WiSARD in the final application. Further experiments concluded that as the brightness of the input images increases, a better WiSARD recognition percentage is achieved due to a higher quality binary image being produced. On the other hand, it was noticed from a further test that the further a detected face image is cropped, the lower the WiSARD measurement and SSIM values become. From this observation, we decided that there would be no further treatment to the images produced by the detector once created.

Application

Through the use of the partitioning software development approach [11], all test applications were amalgamated into a final UI application that would prove the concept of this research.

In this application, a set of pre-selected images are presented as part of the test set selection module and the query window. First the user constructs a test set made up of six images to be used by the WiSARD and SSIM and consequently the system allows the user to select one of twelve query images. Each of the query images may be rotated, if necessary, before being processed for ranking. Once the detection is completed, each detection area is processed through the WiSARD classifier which determines, through a threshold of 20.0, which detected faces will go on for ranking. Subsequently, the SSIM, which is given by Equation 1, is calculated between the detected face and each previously selected test image.

$$S\left(x,y\right)=f\left(\mathbf{l}\left(x,y\right),\mathbf{c}\left(x,y\right),\mathbf{s}\left(x,y\right)\right). \tag{1}$$

Where, $\boldsymbol{l}$ is luminance, $\boldsymbol{c}$ is the contrast and $\boldsymbol{s}$ is the structure.

A ranking window containing the WiSARD value, together with the three most similar faces to the detected face (containing the highest SSIM value) are displayed. Next to each face, a link to the source paining is presented. This is shown below in Figure 5.

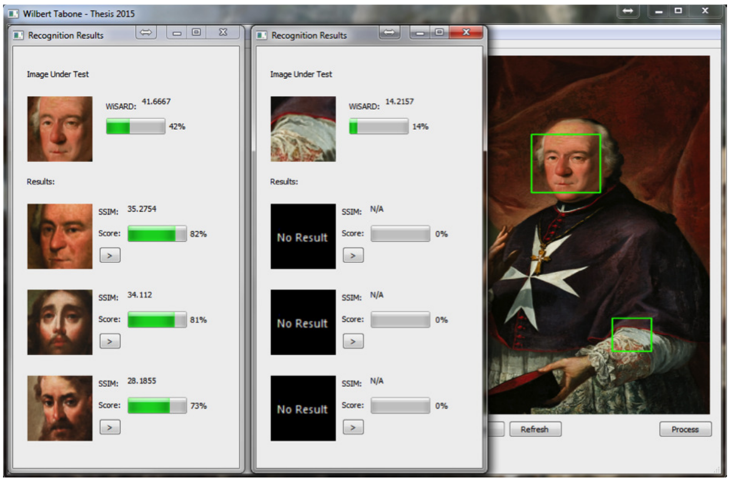


Fig. 5. The rank window for each detection containing the similarity results. Note how the false postitive returned no result due to the thresholds put in place.

As a way to compare the results that were obtained with a fixed questionnaire, a score system was created. The utilised formula returns a score $\mathbf{S}(\%)$ that is based on both the WiSARD and SSIM values, which are both normalised and subsequently the resultant value is divided by two (for the average) and multiplied by 100 for the percentage value as per Equation 2:

$$S = \left(\frac{\left(\frac{WiSARD}{55}\right) + \left(\frac{SSIM}{40}\right)}{2} \right) \times 100 \tag{2}$$

The values for normalisation, i.e. 55 and 40 were obtained through results of experiments conducted to determine the best possible combination.

4 RESULTS

Several results given by the detector and ranking windows were obtained under different circumstances. Overall, the detector gave positive results, with only a minimal number of FPs. Moreover, positive results were obtained by the WiSARD and SSIM recognition system (ranker). In order to better evaluate the system, a fixed questionnaire that aimed to collect both quantitative and qualitative data was devised. The quantitative sections of the questionnaire assessed the human visual system (represented by the respondents) against various components of the developed system described in the previous section. On the other hand, the qualitative data sections required text-input answers from the public.

The majority of the hundred and ten respondents belong to the 19-25 age bracket, with females being the major gender group. Following the demographics, the respondents were asked about their interest in art; a question that returned positive results as 51% reported that they are interested in fine art. The next part, which assessed the detector, required the respondents to point out the amount of non-occluded and prominent faces in the five five paintings presented to them. The application managed to correctly process the paintings with no FPs in four of the five paintings; a result that agreed with the majority of the respondents.

A further section in the questionnaire compared the scores obtained from the rank result window (represented as a progress bar) to the score chosen by respondents on a Likert scale from 1-10, where one denotes minimal similarity and ten denotes maximal similarity. For evaluation purposes, the number chosen by the majority of the respondents was multiplied by one hundred in order to obtain a percentage value that would be directly compared with the percentage score of the application. A sample of the collected results is presented in Figures 6 and 7.

The first part of this section contained side-by-side comparisons of query faces to faces from the test set, which were deemed similar by us through our personal analysis and were ranked as the topmost result by the application. On the other hand, in the second part of this section, the second and third ranked results of the software were presented next to the query face. These latter results are interesting as although they are deemed similar by the software application, they would not look similar on immediate examination by the observers, however prolonged analysis would prove otherwise.

From the tests in the first part, where the first result of the rank window was presented to the respondents, there was agreement between the system and

Rank Window Results	
SSIM:	35.2754
Score:	82%

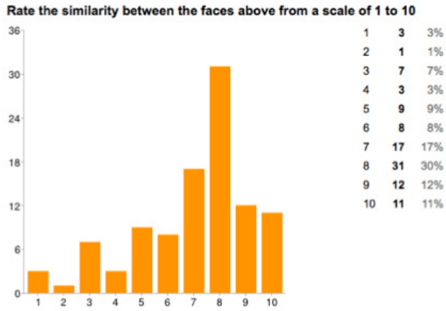
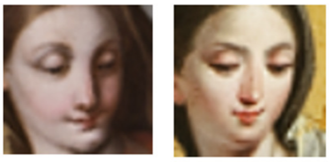


Fig. 6. A sample from the results. Notice how the system developed has returned a value of 82% whilst the HVS (majority of respondents) chose a score of 80%.



Rank Window Results	
SSIM:	24.5925
Score:	71%

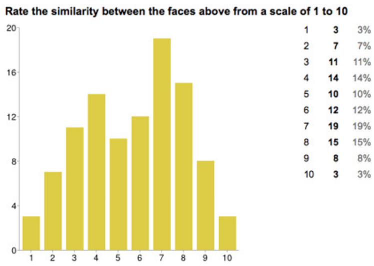


Fig. 7. A further sample from the results. Although a more subjective comparison is presented in this case, the respondents reported a 70% similarity score whilst the application returned a score of 71%.

the scores of the respondents ($\pm$ 10%) in five from the six eligible side-by-side comparisons. Hence, an 83 % success rate was obtained. On the other hand in the second part, these disagreements were more frequent as only in four from thirteen tests did the respondents and ranker agree on an equal score; yet a close score was observed in the majority of results. These disagreements were expected in this case, since as previously mentioned, subjectivity came into place in non-obvious similarity exercises.

Once this data was collected, we presented our findings, together with the developed application to art researchers in a semi-structured interview. With regards to the application and results obtained, the interviewees had positive comments. It was asserted that as the current application stands, it would be mainly useful for research. Furthermore, our application will provide researchers in the area of the history of art a tool that will perform multiple tests that would aid them in gauging if the same human model was used in works of different artists. This is something that very few researchers have done as usually the study is on models that featured in multiple works of the same artist.

The curator of a National fine arts museum stated that in such a setting, the application may be used to link works from such museums to those found in churches across the country/ world (e.g: commissioned by the Church during the Baroque period). Another interesting application would be to genealogically connect pieces in portrait galleries through the recognition system presented here.

Overall, the interviewees were intrigued by the application, asserting that while they see it predominantly as a research tool, there could be various variants of it which can be introduced in an art museum/gallery setting to enhance the experience by the introduction of interactive elements through the use of, for example, a mobile version of the application.

5 DISCUSSION

The system was successful in detecting facial features in the input paintings whilst also correctly utilising its primary recognition system: the WiSARD weightless neural network in order to assess if any similar faces exist in the test set and subsequently utilising the SSIM measurement to pinpoint and rank the actual similar faces by their structural similarity value. We believe that the pre-trained Viola Jones Haar classifier that was utilised provided overall satisfactory results for what we were aiming for, with minimal resource use and efficient execution time.

Moreover, from our analysis we deem the recognition results obtained by the system as satisfactory. Although there were some small discrepancies between the HVS and the system when comparing the second and third rank results, we believe that since the results were still in the region of the application's score and since a high number of agreeable results between the HVS and rank window primary results were attained, it can be concluded that the ranking system is working as expected as a high enough success rate was achieved.

The shortcomings of the system at its current state are that it requires an input of a scan or quality photograph of the painting to be queried in order to produce satisfactory results. Moreover, the current facial rotation system is not automated due to the problem outlined above in Section 3. This is not presently an issue due to the nature of the application being created to run on desktop computers. However, if the system is to be exported to a mobile device, an alternative solution would be utilised.

5.1 Enhancements

From a further question in the questionnaire, it was discovered that people find it easier to discover any similarities between faces when presented with the whole painting. This fact may be exploited further so that more information may be extracted from the painting making it ripe to implement multiple information extractors such as those presented in the work of [14]. Therefore, a system which recognises familiar facial features, links faces to artists and influence between artists can be produced. The final product would be a network that links faces, models and artists in the same or different time periods.

6 CONCLUSION

In conclusion, we believe that such a system is possible to be applied to a wider range of different artworks and media. In the future, we plan to adapt the system in order to support different art forms and pinpoint similarities in both facial and artistic features. Moreover, a system that creates clusters of paintings according to the artist and the characters portrayed may be implemented. These clusters would contain paintings from various sources across the globe, such as museums and private collections. Each item in the collection would have embedded meta-data that would allow the user to search by artist or discover new ones upon examining the meta-data included with the similarity results.

We strongly believe that such a system would help art researchers in amassing collections of links in similarity between works of the same or different artists. This will aid in building upon or commencing new research in their respective areas.

References

1. Aleksander, I., De Gregorio, M., França, F.M.G., Lima, P.M.V., Morton, H.: A brief introduction to weightless neural systems. In: ESANN. Citeseer (2009)
2. Aleksander, I., Morton, H.: An introduction to neural computing, vol. 3. Chapman & Hall London (1990)
3. Beham, M.P., Roomi, S.M.M.: Face recognition using appearance based approach: A literature survey. In: Proceedings of International Conference & Workshop on Recent Trends in Technology, Mumbai, Maharashtra, India. pp. 24–25 (2012)
4. Canny, J.: A computational approach to edge detection. Pattern Analysis and Machine Intelligence, IEEE Transactions on (6), 679–698 (1986)
5. Johnson, M.: A brief introduction to kernel classifiers (2009)
6. Maini, R., Aggarwal, H.: Study and comparison of various image edge detection techniques. International Journal of Image Processing (IJIP) 3(1), 1–11 (2009)
7. Medioni, G., Kang, S.B.: Emerging topics in computer vision. Prentice Hall PTR (2004)
8. Nilsson, M.: Face detection. Presentation by the Mathematical Imaging Group, Centre for Math ematical Sciences, Lund University (2014)
9. Nixon, M.: Feature extraction & image processing. Academic Press (2008)
10. Pereira, F., Mitchell, T., Botvinick, M.: Machine learning classifiers and fmri: a tutorial overview. Neuroimage 45(1), S199–S209 (2009)
11. Pressman, R.S.: Software engineering: a practitioners approach. McGrow-Hill International Edition (2005)
12. Rabbani, M., Chellappan, C.: A different approach to appearance–based statistical method for face recognition using median. International Journal of Computer Science and Network Security 7(4), 262–267 (2007)
13. Robert Fisher, Simon Perkins, A.W.E.W.: Sobel edge detector (2004), accessed: 2014-09-08

14. Saleh, B., Abe, K., Arora, R.S., Elgammal, A.M.: Toward automated discovery of artistic influence. CoRR abs/1408.3218 (2014)
15. Sciberras, K.: Francesco Zahra: His life and art in mid-18th century Malta 1710-1773. Midsea Books (2010)
16. Shrivakshan, G., Chandrasekar, C.: A comparison of various edge detection techniques used in image processing. IJCSI International Journal of Computer Science Issues 9(5), 269–276 (2012)
17. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. vol. 1, pp. I–511. IEEE (2001)
18. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. Image Processing, IEEE Transactions on 13(4), 600–612 (2004)
19. Yang, M.H., Ahuja, N.: Face detection and gesture recognition for human-computer interaction, vol. 1. Springer (2001)
20. Yang, M.H., Kriegman, D., Ahuja, N.: Detecting faces in images: A survey. Pattern Analysis and Machine Intelligence, IEEE Transactions on 24(1), 34–58 (2002)