A COMPARISON OF SEARCH SPACES AND EVOLUTIONARY OPERATORS IN FACIAL COMPOSITE CONSTRUCTION

Joseph James Mist
School of Physical Sciencese
University of Kent, Canterbury, United Kingdom
jm441@kent.ac.uk

Stuart James Gibson
School of Physical Sciencese
University of Kent, Canterbury, United Kingdom
s.j.gibson@kent.ac.uk

Christopher John Solomon
School of Physical Sciencese
University of Kent, Canterbury, United Kingdom
c.j.solomon@kent.ac.uk

Abstract In this paper three experiments concerning the use of interactive evolutionary algorithms in the creation of facial composites are reported. A reduced dimension human evaluation based search space is created from a larger search space using a pairwise face comparison task. The human reduced search space is used in the comparison of two mutation operators and two recombination operators. Finally, three search spaces are compared: large, human reduced, and a mathematically reduced search space. No statistically significant differences are found between the performances of the operators or the search spaces.

Keywords: Facial composites, Interactive evolutionary algorithm.

1. Introduction

An unknown perpetrator may be seen committing a crime by one or more people. In these circumstances it is often useful to create a pictorial likeness of the perpetrator's face based on an eyewitness's description. Such a likeness, known as a *facial composite*, may subsequently be used in a criminal investigation for the purpose of locating a suspect. The traditional method for facial composite construction involves a witness selecting individual facial features from a catalogue or database. A composite operator then assembles the selected features to form a face image. However, psychological research has shown that humans recognise faces not by their individual components but as whole objects [2, 8]. It is also known that people are better able to recognise faces than they can recall and describe them. Accordingly, a holistic method for facial composite construction has been developed that uses the cognitive Gestalt processes involved in face recognition. Two commercial systems based on these principles were developed in the early 2000s; EvoFIT [4] and EFIT-V (originally called EigenFIT) [5]. EFIT-V is now used frequently by the majority of police constabularies in the UK and by law enforcement agencies in many other countries.

The holistic approach requires a multidimensional search space or, more appropriately, a *face-space* in which an approximate likeness to any face can be represented by single point. The location of this point with respect to the origin of the face-space may be encoded as a string of coordinates. This is achieved through the use of a *face model*.

Here, and in our related previous work, we use a face model [1] that is constructed by processing a set of training images and determining their principal components (PCs) (see Section 2.1). The PCs are typically ordered according to the amount of image variance in the training set they explain. Hence, the first component is numerically more important than the second component which is more important than the third etc. Due to the imperfect nature of human face recognition, it is very likely that the required face-space can be constructed using a relatively small number of PCs and the other PCs can be discarded with no perceptible deterioration in performance. A distinction should also be made between an ordering of PCs according to numerical importance and an ordering according to perceptual importance. Using human evaluation to select n PCs may provide an n-dimensional face-space which accounts for more perceptual variation than simply using the first n PCs returned by a principal components analysis as calculated by computer software. This idea is investigated in Section 4.

In simple terms, a facial composite is constructed by searching the face-space for an acceptable facial likeness. The search is achieved by an iterative process whereby faces, generated by the model, are assessed by the witness according to their similarity to the perpetrator. This iterative process immediately suggests the use of an *interactive evolutionary algorithm* (IEA).

IEAs differ from evolutionary algorithms (EAs) in one major respect: human evaluation replaces the fitness function. The use of human evaluation places a number of limitations on the use of IEAs which are generally not present in EAs. The most obvious two effects are that the total number of fitness evaluations that can be performed by the user and the number of possible fitness values it is feasible to assign to an individual are both limited [9]. The IEA used in the second and third experiments of this paper is detailed in Section 2.2.

Intuitively, selection of an appropriate IEA, associated operators, and the values of any associated parameters may have an effect on the composite creation process. Real valued interactive genetic algorithms (IGAs) are used in both EvoFIT and EFIT-V. Design decisions such as the population size, the mutation rates, and the use of elitism were made with the aid of virtual users that attempt to model how human users evaluate individuals in a generation. Very little work has been done to compare the performances of different IEAs for use in the creation of facial composites. A series of small experiments evaluating the performances of various nature-inspired metaheuristic algorithms have been conducted [6, 7]. The results indicate that the choice of algorithm has some effect on the recognition rates of the composites.

The focus of this paper is the comparison of search spaces and evolutionary operators in facial composite construction. In the first of three experiments reported in this paper, a 12-dimensional 'human reduced' face-space is constructed using human evaluation of the differences between pairs of faces from the 'large' 30-dimensional face-space. The second experiment compares the performances of two different mutation operators and two different recombination operators using a task which requires participants to create composites from memory. In the third experiment a 'mathematically reduced' face-space, in which only the first 12 PCs are used, is constructed. The performances of searches using the large, the human reduced and the mathematically reduced face-spaces are compared using the same composite creation task.

2. Theory

2.1 The Face Model

The set of photographs used in the training set to build the face models used in the experiments reported in this paper is composed of 27 males and 63 females of various ages. A number of points common to all of the photographs are landmarked. These common points are facial features such as the corners of the eyes, the bottom of the chin, and the outline of the eyebrows. The set of landmarks on a particular face form a face shape. Thus, there is one face shape for each face in the training set. Each face shape consists of 190 two-dimensional landmarks and thus the resulting shape model has 380 dimensions.

The mean face shape $\bar{\mathbf{s}}$ is found by aligning the face shapes using an iterative Procrastes alignment process. Principal components analysis (PCA) is used to reduce the 380-dimensional shape model to a smaller number of dimensions. Any face shape $\mathbf{s}$ can be approximated to $\widehat{\mathbf{s}}$ in the shape model using

$$\widehat{\mathbf{s}} = \mathbf{P}_s \mathbf{b}_s + \bar{\mathbf{s}} \tag{1}$$

where $\mathbf{P}_s$ are the PCs of the shape model ordered from most important (the PCs which account for the most variance in the data) to least important and $\mathbf{b}_s$ are parameters that determine how the shape PCs are combined to make the face shape.

In order to create the texture model, each photograph in the training set is partitioned using its landmark points and Delaunay triangulation. Piecewise affine transforms are used to map the texture information (the pixel values of the photographs in the training set) from each training photograph's face shape to the mean face shape to form normalised texture patterns. PCA is then used to find a texture model with fewer dimensions than that formed by the tens of thousands of pixels within each normalised texture pattern. As with the face shapes, any face texture $\mathbf{g}$ may be approximated using

$$\widehat{\mathbf{g}} = \mathbf{P}_g \mathbf{b}_g + \bar{\mathbf{g}}. \tag{2}$$

where $\mathbf{P}_g$ are the PCs of the face texture ordered from the most important to least important and $\mathbf{b}_s$ are parameters that determine how the texture PCs are combined to make the face texture. Finally, a face-model is created from the combined shape and texture models using PCA to further reduce the number of dimensions in the final face-space. Thus, the appearance model parameters, $\mathbf{c}$, of any face can be approximated to $\widehat{\mathbf{c}}$ using

$$\widehat{\mathbf{c}} = \mathbf{Q}^T \begin{bmatrix} w\mathbf{b}_s \\ \mathbf{b}_g \end{bmatrix} \equiv \mathbf{Q}^T \begin{bmatrix} w\mathbf{P}_s^T \left(\widehat{\mathbf{s}} - \bar{\mathbf{s}}\right) \\ \mathbf{P}_g^T \left(\widehat{\mathbf{g}} - \bar{\mathbf{g}}\right) \end{bmatrix} \tag{3}$$

where $\mathbf{Q}$ are the appearance PCs of the training set ordered from the most important to the least important and w is a weighting scale that scales the shape parameters such that equal significance is assigned to shape and texture.

New faces can be created by setting the values of an n-dimensional parameter vector $\mathbf{c}$ and performing the above process in reverse. Starting with the extraction of $\mathbf{b}$

$$\mathbf{b} = \sum_{i=1}^{n} \mathbf{q_i} c_i \tag{4}$$

where $\mathbf{q}_i$ is the i-th column of matrix $\mathbf{Q}$ in Equation 3. The shape and texture parameters $\mathbf{b}_s$ and $\mathbf{b}_g$ are extracted from $\mathbf{b}$ and are used in Equations 1 and 2 to find the shape parameters $\mathbf{s}$ and texture parameters $\mathbf{g}$. The pixel intensities in $\mathbf{g}$ are rearranged into a two-dimensional (or three-dimensional for colour images) array of pixels which then form an intermediate face image with mean face shape. Aspects of the edge of the face image which were due to the landmarking process had a dominant unwarranted effect on the perception of the face. To counter this effect the generated face texture was inserted and blended into a softened background. The resulting image was subsequently warped according to the shape parameters, $\mathbf{s}$, to form the final face image.

2.2 The IEA Used

The IEA used is a simple real valued IGA which we refer to as the *simple IGA*. The representation used is an n-dimensional real valued vector where n is the number of dimensions in the face-space.

In the simple IGA each individual in the following generation has two parents and each pair of parents produces only one child. Eight new individuals are needed to fill the following generation (as the best individual from the previous generation is carried through to the following generation). Thus, a mating pool of sixteen parents is required.

Stochastic universal sampling is used to select parents to go into the mating pool. The simple IGA follows Frowd's method [3] and allows only three levels of selection: preferred (best), selected, and not selected. When building the sampling wheel it was decided that all of the selected individuals are assigned equal sized wedges except the preferred individual which is assigned a double sized wedge.

Two recombination methods are used in the experiments reported in this paper: uniform crossover and arithmetic crossover, and two mutation methods are used: nonuniform mutation and Gaussian replacement.

In the implementation of uniform crossover used each gene's value in an offspring has an equal chance of coming from either parent. In the implementation of arithmetic crossover used each gene's value in an offspring is the mean of the values for that gene in the parents.

In the nonuniform mutation used in this paper the mutated gene value c_i' is given by $c_i' = c_i + \sigma_i \cdot m \cdot N(0,1)$ where σ_i is the standard deviation (SD) of the data on the i-th PC, m is the mutation factor set by the user on the interface, and $N(0,1)$ is a random number from the Gaussian distribution. Gaussian replacement is the name given in this paper to an analogous method to the uniform mutation operator. In uniform mutation, there is some probability p_m for each gene in an offspring's genotype that its value will be replaced by a uniformly distributed random value where $c_i, c_i' \in [\text{Lower limit}, \text{Upper limit}]$. The Gaussian replacement operator is similar except that c_i' is a random number taken from $N(0,1)$ and multiplied by the SD of the data on the i-th PC. c_i' has the further restriction that it is bounded by the hyperrectangle which designates the edge of the face space, that is $c_i, c_i' \in [-2.5, 2.5]$ SDs. The mutation probability is set by the mutation slider and is restricted to the range $[0, p_{\max}]$ where $p_{\max} = 5/$ (the dimensionality of the face space).

The population size is limited by three factors: the number of images that can be displayed on the screen simultaneously, the time required to create each image, and the cognitive burden placed on the user when comparing the images. After considering these three factors a population size of nine was chosen.

3. The User Interface for Experiments 2 and 3

A screenshot of the interface developed for the experiment is given in Figure 3. Every generation the participants would choose a preferred composite that best resembled the face they were trying to recreate and select it using the left mouse button. The participants also had the option of selecting any composites that they thought were also good by selecting them using the right mouse button. Anywhere from zero to eight composites could be selected with the right mouse button. A green border was placed around the composite the participant preferred, a yellow border for those composites the participant thought were also good, and a black border for those composites that were not selected. Once they were satisfied that they had selected the best match and any other matches they considered to be good, the participant would go to the next generation by pressing the 'Next' button. The preferred composite was carried forward into the next generation. The participants would continue the process until they thought they had successfully recreated the target face, or until they thought no further improvement was possible, by clicking on the 'Finish' button. It was observed during previous (unpublished) experiments that using a self adaptive step size in an IEA often resulted in the algorithm becoming stuck at a suboptimal solution. This problem was addressed by adding a slider which enabled the user to set the mutation value manually. However, many participants did not adjust the mutation slider. Thus, the mutation slider was decremented by 0.03 per generation by the software (the slider's range was $[0, 1]$). A 'back' button was included which enabled the participant to go back to the previous generation and make alternative selections or adjust the mutation slider if they were not satisfied with the current generation.

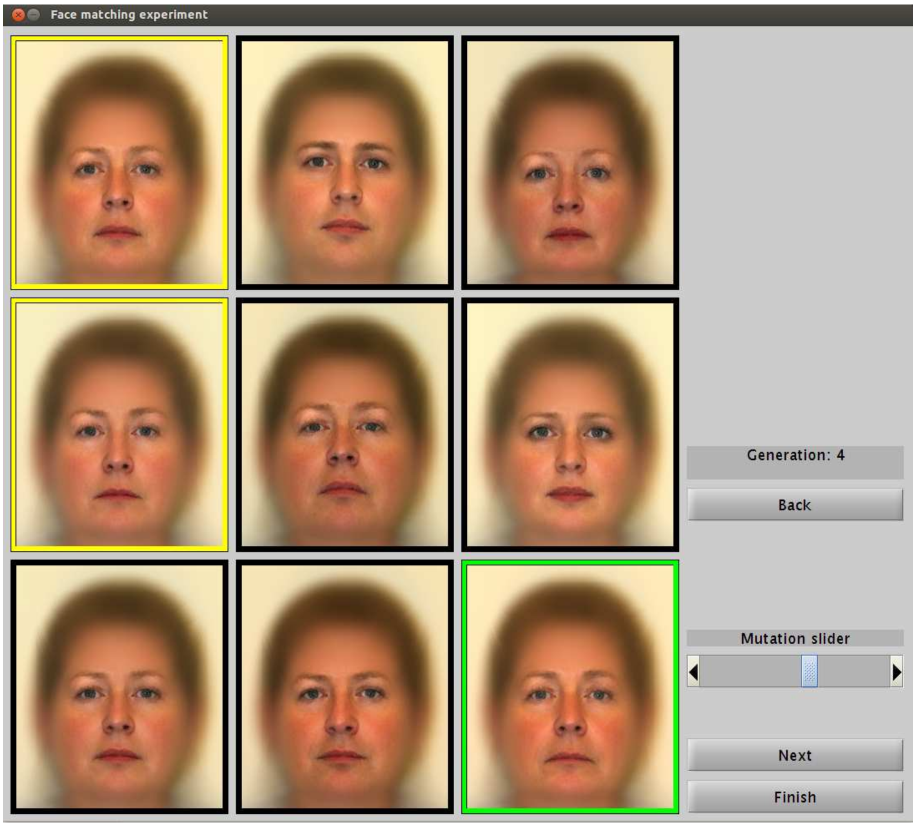

Figure 1. Screenshot of the interface for the facial composite tasks.

4. Experiment 1: Identifying the Most Perceptually Significant PCs

4.1 Method

In this experiment 32 participants were used to determine which 12 of the 30 PCs of the 30-dimensional (large) face-space are perceptually the most significant. Thirty pairs of faces were generated from the large PCA face-space. If a face's representation in the large face-space is given by the point $\mathbf{c} = [c_1, c_2, \ldots, c_i, \ldots, c_{30}]$, then each pair of points

$(\mathbf{c}_{+k}, \mathbf{c}_{-k})$ representing a pair of faces has coordinates defined by

$$c_{\pm i} = \begin{cases} \pm 3 \quad \text{SDs} & \text{if } i = k \\ 0 \quad \text{SDs} & \text{otherwise} \end{cases} \tag{5}$$

The faces were printed in their respective pairs on matt photographic paper. Each pair was 5.8 cm high by 10.2 cm wide. Thirty was the number of pairs chosen because: thirty pairs of faces could fit comfortably on a desk top, the difference between a pair faces becomes more difficult to discern for the higher-dimensional PCs, and the more pairs a participant has to sort through the harder the task becomes. This decision also defined the number of dimensions that were used in the large face-space.

At the start of the experiment the pairs of faces were arranged randomly in a grid six pairs high by five pairs wide. The participants were instructed to group the 12 pairs of faces which 'exhibited the most within pair dissimilarity'. Once the participants had done this they were instructed to sort the 12 pairs of faces from the most similar to the least similar. In preliminary testing, it was observed that the degree of dissimilarity between pairs of faces became very hard to discern beyond the 12 or so most dissimilar pairs. Accordingly, 12 dimensions were used for the small search space.

4.2 Results

The 12 most significant PCs perceptually were found to be 1, 2, 3, 4, 5, 6, 7, 9, 13, 14, 15, and 18. Thus, these are the PCs that were used to build the human reduced face-space. It can be seen that 8 of the 12 PCs in the human reduced face-space are in the first 12 PCs of the large PCA face-space.

5. Experiment 2: Comparison of Recombination and Mutation Operators

5.1 Method

In this experiment 15 participants were used to compare two recombination operators (uniform crossover and arithmetic crossover) and two mutation operators (Gaussian replacement and nonuniform mutation).

The 12-dimensional human reduced face-space was used in this experiment. This face-space was chosen because it was not thought that the face-space used would have an effect on the relative performances of the different recombination and mutation operators. However, it was thought that a lower-dimensional face-space may lead to a face match more quickly and thus induce less fatigue in the participants.

Testing each combination of recombination and mutation operator required $2 \times 2 = 4$ runs per participant. Each participant also did a practice run at the start of the experiment.

The initial population was the same for every run of the experiment and was designed to be roughly evenly distributed in the human reduced face-space. To generate the initial population, 1000 points were generated using a 12-dimensional uniform distribution with the limits being at ± 2.5 SDs on each axis. Matlab's *kmeans* function was used to group the points into nine clusters. The centroids of the nine clusters were used as the genotypes for the initial population of faces.

At the start of each run the participants were given 10 seconds to study the target face which they then had to try to recreate from memory using the interactive evolutionary facial composite process. The target face was not shown to the participants again until the end of the run.

The target faces were chosen to be equidistant from the centre of the human reduced face-space.

At the end of every run, the participants were shown the composite they had just created and were asked to rate its similarity to the target on a scale from 1 to 10. Immediately after rating their composite the participants were shown the target face alongside their composite and asked to rate the similarity between their composite and the target again.

Three sets of objective data were gathered: the time taken to create the composites, the number of generations it took to create the composites, and the number of times the back button was used. If the back button was used more often for a particular operator, it indicates that the operator is not particularly suitable for the task.

5.2 Results

The means and standard deviations of the measure variables (number of generations, time taken, number of times the back button was used, participant rating of their composite without reference to the target, participant rating of the their composite with reference to the target) are given in Table 1. Each of the measure variables were subjected to two-way ANOVA having two mutation operators (nonuniform mutation and Gaussian replacement) and two recombination operators (uniform crossover and arithmetic crossover) (Table 2). It can be seen that the main effects of mutation operator and recombination operator were not significant for any of the measure variables, nor was the interaction of the two operators significant.

Table 1. Means (standard deviations) of the dependent variables in the comparison of mutation and recombination operators in the creation of facial composites.

Mutation	Recombination	Generations	Back count	Time taken	Without rating	With rating
Gaussian replacement	uniform	10.6 (5.10)	0.73 (1.33)	195s (91.5s)	6.27 (1.22)	4.40 (2.10)
Gaussian replacement	arithmetic	12.5 (8.64)	0.47 (0.74)	222s (155s)	5.47 (2.00)	5.07 (2.19)
Nonuniform mutation	uniform	11.5 (4.73)	0.87 (1.41)	220s (71.1s)	6.07 (1.03)	4.60 (2.41)
Nonuniform mutation	arithmetic	9.73 (2.49)	0.47 (0.64)	188s (66.2s)	6.07 (1.49)	4.40 (2.32)

Table 2. Two-way ANOVA of the dependent variables in the comparison of mutation and recombination operators in the creation of facial composites.

	Mutation		Recombination		Interaction	
Variable	$F(1,56)$	p-value	$F(1,56)$	p-value	$F(1,56)$	p-value
Generations	0.43	0.513	0.00	0.946	1.56	0.217
Back Count	0.06	0.813	1.41	0.240	0.06	0.813
Time taken	0.03	0.874	0.01	0.904	1.21	0.275
Without comparison rating	0.27	0.603	1.10	0.300	1.10	0.300
With comparison rating	0.16	0.691	0.16	0.691	0.55	0.461

Table 3. Means (standard deviations) of the dependent variables in the comparison of the large, human reduced and mathematically reduced face-spaces in the creation of facial composites.

Face-space	Gens	Back count	Time taken	Without rating	With rating
Large	10.7 (4.73)	0.50 (0.55)	205s (80.3s)	5.81 (1.13)	4.10 (1.25)
Human reduced	9.38 (4.31)	0.36 (0.42)	186s (91.8s)	6.02 (1.08)	3.95 (1.33)
Mathematically reduced	10.5 (4.75)	0.48 (0.56)	193s (85.6s)	5.86 (1.16)	4.12 (1.82)

6. Experiment 3: Comparison of Face-spaces

6.1 Method

In this experiment 21 participants were used to compare three face-spaces: a face-space constructed from the first 30 PCs of the PCA analysis (the large face-space), a face-space constructed from the first 12 PCs (the mathematically reduced face-space), and a face-space constructed form the 12 most perceptually important PCs identified in the first experiment (the human reduced face-space).

As the results of the second experiment showed no significant difference between the operators on any of the recorded measures, arithmetic crossover and nonuniform mutation were arbitrarily chosen as the operators used for this experiment.

As there were only three test conditions (large face-space, human reduced face-space, and mathematically reduced face-space) each participant performed two runs for each of the test conditions so that they performed $2 \times 3 = 6$ runs.

The initial populations for each of the face-spaces were constructed in the same way as that for the second experiment. The target faces were chosen to be equidistant from the centre of the 30-dimensional face-space.

6.2 Results

The means and standard deviations of the measure variables over all of the runs for each of the algorithms are presented in Table 3.

Performing one-way ANOVA on each of the measure variables (averaged over both runs for each of the test conditions) showed that the differences between the face-spaces were not significant for any of the measure variables (number of generations: $F(2,60) = 0.51, p = 0.604$, number of times the 'back' button was used: $F(2,60) = 0.47, p = 0.629$, time taken: $F(2,60) = 0.28, p = 0.758$, without comparison rating: $F(2,60) = 0.21, p = 0.811$, and with comparison rating: $F(2,60) = 0.08, p = 0.926$).

7. Conclusion

A human evaluation based reduced face-space for use with an IEA in the creation of facial composites was derived from a larger PCA based face-space. The performances of searches for faces in the human reduced face-space was compared to those of a mathematically reduced face-space and the larger face-space. The human reduced face-space was also used in the comparison between different mutation and recombination operators in the simple IGA.

The prioritisation of the PCs with regards to human evaluation was found to be similar to the numerical ordering of returned by principal component analysis itself. The human reduced face-space was found to share eight of its 12 PCs with the mathematically reduced face-space. We note that our data set comprised images captured under conditions of controlled pose, lighting and facial expression. If this were not the case, one might expect greater differences between the perceptual and numerical orderings of PCs.

No significant differences in the performances of the searches conducted using the different operators were detected. The difficulty and uncertain nature of creating a facial composite render any differences in the performances of the operators or the face-spaces insignificant.

Likewise, no significant differences in the performances of the searches conducted in the different face-spaces was observed. The implication of this is that it is possible to reduce the dimensionality of the face model without any loss of performance.

References

[1] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In *Proc. European Conference on Computer Vision (ECCV)*, volume 2, pages 484–498, 1998.

[2] G. Davies and D. Christie. Face recall: An examination of some factors limiting composite production accuracy. *J. Appl. Psychol.*, 67(1):103, 1982.

[3] C. D. Frowd. *EvoFIT: A Holistic, Evolutionary Facial Imaging System.* PhD thesis, Department of Psychology, University of Stirling, 2001.

[4] C. D. Frowd, P. J. B. Hancock, and D. Carson. Evofit: A holistic, evolutionary facial imaging technique for creating composites. *ACM Trans. Appl. Percept.*, 1(1):19–39, 2004.

[5] S. J. Gibson, C. J. Solomon, and A. Pallares Bejarano. Synthesis of photographic quality facial composites using evolutionary algorithms. In *Proc. Britsh Machine Vision Conference*, volume 1, pages 221–230, 2003.

[6] B. Kurt, A. S. Etaner-Uyar, T. Akbal, N. Demir, A. E. Kanlikilicer, M. C. Kus, and F. H. Ulu. Active appearance model-based facial composite generation with interactive nature inspired heuristics. *Lect. Notes Comput. Sc.*, 4105:183–190, 2006.

[7] C. J. Solomon, S. J. Gibson, and J. J. Mist. Interactive evolutionary generation of facial composites for locating suspects in criminal investigations. *Appl. Soft Comput.*, 13(7):3298–3306, 2013.

[8] J. W. Tanaka and M. J. Farah. Parts and wholes in face recognition. *Q. J. Exp. Psychol.*, 46A:225–245, 1993.

[9] D.-M. Yoon and K.-J. Kim. Comparison of scoring methods for interactive evolutionary computation based image retouching system. In *Proc. 14th Annual Conference Companion on Genetic and Evolutionary Computation (GECCO)*, pages 617–618, 2012.