Image and Vision Computing 92 (2019) 103813

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

On the importance of local and global analysis in the judgment of similarity and dissimilarity of faces

Manuele Bicego[a,*], Enrico Grosso[b]

[a] Department of Computer Science, University of Verona, Verona, Italy
[b] Department of Agriculture, University of Sassari, Sassari, Italy

ARTICLE INFO

Article history:
Received 17 January 2019
Received in revised form 16 July 2019
Accepted 9 September 2019
Available online 28 October 2019

Keywords:
Face analysis
Biometrics
Image matching
Similarity-dissimilarity of visual content

ABSTRACT

The ability to recognize faces and to detect differences and similarities between faces has proved to be fundamental in the evolution of humans and in the conditioning of their social behaviors. In this paper, we investigate basic mechanisms underlying this ability, focusing in particular on the relevance of local and global features and on some interesting differences characterizing judgments of similarity with respect to judgments of dissimilarity.

In a first experiment, a set of participants is involved in order to evaluate the human response with respect to a simple judgment protocol based on two-alternative forced choice. Triplets of face stimuli are evaluated first with the aim of identifying (between two candidate faces) the face more similar to a reference face. The protocol is then repeated for the same triplets but involving a different set of participants and asking to identify the face less similar to a reference face. These visual judgments of similarity and dissimilarity are finally analyzed and compared with the results of a closely related computational experiment based on the same set of triplets; in this case, however, the similarity-dissimilarity measure is derived by automatically extracting facial points and matching with regression techniques (LASSO and Elastic Net) two configurations of image descriptors: the first capturing holistic information, the second capturing local information, that is few localized facial features.

Our results suggest that computational models based on holistic cues (emphasizing the concept of the whole as a composed set of interdependent parts) better fit judgments of humans participating to the first experiment (similarity judgments). On the other hand, models based on spatially localized cues do not offer significant accuracy. Vice versa, computational models based on local cues better fit dissimilarity judgments and are less adequate to express similarity information. Notably, our results provide some empirical evidence that local and global cues are both important in face perception, but with different roles. This finding supports the hypothesis that similarity and dissimilarity should not merely be considered as opposing concepts, as they could derive from different processing paths.

1. Introduction

1.1. Overview

In recent years research on face perception led to remarkable advances in the understanding of many different aspects of how face are processed and memorized by the human brain [1,2]. The impact of these findings will be significant on a wide range of disciplines and, notably, in the next-generation of human-machine interfaces, as faces provide primary access to other people's identity [3], and can signal behavioral intentions as well as mental and emotional states that play a crucial role in social interactions [4]. In this paper, we focus on a specific and fundamental feature of face perception, namely the ability to evaluate similarity and dissimilarity from visual cues. This problem has been studied in the last years by researchers from different areas, giving rise to different evidences, interpretations and computational models of face and expression recognition [5-7]. Is the recognition is a unitary event or rather a complex combination of multiple distinct functional components? And does the recognition encompass the evaluation of specific points of the face or rather takes the form of an holistic judgment considering the face as a whole? These issues, which constitute the starting point of our study, are briefly introduced in the next section.

[*] Corresponding author.
E-mail addresses: manuele.bicego@univr.it (M. Bicego), grosso@uniss.it (E. Grosso).

https://doi.org/10.1016/j.imavis.2019.09.004
0262-8856/© 2019 Elsevier B.V. All rights reserved.

1.2. Related works

In the wide literature covering this topic, Multi Dimensional Scaling (MDS) models [8,9] have been extensively used to describe the perception of similarity and dissimilarity. In MDS models, the physical attributes of the stimulus (for example, the hue or the intensity of a color) are coded as different dimensions of a metric space and the stimulus is thus identified with a point in such a multidimensional space. Following this geometric approach, all the considered attributes can easily contribute to a unitary measure of distance between stimuli, representing both similarity (small distance) or dissimilarity (large distance). The extension of MDS models to face similarity is straightforward. To this aim, Valentine [10] suggests that faces can be considered as points in a multidimensional "face space" where dimensions represent specific properties that serve to maximize the discrimination process. Valentine does not provide any scheme to identify the attributes of a face that the dimensions represent: he generically refers to previous works using MDS techniques and assuming that "the principal dimensions needed would represent hair color and length, face shape, and age".

A quite different approach, involving multiple functional components, has been firstly detailed in the seminal work of Bruce and Young [11]. They suggest the existence of distinct types of information that people derive from faces and hypothesize the presence of distinct functional/processing routes contributing to face recognition. Interestingly, among different facial types of information, they identify two types (pictorial and structural) that clearly play different roles. Goldstone and colleagues [12] in a general paper on similarity introduce and detail the key concepts of "primitive attributes" and "relations". Questioning MDS models and other models taking into account set tools [13],they propose a separate pooling of attributional and relational similarities. In particular, they clarify that relations are not global features, even though relations can bind two or more arguments and assume a global meaning.

Vokey and Read [14] address the role of typicality (exemplifying most nearly the essential characteristics of a known group) with a principal components analysis, demonstrating the existence of two independent components: the first coding attractiveness, familiarity, likeability and the second coding memorability. They do not comment about specific attributes of the faces but demonstrate that familiarity generally decreases the discrimination ability while memorability enhances it. Moreover, they argue that familiarity has to do with the similarity of the faces whereas memorability is somehow related to distinctiveness and that the two components work in opposition. O'Toole and colleagues [15] extend the work of Vokey and Read. They train an associative neural network to recognize Caucasian and Asian faces and show that the memorability component of recognition is due to small, local distinctive features, while the familiarity component of recognition is related to more global aspects of the shape of the face. This finding is confirmed by Collishaw and Hole [16]; their results not only support a clear distinction between global and local components but also record similar effects for both familiar (typical) and unfamiliar faces.

The idea that perceived similarity is strongly affected by relational structures, extrinsic to the features compared, is well investigated in a recent work of Jones and Love [17]. They show that in addition to the traditional role of isolated features, the presence of common relational structures (both spatial, causal or dynamic) increases the perception of similarity; this is especially evident when objects involved in the relations play the same role. Simmons and Estes [18] further extend this concept demonstrating the importance of thematic relations and the high plausibility of a dual model process: according to this model "thematic relations are not represented as common features" and "comparison is more heavily weighted for thematically related items". Interestingly, this hypothesis can also explain the well known "non-inversion" effect, related to similarity and difference. In fact, Simmons and Estes report that the thematic effect can be significantly attenuated in difference ratings, thus explaining why perceived similarity is not always inversely related to perceived difference.

The possible existence of different pathways in face perception emerges from the work of Schawaninger and colleagues [19]. They analyze the role of local and global (holistic or structural) representations and conclude arguing about a two-route model of face processing and matching. Similar conclusions are suggested by Lorusso and colleagues [20] investigating the visual judgment of similarity, dissimilarity and kinship. In particular, they report priming effects (selective suppression/enhancement of the dissimilarity judgments following similarity and kinship judgments) that suggest the existence of different processing pathways "perhaps modulated by experience and cultural conditioning".

The importance of the spatial localization of visual cues has been recently investigated also by Dal Martello and colleagues [21]. They suggest a configural nature of the kinship judgment, strictly related to the spatial localization of visual cues. The discussion on how recognition of faces is affected by the adoption of a configural representation becomes particularly interesting considering that face recognition in humans is successful over a variety of appearances due to light, pose and external factors. To this respect, findings of Redfern and Benton [22] suggest that expressions are part of facial representation, thus deeply coded in the basic mechanisms detecting and measuring similarities between faces.

Trying to summarize, there are significant psychological evidences that spatial localization (or configural representation) of visual features play a crucial role in the perception of face similarity. This issue has not been disregarded from researchers in computer science proposing and testing computational models suitable to capture basic abilities of humans. For example, Tistarelli and colleagues [23] focus their work on the selection of relevant features and on the local/global matching strategy bringing to a single similarity score. Again from a technical perspective, Zhan and colleagues [24] focus on the familiarity issue and propose a mathematical approach capable of combine reusable features and to distinguish between different forms of familiarity. Edelman and Shahbazi [25] define a global framework taking into account structural similarity and show how this framework can adequately scale to deal with massive visual data. Martinez [26], trying to strictly formalize the problem of identity recognition, criticizes deep learning [27] and focus on "critical" information of faces and on the contribution that features and surface properties can give to facial and expressions recognition [28].

Despite all these efforts, and probably due to the specificity of the face recognition problem, the relationship between computational models and human judgments for face similarity and dissimilarity has been only slightly investigated. In fact, this kind of perspective requires the adoption of a common dataset and a clear identification of a detailed aspect for which both measures (psychological and computational) can be easily derived. The work presented in this paper makes one step forward along this direction, trying to show that this perspective is viable when evaluating the role of holistic and local information on face analysis.

1.3. The proposed study

Our investigation focuses on the assessment of the relation between judgments of similarity and dissimilarity; in more detail, we show that computational models based on local and global (holistic or structural) representations perform differently with respect to corresponding human judgments, giving strength to the hypothesis of a two-route model of face processing and matching.

The study is organized with both psychological and computational experiments. In particular, in a first psychological experiment (Section 2), participants are asked to judge face-triplets that can vary in age or sex, the judgment consisting on a simple two-alternative forced choice (2AFC) between the central-left, central-right pair. The judgments are both on similarity and dissimilarity, with the goal of assessing the relation between similarity and dissimilarity judgments. In the computational experiment (Section 3), the same set of face triplets are characterized using local and global descriptors which are automatically extracted and described following a two-route model. Such descriptors are then used to approximate the results of the psychological experiment, trying to infer relations between local or global characteristics and perceived similarity and dissimilarity. Please note that the image data used in both experiments were validated through a preliminary calibration procedure whereby an independent group of participants graded face-pairs on a 0–1 point scale: from entirely dissimilar to entirely similar. The reader can refer to Ref. [29] for details concerning this procedure.

Summarizing, the main contributions of this study are the following:

1. a psychological experiment is performed, trying to characterize the relation between perceived face similarity and dissimilarity, and providing some evidences that there are indeed some differences ("similarity" is not just the opposite of "dissimilarity");
2. starting from a classic computational characterization of the face image, two computational models are derived, one encoding a global holistic description of the face, the other encoding a more local-discriminative one; the prediction accuracy of these models is thus tested giving some empirical evidences that the perceived face dissimilarity is more related to local configurations, whereas face similarity is mainly due to global descriptors.

These findings offer a novel perspective, opening the door to the possibility of developing novel automatic face recognition methods; in Section 4, we provide some considerations on a possible computational model which explicitly considers that similarity and dissimilarity are not two faces of the same medal, but rather two complementary and distinct processes, to be possibly simultaneously exploited.

2. Experiment 1 – Perceived similarity-dissimilarity

In this section, the psychological experiment is presented. Starting from a controlled set of 79 face pairs, each associated to a robust (independently computed) perceived similarity index (PSI), the purpose of the first experiment was to understand the role of PSI in the judgment of similarity (JS) and dissimilarity (JD).

2.1. Method

In order to obtain similarity and dissimilarity judgments, a two alternatives forced protocol (2AFC) was applied. Triplets used in the 2AFC trials were composed by three images (two face pairs sharing a common reference image, see Fig. 1). The reference face was always positioned in the middle of the triplet; participants had to make a forced choice between the candidate image on the left and the candidate image on the right side, indicating which of the two candidate faces looked the most similar (or dissimilar) to the reference face. Responses were collected by recording the mouse click (position) and the time elapsed since the presentation of the triplet. Subjects were given unlimited time to respond. The first three triplets presented per participant were considered as a training and were omitted from any subsequent data analysis.

2.1.1. Stimuli

Face pairs. A homogeneous dataset of 79 face pairs, derived from an original set of 54 color photographs, each depicting a face, has been carefully selected for the experiment. Faces had mostly spontaneous expressions, being taken from friends' photo albums, without further processing except for the homogenization of the background. Fig. 1 shows a sample face pair. Twenty-five male and twenty-nine female faces were included in the set, spanning an age range from 25 to 62 years. They were all of Caucasian appearance. Forty-four of the total declared themselves to be in kinship relation (either parents-offspring or siblings) with one another. In total, 14 distinct family sets each containing three or four members were so declared.

Please note that face-pairs used for the experiment are all associated to a perceived similarity index (PSI) which is the result of a complex calibration procedure well described by Lorusso and colleagues [29]. This index is somehow analogous to the index that can be obtained by rating scales, an approach commonly used in psychophysics; however, the procedure proposed by Lorusso and colleagues proved to be much more reliable, and useful to take into account some context-based and non-metric behaviors that characterize the human judgment. Without going into the details of the calibration, it is worth noting for our purposes that the rating of each face-pair is the final outcome of an experimental protocol based of 2AFCs, thus very similar to that adopted in this experiment, but involving an exhaustive comparison of the face-pair with all the images of the dataset. The rating of each selected face-pair (from the dataset of 54) finally required the evaluation of 52 different triplets where the pair-reference face was always positioned in the middle of the triplets while the pair-candidate face and the varying candidate were randomly positioned to the left or right side. To improve the statistical significance of the computed PSI, each single triplet was judged by six participants; as a consequence, each face-pair received a total of 312 evaluations (six evaluations for each of the 52 triplets).The final PSI value for the face-pair, in the range 0–1, was derived taking into account both the accordance of the six evaluations and the distribution of the judgments for all the 52 triplets.

Due to the very large number of judgments required in order to rate a single face-pair, only a limited subset of all possible face-pairs was rated by this procedure. In particular 79 items were selected, with the assistance of expert psychologists, in order to guarantee a good representativeness of the subset with respect to i) potential differences in rating ii) balanced presence of kin and non-kin pairs. In total, the calibration experiment involved 24,648 single judgements (79 × 52 × 6). Finally note that, despite a balanced presence of kin and non-kin face-pairs, PSI values above 0.5 (on a range 0–1) tended to be more frequent, suggesting some kind of perceptual saturation in the human judgments. As expected, this effect was much more evident for kin pairs, where 85% of the pairs have shown PSI values above 0.5.

Face triplets. In analogy with the selection of the 79 face pairs, the generation of the triplets proved to be a very difficult task that required the assistance of expert psychologists. Starting from a repeated random selection of a first face-pair, the procedure tried to iteratively associate a second face-pair but always keeping in mind both the family relationships and the differences in PSI. More in detail, three simple constraints were considered:

- large coverage of the available face-pairs (i.e. well distributed use of all the face pairs), provided the existence of a common pair-reference image for each triplet;
- large coverage of differences in PSI (i.e. generation of triplets spanning all possible values of difference in PSI, from −1 to 1);
- large coverage of different types of stimulus (i.e. balanced use of triplets including people related or not-related by kin).

To all the constraints mentioned above it has been assigned the same cost, in such a way as to ensure a fairly uniform distribution

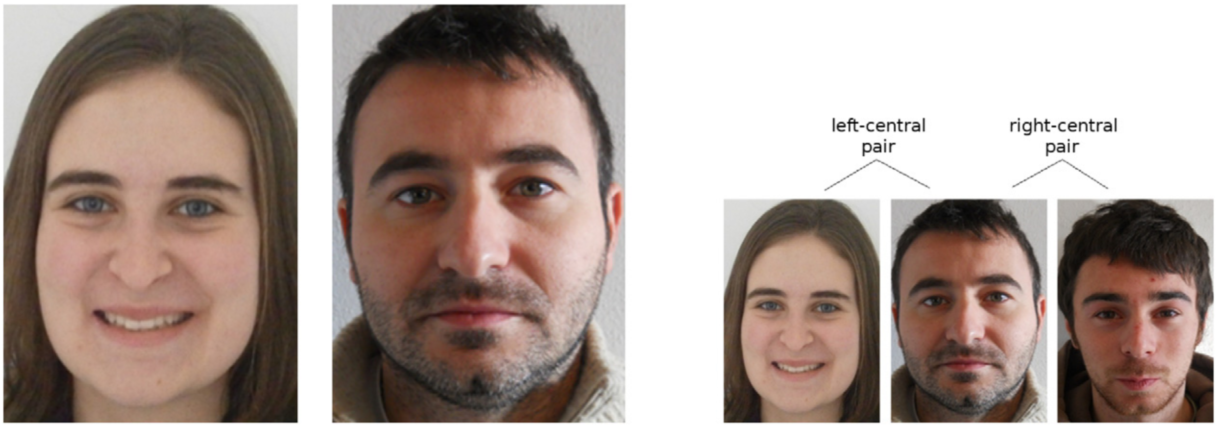

Fig. 1. (Left): a face-pair selected from the original dataset of 54 images. (Right): an example of triplet used in the 2AFC trials.

of the features that characterize the stimuli. Despite the great number of potential triplets theoretically derivable from the 79 face-pairs (3081), the application of cost constraints and the need to identify a common pair-reference image for each triplet drastically reduced (under 300) the number of triplets passed to the manual selection. Also in this case, the assistance of expert psychologists was essential in order to proceed to the final selection and to guarantee a good representativeness of the subset. From now on, the difference in PSI between face-pairs (left-central, right-central) of a triplet will be called PSI delta (PSID). As already stated, triplets included people related or not-related by kin; in particular three groups of triplets were so considered:

- Mixed — where only one of the two face-pairs of the triplet was of people related by kin;
- Non-mixed K — where both face-pairs belonged to people related by kin:
- Non-mixed NK — where both face-pairs belonged to people not-related by kin.

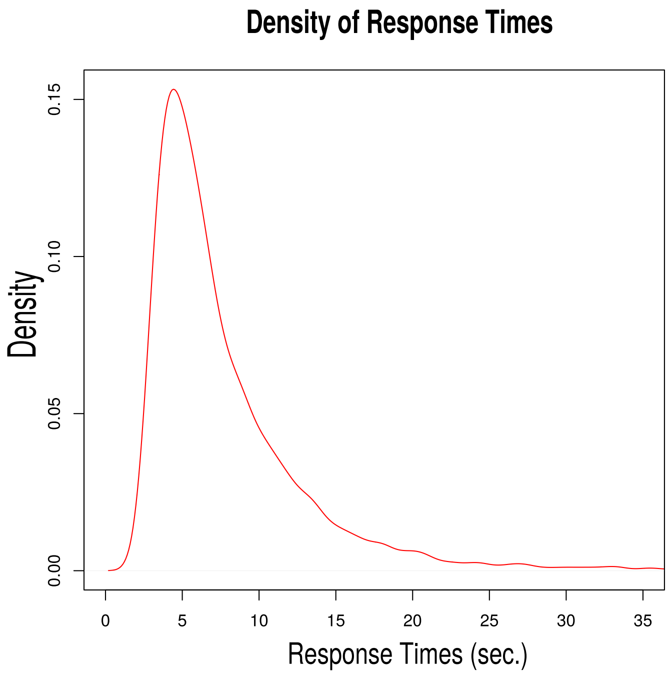

Fig. 2. Distribution of the time response.

The total number of triplets involved in the experiment was 69.

2.1.2. Participants

As usual in psychological studies [30,21], participants in the experiment have been recruited among students of the University. In particular, participants were 64 undergraduates at the University of Sassari (39 females and 25 males), all reported to have normal or corrected-to-normal vision. Mean age of the participants was 23 years.

2.1.3. Equipment and detail of the procedure

Stimuli were presented on a computer monitor, screen resolution 1280 × 1024, refresh rate 60 Hz. Pictures were presented in triplets over a gray background. Participants completed one of the two possible trials (assigning similarity or dissimilarity), a trial being a randomized sequence of 2AFC judgments for the same set of 69 triplets. The order of presentation of the triplets was randomized in order to guarantee the presence of different faces in two consecutive triplets, so avoiding memory effects.

In order to improve the statistical significance of the judgments, the 64 participants were randomly divided in two groups; we finally had 32 subjects assigned to task Judgments of Similarity (JS) and 32 to task Judgments of Dissimilarity (JD). Each of the 64 participants evaluated only 69 triplets, expressing a single kind of judgment. In summary, each single triplet has been involved in 64 2AFC trials (32 JS + 32 JD), for a total number of 4416 single judgments (69 × 64).

2.2. Results

As a first result, Fig. 2 and Table 1 show the distribution of the response times recorded during the experiment. Response times are clearly condensed in the range 0–20 s with a peak around 6 s. Moreover, a non-normal distribution of response times is confirmed; this aspect is not surprising, as it has been reported in different works and faithfully reflects well known previous models [31] based on the so-called ex-Gaussian distribution.

Fig. 3 plots the results obtained from our psychological experiments, for both JS and JD, on the three groups of face-triplets (Mixed, Non-mixed K, Non-mixed NK). In particular, in the plot related to JS (top), the relative amount of choices for the left face-image has been

Table 1
Median of the response times and 95% tolerance interval with a coverage area of 99%.

Num	Median RT	Toler. $\alpha = 0.05, P = 0.99$	
		Lower side	Upper side
2408	6.169	2.632	32.070

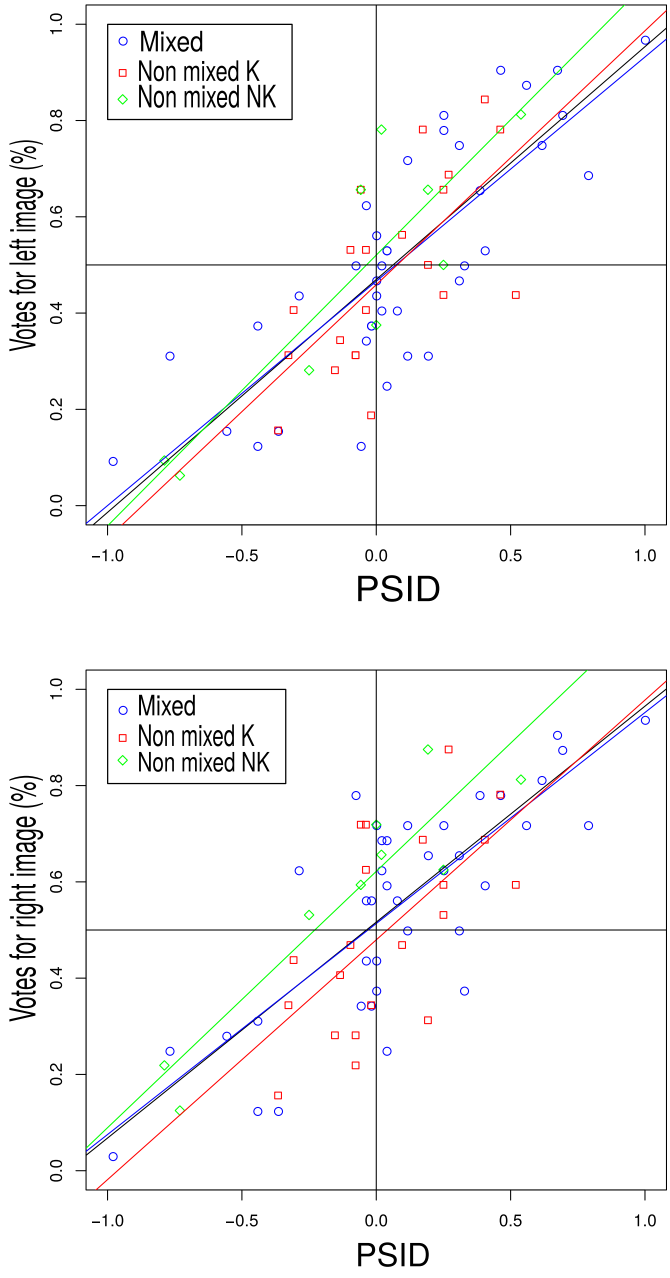

Fig. 3. Results for the perceived similarity (top) and perceived dissimilarity test (bottom). Each point represents the ratio between votes for the left (right) pair and the total number of votes. In blue (circles) the Mixed triplets (Mixed), in red (squares) the Non-mixed triplets where both face-pairs came from the same kin set (Non-mixed K), in green (diamonds) the Non-mixed triplets where both face-pairs came from different kin sets (Non-mixed NK). Lines represent linear regression for Mixed (blu), Non-mixed K (red) and Non-mixed NK (green) triplets. In black the linear regression computed for all the available points. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

plotted in function of PSID; in the plot related to JD (bottom) the relative amount of choices for the right face-image has been preferred in order to facilitate the comparison between JS and JD. Note that for both judgments the percentage of votes seems to linearly increase with the increase of PSID. This is somehow expected because PSID represents the "amount" of similarity that an independent set of observers judged in favor to the left pair of the triplet. In other words, in presence of a positive PSID the choice of the left image is expected for JS, and similarly the choice of the right image is expected for JD. This result can be also considered as a further validation of the method used to compute PSI between faces.

Table 2
Result of the Pearson's correlation test for the similarity (JS) and dissimilarity (JD) judgments with respect to available PSID. "Group" represents the different splitting of tests, "dof" are the degrees of freedom of the test, "coef" is the Pearson correlation coefficient for Judgements of Similarity (JS) and Dissimilarity (JD), together with the corresponding *P*-value ("*P*-val").

Group	dof	Coef (JS)	*P*-val (JS)	Coef(JD)	*P*-val (JD)
Mixed	39	0.802	8.21E−10	0.795	1.47E−09
Non-mixed K	21	0.677	0.0007	0.626	0.0024
Non-mixed NK	9	0.873	0.0021	0.926	0.0003
ALL	69	0.785	1.33E−15	0.758	4.39E−14

Table 2 shows the Pearson correlation coefficients of the responses for all the groups and tasks considered, as a function of the PSID. Correlation coefficients of the ALL group show that the two judgments (JS and JD) depend linearly on the PSID; in both cases (JS and JD) we observe that this dependence is retained by the Mixed group while a variation in the dispersion of data is present for both the Non-mixed groups (an increase of dispersion for Non-mixed K, a decrease for Non-mixed NK). This effect seems more pronounced for JD. On one hand this finding could suggest a limited role of kinship; in fact the presence of kinship increases going from the Non-mixed NK (no kin pairs) to the Mixed (1 kin pair) to the Non-mixed K group (2 kin pairs). On the other hand, however, this result could testify in favor of a non-perfect opposition between the concepts of similarity and dissimilarity (JD and JS seem to be differently affected by Non-mixed triplets).

Table 3 shows the estimated values of the linear regression for the different tasks and groups considered. Standard errors and p-values of the estimates are also reported. Looking carefully to the two parameters characterizing the linear regression, the line's *slope* and *intercept*, some additional aspects emerge. First of all the variation between the Mixed and Non-mixed groups (both in slope and intercept) shows similar trends for JS and JD. This result means that a role of kinship, if proved, would have similar consequences (presumably a mix of "enhancement-suppression" and "bias" effects) for both judgments. Second, note that for all the groups, slope decreases and intercept increases in JD with respect to JS, with a noticeable effect on the intercept. In order to better understand this difference

Table 3
Slope and intercept values for the similarity (JS) and dissimilarity (JD) judgments.

Parameter	Value	Std err	*P*-value
Mixed (DOF: 37)			
Slope JS	0.465	0.056	8.216E−10
Intercept JS	0.466	0.023	3.556E−21
Slope JD	0.438	0.054	1.476E−09
Intercept JD	0.513	0.022	3.700E−23
Non-mixed K (DOF: 19)			
Slope JS	0.527	0.131	0.0007
Intercept JS	0.459	0.033	2.073E−11
Slope JD	0.498	0.142	0.0024
Intercept JD	0.480	0.035	3.663E−11
Non-mixed NK (DOF: 7)			
Slope JS	0.562	0.119	0.0021
Intercept JS	0.520	0.050	1.675E−05
Slope JD	0.532	0.082	0.0003
Intercept JD	0.622	0.035	4.209E−07
All (DOF: 67)			
Slope JS	0.483	0.046	1.330E−15
Intercept JS	0.469	0.017	1.166E−37
Slope JD	0.448	0.047	4.392E−14
Intercept JD	0.516	0.017	6.660E−40

we thus considered the two linear regressions of the whole set (ALL group) and statistically compared the slope and the intercept values. Note that both these parameters follow the *t* distribution and that for both the *t*-value can be computed by the expression:

$$t = \frac{p_1 - p_2}{\sqrt{S_1^2 + S_2^2}} \tag{1}$$

where p_1 and p_2 are the considered parameter (slope or intercept) for the two populations and S_1, S_2 the corresponding standard errors.

Table 4 shows the result of the test; note that with a confidence interval at 95% the null hypothesis (same populations) cannot be rejected but for the intercept the *t*-value is very close to the critical value; the *p*-value on the last column better explains the fact that the null hypothesis would be rejected at significance level only slightly higher. For our purposes, this result testifies in favor of a possible (significant) distinction between JS and JD and suggests the presence of different or partly different functional components that participate to the formation of the final judgment. We then proceeded in the investigation of a computational model taking into account different processing pathways for JS and JD. All these aspects are better developed and discussed in the following section.

3. A computational comparison of similarity and dissimilarity

The main goal of this section is to propose a computational approach to the problem of comparing similarity and dissimilarity of faces. Clearly, we are aware that in recent years a huge amount of computational methods have been proposed for measuring similarities (or dissimilarities) between images of faces, mainly for biometrics purposes [5,32,33,7]. However, it is important to note that in this work we are not interested in proposing a novel method for face recognition, neither to implement the most accurate and recent approaches in this field. The main goal here is to derive two simple computational models usable to model perceived similarity and dissimilarity measures derived from the previous experiment; the final aim is to provide some computational evidence that these measures are different in terms of the information of the face which is exploited. To this end, the two computational models get inspiration from two opposite and widely known paradigms in face recognition: holistic and local paradigms [34]. In the first paradigm (also referred to as global, configural, relational, and monolithic), faces are perceived as units, and characterized with features extracted from the entire face. On the contrary, the second paradigm (also referred to as part-based, analytic, feature-based, piecemeal, componential and fine grained) is focused on the exploitation of specifically localized parts of the face. Our goal here is to investigate in which way these two complementary processing schemes are related to similarities and dissimilarities between faces.

In short, our computational experiment is composed of four steps:

- a set of iconic features is extracted in specifically relevant locations, from all the faces of the triplets;
- this set of iconic features is encoded following two different strategies, one based on a holistic configuration (using features from all parts of the face) and one based on a local features (using only few distinctive parts);
- the two coded sets of features are used as input to two regressors, one fitting perceived similarity and the other fitting perceived dissimilarity. At the end of this process, four possible cases are considered: using Holistic Features to model similarity (HolF-Sim), using Holistic features to model dissimilarity (HolF-Dis), using Local features to model similarity (LocF-Sim) and at last using Local features to model dissimilarity (LocF-Dis);
- via a cross validation experiment, the four cases above are tested, and experimental evidence is provided that when fitting the similarity the model based on Holistic Features is better than the model based on Local Features. On the contrary, when the aim is to fit the dissimilarity, the model based on Local Features is better than the one based on Holistic Features.

Table 4
Statistical comparison for slope and intercept values between similarity (JS) and dissimilarity (JD) judgments (ALL group).

Parameter	dof	t	$t*$ ($\alpha = 0.05$)	P-value
Slope	134	0.530	±1.896	1.403
Intercept	134	−1.882	±1.896	0.061

Before going into the details of the method, let us stress that we are not proposing a novel model for visual analysis of a face, but we simply propose two opposite modalities of exploiting the perceptual results of the previous section, empirically showing how these modalities are related to similarity and dissimilarity. Overall, we provide additional evidence in favor of a multi-route model for judging similarity-dissimilarity of faces and we show that this model should take into account both local and global information.

3.1. Facial features

Many open tools exist for the automatic extraction of facial features. We adopted Openface, an open source, state of the art tool [35] capable of facial landmark detection and head pose estimation. Facial landmarks detection is performed over 68 points, see Fig. 4. Original images are then cropped, rotated and scaled taking the interocular distance as the only normalizing factor. The extraction of the iconic information is performed for all the 68 points by resampling through a log-polar scheme. The log-polar transformation applied here is that described in Ref. [30] but with a very limited number of receptive fields (8 radial fields along 12 angular directions). Moreover, mean and Laplacian of Gaussians filters are applied to each receptive field. With reference to Fig. 5, each descriptor finally results into a 96 × 2 vector of floats; only 96 of these values, the first part of the vector related to the mean filter, are used in the following experiments.

In summary, each face A is characterized by a set $\mathscr{F}(A)$, which contains M descriptors, i.e.,

$$\mathscr{F}(A) = \{\mathbf{f}_1(A), \mathbf{f}_2(A), \cdots \mathbf{f}_M(A)\}, \tag{2}$$

where M represents the number of landmarks extracted from the face (in our experiments $M = 68$); each descriptor $\mathbf{f}_i(A)$ is itself a vector of Z dimensions (where $Z = 96$ represents the total number of receptor fields of the log-polar mapping). Note that the set of landmarks $\mathscr{F}(A)$ is the basis for both the local and the global (holistic) models described in the following sections. In particular, even though single landmarks come from local points (and single descriptors characterize local areas of the image), the full set of descriptors can be considered as a good representation of the whole image while a local model will only keep part of this information [36].

3.2. Holistic vs local models

A single triplet of the perceptual experiment is involved in 32 similarity and 32 dissimilarity judgments. Let us denote by the letters L(eft), C(entral) and R(ight) the three faces considered; this process provides a final similarity score for the left-central pair ($S_p(L,C)$) and for the right-central pair ($S_p(C,R)$). Analogously, it provides final dissimilarity scores $D_p(L,C)$ and $D_p(C,R)$. The subscript p indicates that these measures are derived from the perceptual experiment, possibly being even not metric. By construction $D_p(L,C) = 1 - D_p(C,R)$

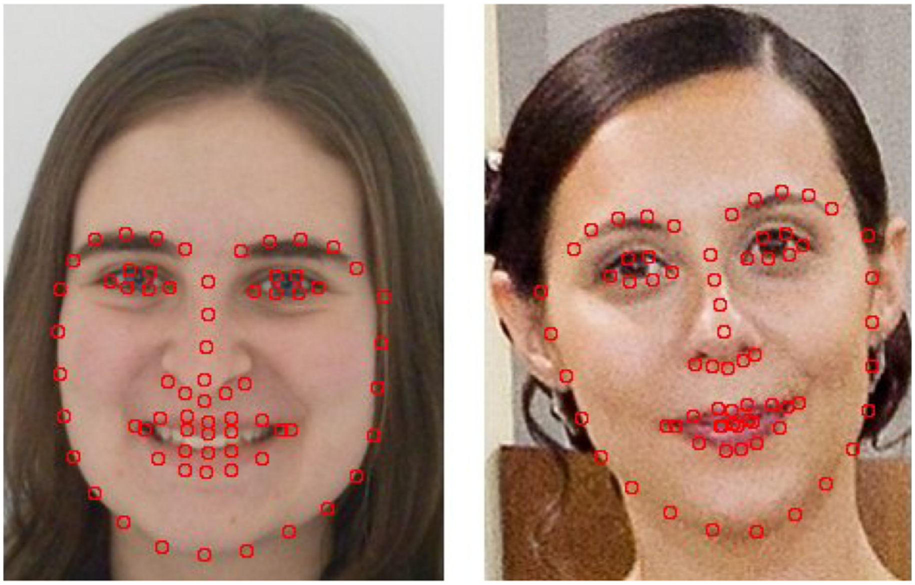

Fig. 4. The automatic extraction of facial landmarks provided by the Openface toolkit: note that the tool is robust to rotation and small scaling factors.

and $S_p(L,R) = 1 - S_p(C,R)$. This is because, when assessing the similarity-dissimilarity in the perceptual experiment, the subjects have to decide between two pairs of faces: (L,C) and (C,R). As said before, we devised two scenarios, one more linked to a holistic configuration, and one linked to a local configuration. These scenarios are aimed at forming the vector $\mathbf{x}_{LCR}$ to be used for the regression of the similarity-dissimilarity.

3.2.1. Holistic model

In this case, the vector $\mathbf{x}_{LCR}$ is defined in two steps:

1. Starting from the set of features $\mathcal{F}(L)$, $\mathcal{F}(C)$ and $\mathcal{F}(R)$, we first compute the distance between all pairs of corresponding features of face L and C (please remember that the M points are ordered). In other words, we compute the vector of distances $\mathbf{d}_f(L,C)$ (the subscript f indicates that this represents a distance between features):

$$\mathbf{d}_f(L,C) = \begin{bmatrix} d(\mathbf{f}_1(L), \mathbf{f}_1(C)) \\ d(\mathbf{f}_2(L), \mathbf{f}_2(C)) \\ \vdots \\ d(\mathbf{f}_M(L), \mathbf{f}_M(C)) \end{bmatrix} \tag{3}$$

$d(\cdot,\cdot)$ is any distance between vectors (Euclidean, based on correlation or others). In the same way, we compute the vector $\mathbf{d}_f(C,R)$.

2. The vector $\mathbf{x}_{LCR}$ is obtained by concatenating the two vectors $\mathbf{d}_f(L,C)$ and $\mathbf{d}_f(C,R)$:

$$\mathbf{x}_{LCR} = \begin{bmatrix} \mathbf{d}_f(L,C) \\ \mathbf{d}_f(C,R) \end{bmatrix} \tag{4}$$

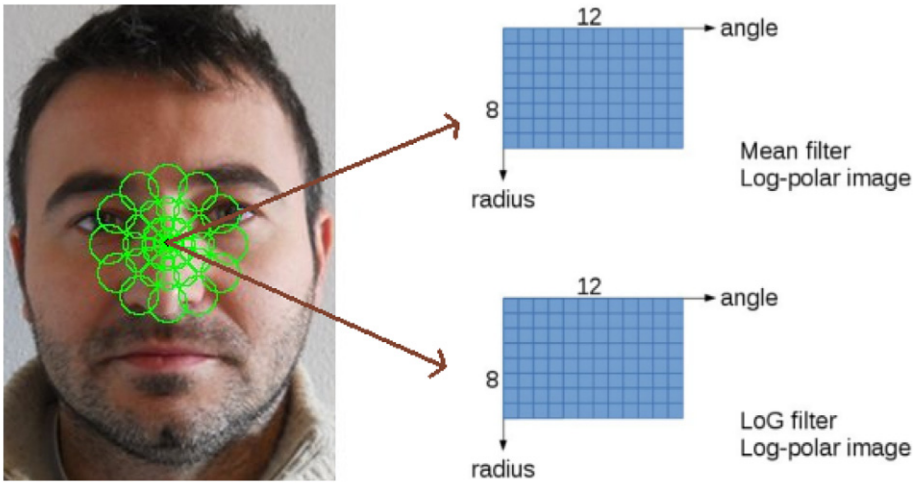

Fig. 5. A single facial landmark is encoded through a log-polar mapping; mean and LoG filters are applied to each receptive field.

The idea is that in the experimental set up to assess the similarity-dissimilarity a human has to use both the comparison between (L,C) and that between (C,R).

This represents an holistic approach, due to the nature of the log polar features (which encode a region of the face) but mainly due to the fact that we are using all the available information: actually the similarity is measured by comparing all the different parts of the face.

3.2.2. Local model

In the Local Model, we start from the same set of features $\mathcal{F}(L)$, $\mathcal{F}(C)$ and $\mathcal{F}(R)$, but take a different approach. In particular, we exploit the observation that what makes two faces different is encoded in few points, namely the points where the two faces differ *the most*. To encode this fact, we extract from vectors $\mathbf{d}_f(L,C)$ and $\mathbf{d}_f(C,R)$ *the K largest values*, which are relative to the most different parts of the face. In practice:

1. in this step, as in the Holistic Model, we compute the vectors $\mathbf{d}_f(L,C)$ and $\mathbf{d}_f(C,R)$
2. we sort (in descending order) the two vectors, obtaining the vectors $\mathbf{d}_f^{ord}(L,C)$ and $\mathbf{d}_f^{ord}(C,R)$
3. the vector $\mathbf{x}_{LCR}$ is obtained by concatenating only the largest K values of $\mathbf{d}_f^{ord}(L,C)$ and $\mathbf{d}_f^{ord}(C,R)$, i.e.

$$\mathbf{x}_{LCR} = \begin{bmatrix} \mathbf{d}_f^{ord}(L,C)_{[1\ldots K]} \\ \mathbf{d}_f^{ord}(C,R)_{[1\ldots K]} \end{bmatrix} \tag{5}$$

where the suffix $[1\ldots K]$ indicates the first K elements of the vector.

In our Local Model, we only use few points to represent the experiment, in particular the most distinctive points between the pairs (L,C) and (C,R). Please note that the selected features can be in principle different for every face, since we are extracting the most distinctive ones for the given comparison (in one case, the difference between the two faces can be seen in the nose, in another in the lips).

3.3. Regression

Once encoded the set of psychophysical experiments into a set of n vectors $\{\mathbf{x}_i\}$, the goal is to find a function of $\mathbf{x}$ which approximates the similarity (dissimilarity) values. This is clearly a regression problem, which can be addressed with different approaches. In general, most of them can be formulated in the following way: given a set of training points $\{\mathbf{x}_i, y_i\}$ (where y_i represents the target value for the input $\mathbf{x}_i$), the goal is to find a function $\mathcal{R}(\mathbf{x})$ which approximates the target values y_i. Since the goal is to not to memorize the training set but rather to capture the general behavior of the function (i.e. generalization), typically a restriction is imposed on the function (the so-called regularization). In our work, we used two kinds of regression approaches: LASSO (Least Absolute Shrinkage and Selection Operator [37]) and Elastic Net [38], all based on a *linear* approximation of the input, i.e.,

$$y = \mathbf{x}^T\beta + \epsilon, \tag{6}$$

where β represent the regression coefficients (a vector of dimension p, where p is the dimension of the input space where $\mathbf{x}$ lives) and ϵ is the residual error. The difference between the two approaches lies in the way they estimate the regression coefficients β. Given a set of n examples $\mathbf{x}_i$ encoded in the training matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n]$, and given the corresponding target vector $\mathbf{y} = [y_1, y_2, \cdots, y_n]^T$, the regression coefficients β are retrieved via the minimization of two different functions:

$$\text{LASSO}: \beta_{LASSO}(\lambda) = \arg\min_{\beta} \parallel \mathbf{y} - \mathbf{X}^T\beta\|^2 + \lambda \parallel \beta\|_1 \tag{7}$$

$$\text{Elastic Net}: \beta_{EN}(\lambda,\delta) = \arg\min_{\beta} \parallel \mathbf{y} - \mathbf{X}^T\beta\|^2 + \delta \parallel \beta\|^2 + \lambda \parallel \beta\|_1 \tag{8}$$

Both techniques minimize the approximation error ($\|\mathbf{y} - \mathbf{X}^T\beta \parallel^2$), with two different regularization terms: in all cases, the goal is to force the shrinking of coefficients of β toward to zero, in order to remove irrelevant contributions. The LASSO penalty term forces the regularization coefficients to be *exactly zero* (i.e. to have a sparse solution); the Elastic Net extends the LASSO approach by adding a L_2 regularization term (similarly to what is used in Ridge Regression [39]), which permits to derive a close form solution for the minimization problem. For more information on these approaches we refer the interested readers to Refs. [37-39].

3.4. Empirical results

In this section, we describe the experiments aimed at evaluating the proposed models. In particular, the scheme is applied to the 69 triplets of the perceptual experiment, each one involving a face triplet. For each face we extracted 68 fiducial points (i.e. $M = 68$), and we computed the log-polar descriptor (96 values) for each fiducial point — see Section 3.1. We then computed the Euclidean distance between feature descriptors $d(\mathbf{f}_i(A), \mathbf{f}_i(B))$, $(1 \leq i \leq M)$. Before estimating the regression coefficients, we employed Principal Component Analysis to remove the redundancy present in the vectors $\{\mathbf{x}_i\}$: in particular we reduced the space via PCA by keeping the 99.99% of the variance explained by the data: in this way, we had a number of coefficients which was inferior to the number of objects (to get more accurate regression solutions). Regression has been performed using the SPASM software[1], which implements, among others, the LASSO and the Elastic Net regression methods. In order to estimate the parameters of these models (i.e. λ for LASSO, (λ, δ) for Elastic Net) a sequence of models is estimated, each one with a different parameters configuration: the best model is then chosen according to the Akaike's Information Criterion, which is estimated for each model (k) as

$$AIC^{(k)} \parallel \mathbf{y} - \mathbf{X}^T\beta^{(k)}\|^2 + 2\sigma_\epsilon^2 df^{(k)} \tag{9}$$

where $df^{(k)}$ represents the number of non-zero elements in the vector of coefficients $\beta^{(k)}$, and σ_ϵ^2 represents the residual variance of a low-bias model defined as

$$\sigma_\epsilon^2 = \frac{1}{n} \parallel \mathbf{y} - \mathbf{X}^\dagger\mathbf{y}\|^2, \tag{10}$$

where $\mathbf{X}^\dagger$ is the Moore-Penrose pseudo-inverse of $\mathbf{X}$. The best model is the one minimizing Eq. (9).

For what concerns K, which represents the number of *maximally distant* points considered to build the Local model, we set this value to 34. This value, in the middle of the range of possible values [1–68], has been chosen as a compromise between two extremes: from one hand, a very small K does not permit a good regression, since the $\mathbf{x}$ vector contains too few elements, becoming not informative. On

[1] Available at http://www2.imm.dtu.dk/projects/spasm/.

the other side, with a too large *K*, we loose the local behavior we are interested in, coming back to the Holistic approach.

In order to test the two different models (Local and Holistic) with the two different measures (Similarity and Dissimilarity), we estimated four models: one using Holistic Features to model similarity (HolF-Sim), one using Holistic features to model dissimilarity (HolF-Dis), one using Local features to model similarity (LocF-Sim) and the last using Local features to model dissimilarity (LocF-Dis). The goal is to show that Holistic Features are better when modelling similarity, whereas Local Features are better when estimating dissimilarity. To have a robust comparison we performed a Leave One Out Cross Validation procedure: we used the whole set, except one, to estimate the parameters of the regressor, using the left one to test it; we repeated this procedure until all elements of our perceptual experiment have been left out. To test a regressor we computed the Mean Squared Error (MSE), defined as the squared difference between the true target and the estimated one:

$$MSE = \frac{1}{n}\sum_i (y_i - \mathbf{x}_i^T\beta)^2 \qquad (11)$$

The final Leave One Out MSE measure (LOO-MSE) is then obtained by averaging the MSE on the different trials. In Table 5, we reported the results, for the two different regression models.

In order to provide more support to our conclusions, we repeated the regressions with the local model by letting *K* vary in the range [16–50] (step 2). We then report the LOO-MSE value, averaged over all different *K* (last two rows of Table 5). From the Table, we can observe that the best fitting for the similarity is obtained with the holistic model, whereas the best fitting for the dissimilarity is obtained with the Local model: this is more evident with the chosen *K* ("Fixed K"), but still holds when averaging results for multiple *K* ("ALLK"). To better investigate this aspect, we reported in Fig. 6 the LOO-MSE error when fitting the Similarity with Local Features (LocF-Sim) and the LOO-MSE error when fitting the Dissimilarity (LocF-Dis), for the different values of *K* (the plot is the average between LASSO and Elastic Net). This plot supports the conclusion: LocF-Dis is consistently better (lower LOO-MSE) that LocF-Sim among the whole range. Moreover, it can be seen that the largest difference is in the middle of the range (as hypothesized before). Please note that all these observations are valid for both regressors. As expected, regression with Elastic Net is more accurate than the one obtained with LASSO.

In order to have a statistical significance, we performed a *t*-test(with a significance level of 0.05) to compare the two pairs of results (i.e. HolF-Sim vs HolF-Dis and LocF-Sim vs LocF-Dis). In all cases, we performed a paired *t*-test comparing all the corresponding LOO MSE, under the null hypothesis that the two samples come from distributions with the same means. In all cases, this hypothesis has been rejected, thus meaning that the difference is statistically significant. The corresponding *p*-values are reported in Table 6: again, we report results for fixed *K* as well as for all *K*.

Table 5
Result of the leave one out experiment.

Method	LASSO LOO-MSE	Elastic Net LOO-MSE
(HolF-Sim)	0.0342	0.0296
(HolF-Dis)	0.1177	0.0779
(LocF-Sim) — Fixed K	0.0683	0.0584
(LocF-Dis) — Fixed K	0.0367	0.0379
(LocF-Sim) — ALL K	0.0580	0.0506
(LocF-Dis) — ALL K	0.0428	0.0383

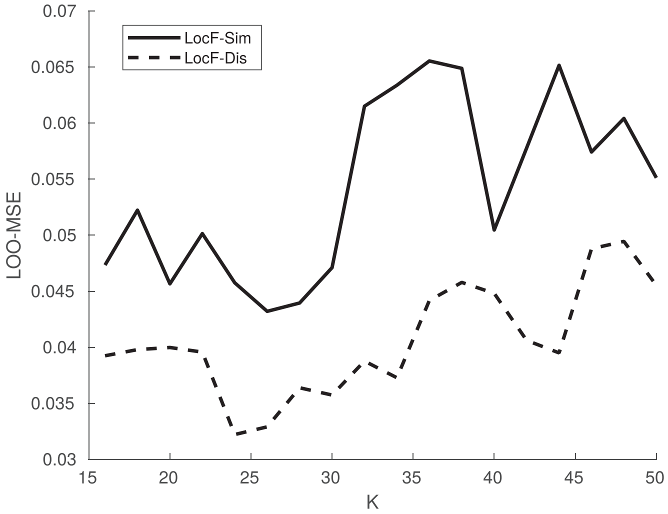

Fig. 6. LOO-MSE when varying *K* in the Local Model.

4. Suggestions for a two-routes computational model

Results proposed in the previous sections point toward the definition of a new two-routes computational model for face recognition, suitable to simultaneously exploit similarity and dissimilarity information. This perspective, which has been considered for domains like shape categorization [40], spectral clustering [41] and person re-identification [42] has been only partly applied to face recognition, as most of the attention has been independently devoted to similarity or dissimilarity measures (see for example Ref. [43]).

Following ideas emerged in our manuscript, we hypothesize in Fig. 7 a possible model for the implementation of a two-route model. In our perspective, developers should take into consideration at least four points: i) feature extraction, ii) feature coding, iii) metric for similarity/dissimilarity estimation, and iv) basic mechanisms of fusion/integration of the two routes. Concerning the first aspect (feature extraction), clearly one can resort to many well established methods usable to extract salient points [44]: this should be probably preserved as a first, low level stage of feature extraction. These descriptors could be immediately used in order to code dissimilarity (local) features. However, the coding of holistic features will probably require an additional effort to better understand and define relational structures that condition holistic perception. Concerning similarity/dissimilarity estimation, an important point is the adopted metric. To this purpose, different metrics could be used for similarity and dissimilarity. More importantly, in line with some literature results [45], we suggest to implement some enhancement/suppression mechanisms giving rise to non-linear responses of the two paths, depending on the strength of the similarity/dissimilarity measures. This would be also in line with the old proposal of Tversky (the so called Tversky contrast model [13]) that suggests a visual description in terms of qualitative features and the comparison in terms of presence or absence of such specific features. Concerning the last aspect

Table 6
Statistical test. "FK" refers to local models which use a single fixed *K* (Fixed K), whereas "AK" is related to results obtained using all *K* (ALL K).

Comparison	LASSO		Elastic net	
	Result	*P*-value	Result	*P*-value
(HolF-Sim) vs (HolF-Dis)	Reject	2.6823e−06	Reject	2.7277e−05
(LocF-Sim-FK) vs (LocF-Dis-FK)	Reject	0.0019	Reject	0.0146
(LocF-Sim-AK) vs (LocF-Dis-AK)	Reject	2.613e−11	Reject	2.432e−10

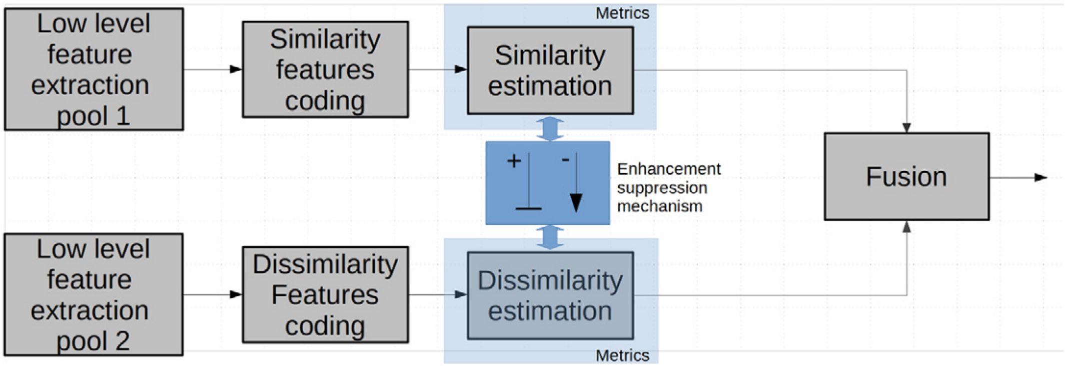
Low level feature extraction pool 1
Similarity features coding
Metrics
Similarity estimation
Enhancement suppression mechanism
+
-
Fusion
Low level feature extraction pool 2
Dissimilarity Features coding
Dissimilarity estimation
Metrics

Fig. 7. A possible two-routes computational model.

(fusion/integration of the two routes) traditional approaches (voting, ranking, weighted average) could be applied. However, we see a potentially interesting approach also in the application of metric learning techniques, i.e. in methods which try to derive a global proximity measure which can be meaningful for the specific classification problem. An interesting example, in this sense, can be found in Ref. [46], where, in a clinical scenario, different metrics derived from different clinicians are fused together to derive a global measure of similarity between patients.

5. Conclusions

In this paper, we investigated basic mechanisms underlying the ability of humans to discriminate similarities and differences between faces. We focused in particular on the relevance of local and global features and on some interesting peculiarities characterizing judgments of similarity with respect to judgments of dissimilarity. First, a psychological experiment has been designed and performed, with the aim of verifying whether similarity and dissimilarity are two aspects of the same process; results provide evidences that there are indeed some differences, which emerge independently of the stimulus type. Second, we tested two computational models, based on a classical pipeline for the characterization of faces. Through these models, we provided some empirical evidences about the origin of this difference, suggesting that the perceived face dissimilarity is more related to a local analysis of the face, whereas face similarity is mainly due to global descriptors. Based on these observations, we also suggest a possible novel two-routes computational model, aimed at explicitly considering that similarity and dissimilarity are not two faces of the same medal, but rather two complementary and distinct processes, probably exploiting different image cues.

Future work will be devoted to the implementation of the proposed computational two-routes model for face recognition. Being conscious that the size of the dataset so far adopted for the computational experiment is limited, especially if compared with recent benchmarks developed for assessing automatic face recognition methods, additional effort will be also addressed to the extension of the dataset involved, for instance considering all the possible triplets that can be generated starting from the available set of face pairs. Further work will be finally addressed to better investigate the variegated set of local descriptors proposed in the literature, and to better evaluate the dependence of the recognition from the quality and the amount of local information involved in the process.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This research has been sponsored in part by the University of Sassari through the grant "Fondo di Ateneo per la ricerca 2019".

References

[1] A. Little, B. Jones, L. DeBruine, The many faces of research on face perception, Philos. Trans. R. Soc. Lond. B Biol. Sci. 366 (1571) (2011) 1634–1637.
[2] N. Rule, Introduction to the special issue on face perception, Curr. Dir. Psychol. Sci. 26 (3). (2017)211–211.
[3] A. Calder, A. Young, Understanding the recognition of facial identity and facial expression, Nat. Rev. Neurosci. 6 (8) (2005) 641–651.
[4] P. Ekman, E. Rosenberg, What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS), 2nd Ed. ed., Oxford University Press. 2005.
[5] S. Li, A. Jain, Handbook of Face Recognition Springer, 2011.
[6] J. Kumari, R. Rajesh, K. Pooja, Facial expression recognition: a survey, Procedia Comput. Sci. 58 (2015) 486–491.
[7] C. Ding, D. Tao, A comprehensive survey on pose-invariant face recognition, ACM Trans. Intell. Syst. Technol. 7 (3) (2016) 37:1–37:42.
[8] R. Shepard, The analysis of proximities: multidimensional scaling with an unknown distance function. ii, Psychometrika 27 (3) (1962) 219–246.
[9] N. Jaworska, A. Chupetlovska-Anastasova, A review of multidimensional scaling (mds) and its utility in various psychological domains Tutor, Quant. Methods Psychol. 5 (2009)
[10] T. Valentine, A unified account of the effects of distinctiveness, inversion, and race in face recognition, Q. J. Exp. Psychol. Sect. A 43 (2) (1991) 161–204.
[11] V. Bruce, A. Young, Understanding face recognition, Br. J. Psychol. 77 (3) (1986) 305–327.
[12] R. Goldstone, D. Medin, D. Gentner, Relational similarity and the nonindependence of features in similarity judgments, Cogn. Psychol. 23 (2) (1991) 222–262.
[13] A. Tversky, Features of similarity, Psychol. Rev. 84 (4) (1977) 327–352.
[14] J. Vokey, J. Read, Familiarity, memorability, and the effect of typicality on the recognition of faces, Mem. Cogn. 20 (3) (1992) 291–302.
[15] A. O'toole, K. Deffenbacher, D. Valentin, H. Abdi, Structural aspects of face recognition and the other-race effect, Mem. Cogn. 22 (2) (1994) 208–224.
[16] S. Collishaw, G. Hole, Featural and configurational processes in the recognition of faces of different familiarity, Perception 29 (8) (2000) 893–909.
[17] M. Jones, B. Love, Beyond common features: the role of roles in determining similarity, Cogn. Psychol. 55 (3) (2007) 196–231.
[18] S. Simmons, Z. Estes, Individual differences in the perception of similarity and difference, Cognition 108 (3) (2008) 781–795.
[19] A. Schwaninger, J. Lobmaier, C. Wallraven, S. Collishaw, Two routes to face perception: evidence from psychophysics and computational modeling, Cognit. Sci. 33 (8) (2009) 1413–1440.
[20] L. Lorusso, G. Brelstaff, L. Brodo, A. Lagorio, E. Grosso, Visual judgments of kinship: an alternative perspective, Perception 40 (2011) 1282–1289.
[21] M.D. Martello, L. DeBruine, L. Maloney, Allocentric kin recognition is not affected by facial inversion, J. Vis. 15 (13) (2015) 5.
[22] A. Redfern, C. Benton, Expression dependence in the perception of facial identity, i-Perception 8 (3). (2017)2041669517710663.
[23] M. Tistarelli, M. Bicego, E. Grosso, Dynamic face recognition: from human to machine vision, Image Vis. Comput. 27 (3) (2009) 222–232.
[24] C. Zhan, W. Li, P. Ogunbona, Measuring the degree of face familiarity based on extended nmf, ACM Trans. Appl. Percept. 10 (2) (2013) 8:1–8:22.
[25] S. Edelman, R. Shahbazi, Renewing the respect for similarity, Front. Comput. Neurosci. 6 (2012) 45.
[26] A. Martinez, Computational models of face perception, Curr. Dir. Psychol. Sci. 26 (3) (2017) 263–269.

[27] Y. Taigman, M. Yang, M. Ranzato, L. Wolf, Deepface: closing the gap to human-level performance in face verification, 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014. pp. 1701–1708.
[28] M. Sormaz, A. Young, T. Andrews, Contributions of feature shapes and surface cues to the recognition of facial expressions, Vision Res. 127 (Supplement C) (2016) 1–10.
[29] L. Lorusso, L. Pulina, E. Grosso, The measure of perceived similarity between faces: old issues for a new method, Rev. Philos. Psychol. 6 (2) (2015) 317–339.
[30] M. Bicego, E. Grosso, A. Lagorio, G. Brelstaff, L. Brodo, M. Tistarelli, Distinctiveness of faces: a computational approach, Trans. Appl. Percept. 5 (2) (2008) 11:1–11:18.
[31] R. Whelan, Effective analysis of reaction time data, Psychol. Rec. 58 (2008) 475–482.
[32] X. Tan, S. Chen, Z.H. Zhou, F. Zhang, Face recognition from a single image per person: a survey, Pattern Recogn. 39 (9) (2006) 1725–1745.
[33] S. Ouyang, T. Hospedales, Y.-Z. Song, X. Li, C.C. Loy, X. Wang, A survey on heterogeneous face recognition: sketch, infra-red, 3d and low-resolution, Image Vis. Comput. 56 (2016) 28–48.
[34] W. Zhao, R. Chellappa, P. Phillips, A. Rosenfeld, Face recognition: a literature survey, ACM Comput. Surv. 35 (2003) 399–458.
[35] T. Baltrušaitis, P. Robinson, L.-P. Morency, Openface: an open source facial behavior analysis toolkit, IEEE Winter Conference on Applications of Computer Vision, 2016.
[36] M. Tistarelli, E. Grosso, Active vision-based face authentication, Image Vis. Comput. 18 (2000) 299–314.
[37] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. B 58 (1) (1996) 267–288.
[38] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, J. R. Stat. Soc. B 67 (2) (2005) 301–320.
[39] A. Hoerl, R. Kennard, Ridge regression: biased estimation from nonorthogonal problems, Technometrics 12 (1) (1970) 55–67.
[40] F. Mathy, H. Haladjian, E. Laurent, R. Goldstone, Similarity-dissimilarity competition in disjunctive classification tasks, Front. Psychol. 4 (2013) 26.
[41] B. Wang, L. Zhang, C. Wu, F. Li, Z. Zhang, Spectral clustering based on similarity and dissimilarity criterion, Pattern. Anal. Applic. 20 (2) (2017) 495–506.
[42] M. Ye, C. Liang, Y. Yu, Z. Wang, Q. Leng, C. Xiao, J. Chen, R. Hu, Person reidentification via ranking aggregation of similarity pulling and dissimilarity pushing, IEEE Trans. Multimed. 18 (12) (2016) 2553–2566.
[43] M. Hernández-Durán, Y. Plasencia Calana, H. Méndez-Vázquez, Metric learning in the dissimilarity space to improve low-resolution face recognition, Proc. Ib. Conf. on Image Recognition and Processing, 2016. pp. 217–224.
[44] H. Wang, J. Hu, W. Deng, Face feature extraction: a complete review, IEEE Access 6 (2017) 6001–6039.
[45] K. Ruys, R. Spears, E. Gordijn, Different and similar at the same time: when the activation of dissimilarity makes people look more similar, Eur. J. Soc. Soc. 38 (3) (2008) 576–585.
[46] F. Wang, J. Sun, S. Ebadollahi, Composite distance metric integration by leveraging multiple experts' inputs and its application in patient similarity assessment, Stat. Anal. Data Min. 5 (1) (2012) 54–69. special Issue: Best Papers of SDM'11.