Blind Facial Image Quality Enhancement Using Non-Rigid Semantic Patches

Ester Hait and Guy Gilboa, *Member, IEEE*

***Abstract*— We propose a new way to solve a very general blind inverse problem of multiple simultaneous degradations, such as blur, resolution reduction, noise, and contrast changes, without explicitly estimating the degradation. The proposed concept is based on combining semantic non-rigid patches, problem-specific high-quality prior data, and non-rigid registration tools. We show how a significant quality enhancement can be achieved, both visually and quantitatively, in the case of facial images. The method is demonstrated on the problem of cellular photography quality enhancement of dark facial images for different identities, expressions, and poses, and is compared with the state-of-the-art denoising, deblurring, super-resolution, and color-correction methods.**

***Index Terms*— Prior-based image quality enhancement, similarity measures, non-rigid registration, denoising, deblurring, super-resolution.**

I. Introduction

IN THIS paper we propose a new way to solve the following very general and challenging blind inverse problem:

$$f = T(g) + N(g), \tag{1}$$

where f is the degraded input image and g is the unknown original image to be recovered. T is an unknown complex degradation transformation, which may include multiple degradations: resolution reduction, blur and contrast and color changes. The degradation T can be spatially varying and may include nonlinearities, so it cannot be modeled by a convolution kernel. N is noise, which can also be of various characteristics; It may be signal-dependent and with spatially-varying statistics. Thus a parametric model is very hard to establish for this general case. Our main assumption is that the degradations are *structure-preserving*, such that significant edges and structures are retained. This assumption will be made more formal hereafter. As problem (1) is highly challenging, it was not frequently tackled in image processing; it is extremely ill-posed and cannot be solved without additional strong priors or external data.

In the past decades, handling common image flaws has gradually improved with the use of more sophisticated image priors and models. Early methods used pixel-based statistics, such as smoothness [1], piecewise smoothness [2], total-variation [3], pixel correlation [4], or wavelet decomposition [5] for image reconstruction. In recent years, nonparametric patch-based methods, such as Nonlocal Means [6] and BM3D [7], exploited local and nonlocal self-similarities. Other patch-based, training-based methods were using Markov Random Fields [8] and dictionary learning [9].

In recent years, using *generic* image priors has started to reach an optimality bound; e.g., for super-resolution [10] and denoising [11]. Therefore, problem-specific priors were used to solve specific problems. For example, *facial priors* were used for facial image processing, such as face hallucination [12]; face compression using K-SVD [13]; or face deblurring, using the most similar different-identity face example [14]. Despite this progress, today's main state-of-the-art methods are still based on square patches with little if any semantic context [15]–[17]. We follow the concept of using problem-specific priors, but propose an alternative concept of using large, highly-semantic non-rigid patches.

We show how such a hard problem can be solved, without explicitly estimating the degradation, by using suitable prior data and non-rigid registration, which is robust enough to T and N. We illustrate our algorithm on facial data quality enhancement, where the model and underlying assumptions are valid (Fig. 1). We use a mechanism which is invariant to low-to-moderate quality reductions to solve the problem indirectly. Given today's highly available mobile photography devices, our model assumes available *high-quality personal priors*, but *no knowledge of the degradation model*. We use *non-rigid* processing of *semantic patches* of facial features, while preserving structure and context coherency. We also assume that no matches of high quality (HQ) and low quality (LQ) data are available for learning. As there is no degradation model, one also cannot faithfully generate LQ images by degrading HQ images (e.g. adding noise to a clean image).

We can conceptually describe our algorithm as an approximation of the following mathematical formulation. Let f, g, T, N be as described in problem (1), and let u be the recovered image. $u, f, g, v \in X$ belong to an image space X (we may handle images or semantic image patches).

Let $V = \{v_i\} \subset X$, $i = 1..M$, denote the space of HQ data-specific priors. Let $D(q, w) : X \times X \to X$ denote a non-rigid registration from image q to image w, associated with a displacement field $\Phi_{q,w}$, such that $\Phi_{q,w} \circ q = D(q, w)$. As described in problem (1), we denote the degradation transformation as T, and assume it to be unknown and complex, yet **structure-preserving**. We assume the displacement field

Manuscript received January 5, 2017; revised March 9, 2017; accepted March 15, 2017. Date of publication March 22, 2017; date of current version April 11, 2017. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jie Liang.

The authors are with the Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa 32000, Israel (e-mail: etyhait@campus.technion.ac.il; guy.gilboa@ee.technion.ac.il).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2017.2686003

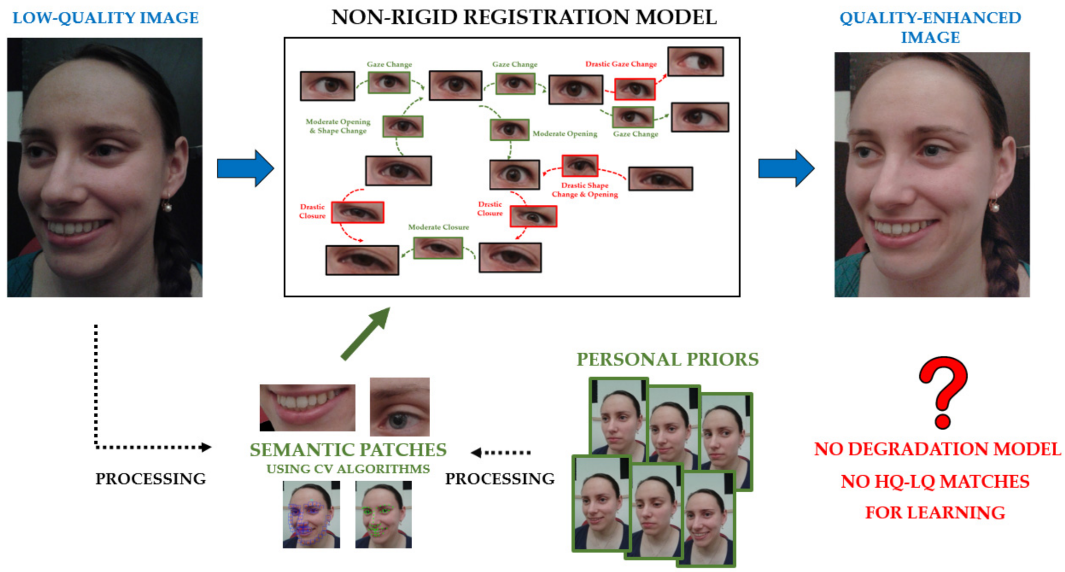

Fig. 1. Problem and assumptions of model: blind quality enhancement of LQ facial images using prior data. Semantic patches of facial features are extracted to preserve structure and context coherency. Our model assumes available HQ personal priors, but no knowledge of the degradation model and no matches of HQ and LQ data for learning.

is robust to T and N:

$$\Phi_{f,v} \approx \Phi_{g,v}, \quad \Phi_{v,f} \approx \Phi_{v,g}, \quad \forall v \in V. \tag{2}$$

Let $d(q, w) : X \times X \to \mathcal{R}^+$ be a distance-type measure, derived from the registration $D(q, w)$, such as $\|D(q, w) - w\|$, where $\|\cdot\|$ is some norm. Thus, projecting g onto the space V using d can be approximated by projecting f onto V:

$$\underset{v \in V}{\operatorname{argmin}}\, d(g, v) \approx \underset{v \in V}{\operatorname{argmin}}\, d(f, v). \tag{3}$$

This structure-preserving property under moderate complex degradations will allow us to recover $u \approx g$, **without explicitly finding T**. We therefore define $v^*(f)$ as the image in V which is the most similar to f, in the sense of the measure d:

$$v^*(f) = \underset{v \in V}{\operatorname{argmin}}\, d(f, v). \tag{4}$$

We can thus accurately select relevant prior information, while preserving structure (thus preserving expressions in the case of facial images).

To recover the accurate structure of u, the selected prior v^* must be properly non-rigidly registered to the desired structure of f:

$$u = D((v^*(f), f)). \tag{5}$$

Thus u is an approximated non-degraded version of the distorted image f, which practically minimizes a fidelity term, **without explicitly finding** T, N:

$$\|f - T(u)\|. \tag{6}$$

For example, let us assume an optimal case where $g \in V$. Therefore from (3) and (4) we can assume with high probability that $v^*(f) = g$. Thus

$$u = D(v^*(f), f) = D(g, f) \approx g.$$

The rest of the paper is organized as follows. We first present the related work and our resulting insights, as well as our contributions and novelties (Sec. I). We then methodologically explore the non-rigid Demon registration [18] to show its robustness to a variety of degradations and to non-rigid variations, and thus its compliance to this concept (Sec. II). Relying on this, we show an end-to-end application, approximating our concept for the problem of blind quality enhancement of facial images (Sec. III). We then demonstrate our experimental results for dark cellular facial images, and discuss drawbacks and future work (Sec. IV). Sec. V concludes our work.

A. Related Work

Capel and Zisserman [19] observed that better learning is obtained when considering different facial regions, rather than the whole face, and that better representation is needed when handling high-detail facial regions that attract human attention. A separate PCA basis was learned for different key facial regions. Unlike our proposed method, they used linear PCA decomposition and training sets of *multiple* people.

Jia and Gong [20] performed face hallucination of a single modality (expression, pose and illumination) into a set of high-resolution (HR) images of different modalities, but used multiple people's images as priors. Interestingly, they deduced that hallucinating the same expression as in the test image was better than hallucinating other expressions.

Lee *et al.* [21] represented multiple-pose facial images as a low-dimensional appearance manifold in the image space, for video face recognition. The appearance manifold, learned from training, consisted of pose manifolds and their connectivity matrix, encoding transition probabilities between images.

Yu *et al.* [22] incrementally super-resolved 3D facial texture from video under changing light and pose, but used temporal information from sequential frames and a generic 3D face model. They also handled facial non-rigidity using a local region-based approach: using a match statistic to detect significant facial regions expression changes between frames.

Shih *et al*. [23] performed noise level estimation for denoising, by maximizing the joint noise probability across same-identity facial images of different noise levels. The estimated noise level can then be used for state-of-the-art denoising algorithms requiring it, such as BM3D.

Joshi *et al*. [24] were the first to suggest the use of "personal priors" to enhance the quality of a particular person's image, performing both global and face-specific corrections. They relied on the growing available amount of personal images. Their algorithm derived its strength from using multiple same-identity example images, which, as they observed, can span a smaller space than that spanned by images of multiple people.

They performed global corrections of non-facial regions (such as deblurring, color and exposure corrections) using mean and basis vectors generated using PCA decomposition (of different image layers), to derive priors for MAP estimation. They also performed local corrections of face regions (hallucination for sharpening; or inpainting for exposure correction), by transferring desired properties from HQ images in the gradient domain, using the Poisson equation.

The major drawback of this algorithm is its simplistic model which can address only frontal images with little expression variations and large non-facial regions. We wish to focus on a more high-quality quality enhancement of *facial* regions, and handle a variety of *subtle* expression variations.

Following this, Loke *et al*. [25] suggested to super-resolve very LR facial images by selecting a set of the most similar HR same-identity training images, in the sense of pose and expression. A similarity measure, based on pose estimation and an expression descriptor, relying on shape and texture, was used for selection. After aligning the selected images using triangulation and affine warping, patches of them were used to hallucinate the face using a MRF model, based on color and edge constraints and a smoothness term.

Drawbacks of this work include the selection process, based on a rough match of some facial regions to the query; we wish to handle more *subtle* expression variations. Replacing LR patches with HR ones results in noticeable artifacts, seams and change of color, since this patch-based method does not account for the human observer's sensitivity to certain facial regions and their expressions. Other drawbacks are using a very large HR dataset (thousands of images), their small size, and the manual labeling of feature points in the LR image.

B. Insights

Previous works and early experiments point out important insights regarding facial images of a specific individual.

- Observing the impact of facial expression variations, we note that the *non-rigid* behavior of faces and facial features under such variations requires *non-rigid* registration.
- The space spanned by same-identity facial images, depicting a limited range of expressions, is significantly smaller than that spanned by multiple-identity images [24]. Using *generic* faces as priors, on the other hand, introduces artifacts and possible changes in identity and expression.
- A change in identity or expression is visually disturbing to a human observer; Thus, only the most structurally similar examples should be used for reconstruction (which can also be deduced from [20], [24], [25]).
- Better learning is obtained when considering different facial regions, rather than *the whole face* [19]. Artifacts may appear when using a *patch-based* method, which does not consider human sensitivity to certain facial regions and expressions [25]. Better representation is also needed when handling high-detail facial regions, such as eyes, compared to smooth regions, such as cheeks [19].
- We observe that decomposing the face into facial regions increases the versatility in generating a variety of expressions (combining different eye expressions and mouth expressions), while decreasing the number of samples required. This allows to construct and search datasets of small yet semantic facial regions, rather than large whole-face images, saving both memory and computation time.

C. The Proposed Method

In our work we use personal priors to enhance the quality of facial images of a particular person. We obtain new data-driven facial features spaces, based on only tens of HQ, same-identity, same-pose examples, differing in facial expression; and define a new affinity measure to match them to given LQ queries.

For two key facial features and for different head poses, we construct HQ, identity-specific affinity spaces, representing various "principal modes" of the feature (different structures and expressions). We define a new affinity measure for image matching under non-rigid variations, which derives from the distance between images, in the sense of the non-rigid transformation [18] required to register them.

This measure corresponds to how natural, real-world interpolated images appear to a human observer, or their "visual validity". The registration can thus interpolate real-world looking images, that can expand the affinity space. It also provides a useful tool for fine registration of non-rigid facial features.

The measure's robustness to quality degradation and non-rigid variations enables to accurately match a query feature to the most similar example from the suitable affinity space. HQ information selected is then registered and embedded into the LQ image to obtain a HQ facial image.

D. Contributions and Novelties

1) We solve an extremely difficult inverse problem of simultaneous recovery of multiple degradations, with no standard solutions in the literature. We assume a complex, blind degradation model; thus, one cannot generate LQ images by degrading HQ ones; also no matches of such data are available for learning.
2) We integrate non-rigid registration as a tool for image quality enhancement. Defining a new registration-based affinity measure to search suitable data, its robustness to moderate quality degradations and non-rigid variations allows accurate matching.

3) We illustrate our model and present an end-to-end solution to a real-world problem arising from the common scenario of cellular shooting in the dark.
4) We assume that personal priors, that is facial images of the same person, are at hand, given today's massive amount of personal images on personal image collections. There is very little previous work using this prior; we present novel models and concepts to solve it.
5) We use non-rigid semantic patches. We take into consideration the sensitivity of human observers to unique facial regions and expressions, and their non-rigid nature. We thus use semantic patches of adaptive size and location, but of coherent structure and context.

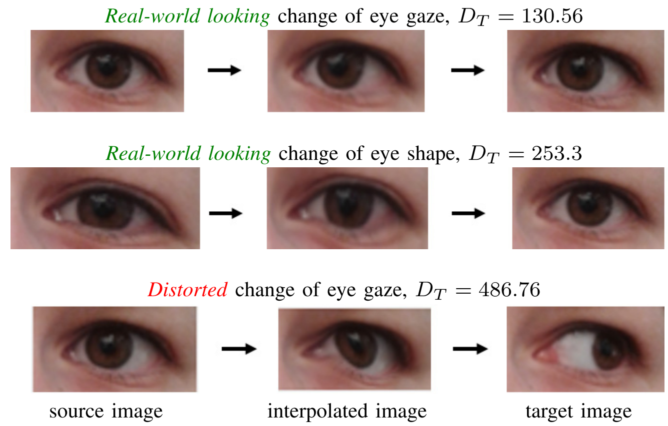

Fig. 2. Examples of deformation paths between HQ same-identity eye images, and their Demon measures. Compare the **low** Demon distances for the *visually valid* paths to the **higher** distance for the *visually non-valid* path.

II. Registration-Based Affinity Space

A. Demon Diffusion-Based Non-Rigid Registration

We use the Demon registration as a registration tool for our quality enhancement model. Demon registration, first introduced by Thirion [18], [26], describes the gradual diffusion of an object, represented by a deformable grid, into another object, represented by a semi-permeable membrane, through its boundaries, by the action of Demon effectors.

Thirion showed the translation of this concept into a simple gradient-based displacement field $\vec{u}$ from the *moving image* m to the *static image* s. Further improvements suggested by Wang *et al.* [27] and Cachier *et al.* [28] yield the following:

$$\vec{u} = (m-s) \times \left(\frac{\vec{\nabla} s}{|\vec{\nabla} s|^2 + \alpha^2 (s-m)^2} + \frac{\vec{\nabla} m}{|\vec{\nabla} m|^2 + \alpha^2 (s-m)^2} \right), \quad (7)$$

where $\vec{\nabla}$ denotes image gradient, and α is a normalization factor accounting for adaptive force strength adjustment.

This registration method was so far usually used for medical image registration (e.g. [29], whose implementation we use). A detailed explanation is given in Appendix A.

We now explore this registration tool and show it meets the following requirements to be used in our non-rigid concept:

- Robust to moderate non-rigid structure variations; relating to local structure (rather than global color statistics).
- Robust to moderate multiple blind quality degradations.
- Interpolating visually valid and non-valid intermediate phases between principal modes, in correlation with a distance measure between them.
- Fast; easy to understand and implement.

Naturally, this does not exclude other registration methods, which may be used alternatively for the same concept. This may be an interesting direction for future work. We give elaborated comparisons of the Demon registration and derived Demon measure to other registration and matching methods in Appendix A and Figs. 28, 29.

B. Registration-Based Affinity Measure

We aim to characterize the sequence of intermediate images generated during Demon diffusion (the "deformation path"), using the newly-defined concept of "visual validity": how natural, real-world and undistorted the path appears to a human observer (Fig. 2). We thus define an affinity measure and show its resulting high correspondence to this visual validity.

Fig. 2 shows examples of visually valid and visually non-valid deformation paths between same-identity HQ eyes. Moderate, interesting variations in gaze, shape or closure between source and target images result in visually valid interpolated images: naturally-looking, real-world facial features. However, drastic changes result in visually non-valid images: distorted images that cannot be considered as real-world features.

To quantitatively characterize a deformed image by its visual validity, we define a new Demon-based affinity measure (Eq. (8)) between images under non-rigid variations, derived from the transformation required to register them. We use it as our matching criterion $d(u, v)$ for selecting structurally-similar principal modes of a facial feature (recall Sec. I, Eq. (4)).

The distance measure between two images is proportional to the mean absolute error between the *deformed* image m at a *fixed* time point T in the registration process (taken in our implementation as 200 iterations), and the target image s:

$$D_T(m, s, \alpha_l) = C \| m_{HSV,\alpha_l}(T) - s_{HSV,\alpha_l} \|_{L_1} \quad (8)$$

Where α_l indicates feature-dependent HSV color space channel selection: hue channel for mouths, value channel for eyes. Intuitively, it is a measure of the distance "left to go" from m to s; taking into consideration not only their naïvež pixel-to-pixel similarity, but also Demon's *ability* to successfully deform one into another *in a given time* (as opposed to Cachier's minimization criterion, see Appendix A). It relates to local structure and shape, as opposed to histogram distances or EMD, relating to *global*, non-spatial color information. Differentiation between distances corresponding to visually valid and non-valid images was empirically found to be somewhat better using MAE rather than MSE. Better differentiation was also achieved using HSV rather than other color spaces.

Fig. 2 shows how *visually valid* deformation paths correspond to lower Demon distances; whereas *visually non-valid* deformation paths correspond to higher Demon distances. Mouth images of a certain identity behave similarly. Note the

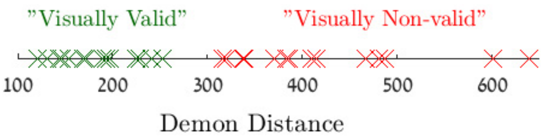

Fig. 3. Correspondence between visual validity and the Demon distance (for HQ, same-identity eye images). Compare the **lower** Demon distance for the *visually valid* paths to the **higher** distances for the *visually non-valid* paths.

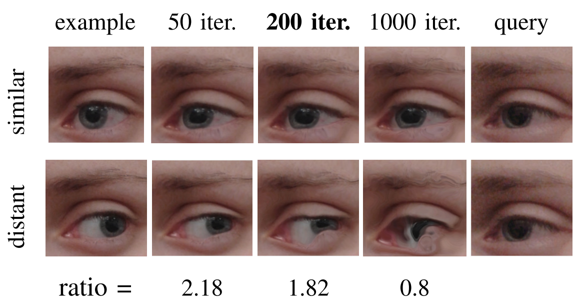

Fig. 4. Effect of time parameter T: As T increases, the distance of a HQ example (source) to a LQ query (target) decreases; The ratio between distances, when comparing distant vs. similar examples, becomes smaller, thus making differentiation less reliable. We thus choose a relatively small T. However, a too small T (50 iter.) might not allow significant deformation yet.

similar behavior shown in Appendix B when deforming synthetic images: for moderate variations, deformation succeeds and the measure *moderately* increases with variation. But for more drastic variations, the deformed image becomes too different or distorted; and the measure *drastically* increases.

Fig. 3 shows the correspondence between visual validity and the Demon distance for HQ, same-identity eye images. Visual validities were determined using the concept described above. **Lower values of the Demon distance correspond to visually valid deformation paths, and vice versa**. Thus the measure better reflects human visual judgment of visual validity, allowing automatic differentiation between different visual validity categories. Note that we later use a nearest-neighbor scheme to choose the most relevant patch, so no actual threshold of validity is needed.

A visually valid deformation path allows using interpolated images as new naturally-looking intermediate phases of subtle variations between existing principal modes, thus potentially expanding the dataset. The measure cannot be considered as a distance or metric in the mathematical sense, as a triangle inequality cannot be shown. Fig. 4 shows the effect of the time parameter T: as T increases, distances decrease. When comparing the distances of a LQ query to a distant HQ example, vs. a similar example, then the following ratio between distances:

$$\frac{D_T(example_1, query) - D_T(example_2, query)}{D_T(example_2, query)}$$

becomes smaller as T increases, thus making differentiation less reliable. We thus choose a relatively small T (also to save running time).

Illumination consistency between images has much influence on Demon registration distortion (Fig. 5). Registering images similar in shape and structure, but differing

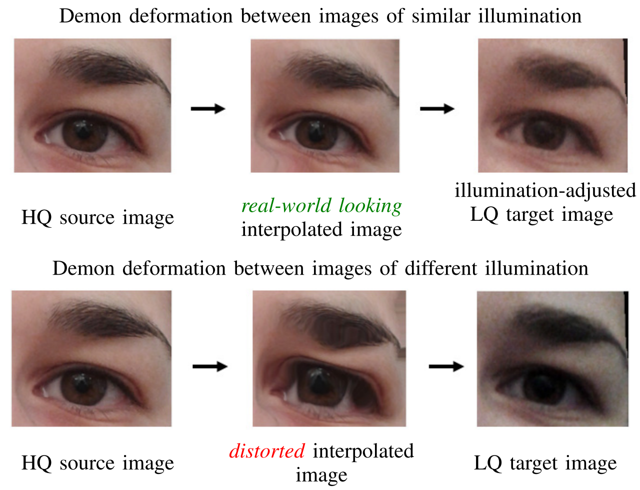

Fig. 5. Demon deformation of similar-structure HQ to LQ images, and the effect of illumination adjustment: the naturally-looking, HQ image obtained when deforming same-illumination images (top); compared to the distorted image obtained when deforming different-illumination images (bottom).

in illumination, results in a distorted interpolated image, whereas a naturally-looking, same-structure, HQ result is obtained when first performing illumination adjustment (we use the simple histogram equalization). This example also illustrates the two useful qualities of this registration: **robustness to quality degradation** and **preservation of source quality**. When deforming a HQ image into a similar-illumination, similar-structure LQ image, the deformed image is both naturally-looking and of quite high quality.

Concluding Demon's important characteristics:

1) **Correspondence between Demon measure and visual validity**: The measure relates to the registration's ability to bring one image close to another under moderate non-rigid variations, while preserving real-world appearance.
2) **Robustness to quality degradation**: This holds given consistent illumination, e.g. for noise and resolution reduction; Therefore, in HQ to LQ image registration, low distances still correspond to similar structures.
3) **Preserving source quality when registering different-quality images**: registering a HQ image to a similar-shape LQ image preserves its high quality, while adjusting to the desired shape, as can be seen in Fig. 5.

Combining these characteristics allows performing a measure-based Nearest Neighbor search to match a LQ query to the most structurally-similar HQ dataset example. It then enables *fine non-rigid, quality-preserving registration* of the HQ match to the LQ query, such that the result is naturally-looking and of desirable shape and quite HQ.

C. Registration-Based Facial Features Affinity Spaces

Fig. 8 illustrates the concept of an affinity space based on the Demon affinity measure, with a geodesic of visually valid deformation paths, depicting intermediate deformation steps between principal modes. Visually non-valid paths are not allowed, as they do not generate new naturally-looking images. Fig. 9 shows part of a real affinity space of same-identity, same-pose eyes, automatically constructed using the Demon measure. Note that the outlier of a non-frontal

Fig. 6. Examples of our personal priors image set, which includes 7 sets of different identities and poses; each consists of 20-30 same-identity, same-pose, multiple-expression HQ cellular images.

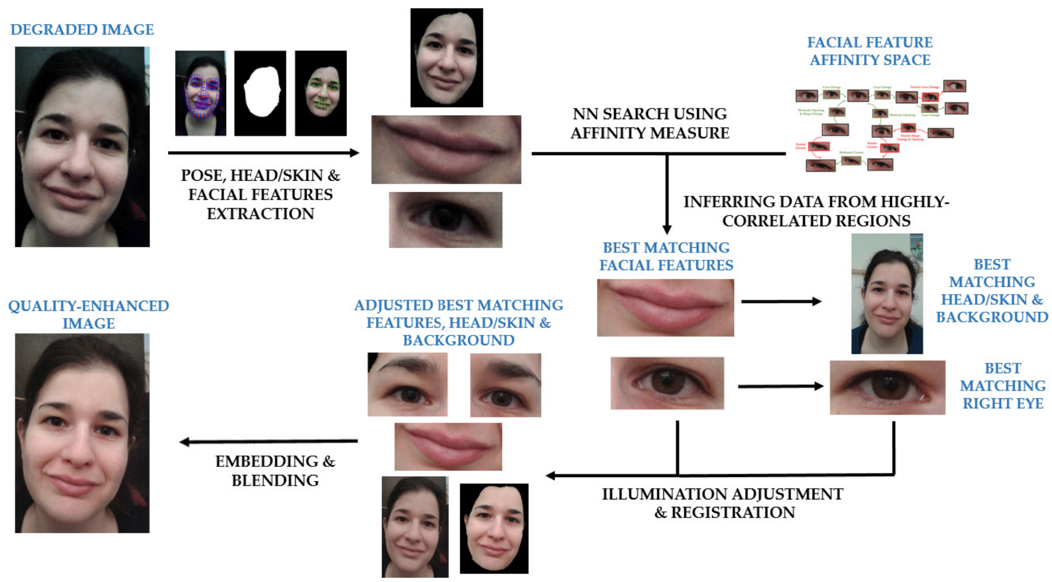

Fig. 7. Algorithm's flowchart: Facial image quality enhancement using registration-based affinity measure and affinity spaces.

eye (uppermost right), wrongly classified as frontal during preprocessing (see Sec. III-A), is unconnected to all others.

We use only tens of personal priors images to *automatically* construct 14 data-driven spaces. Each space consists of multiple (about 20-30) HQ, same-identity, same-pose examples (2 identities with 2 poses each; one identity with 3 poses) of a specific facial feature (left eye or mouth). Note that we **do not** construct spaces for **right eyes**, see Sec. III-C. Fig. 6 shows examples of the personal priors image set. As opposed to previous works, affinity spaces describe many different *subtle* expression variations, such as different eye gaze, closure and shape, or different mouth closure, shape and expression.

In the future, it might be possible to use the subset of the most similar principal modes and visually valid interpolated images between them as priors or constrains within a more generalized framework of image restoration. Another option is making the selection process more time-efficient, by performing a more sophisticated initial projection of the query onto the set, before performing a more comprehensive search.

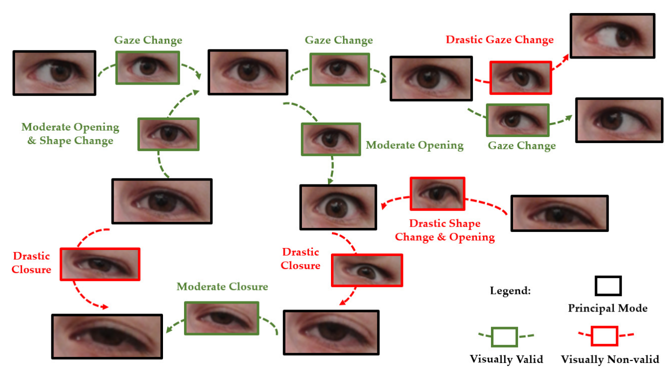

Fig. 8. Illustration of an eye affinity space, constructed based on the Demon-based affinity measure. Visually valid deformation paths and interpolated images appear in green, whereas visually non-valid ones appear in red.

III. Facial Image Quality Enhancement

Relying on the concept of non-rigid registration of semantic patches of personal priors, we present an end-to-end application which approximates this concept for semantically-aware quality enhancement of facial images, using a Nearest Neighbor search. Fig. 7 illustrates the proposed method.

The details of the algorithm are as follows:

A. Preprocessing: Facial Features Extraction

We use [30] to detect the facial contour, whose convex hull is used as input to image matting [31]. Thresholding and

Fig. 9. Example of part of a real affinity space of same-identity, same-pose eyes, automatically constructed using the Demon affinity measure.

Algorithm 1 Facial Image Quality Enhancement

Data: A Degraded facial image (blind degradation model), HQ personal priors.

Result: A quality-enhanced facial image.

1 Extract facial features, select relevant HQ patch affinity spaces according to identity and pose. See Sec. III-A;

2 For each semantic patch (eye, mouth): select the most similar HQ patch in the space, using illumination adjustment and the Demon measure. See Sec. III-B;

3 Infer data regrading other highly-correlated facial regions. See Sec. III-C;

4 Embed high-quality image details, using registration, color-correction and blending. See Sec. III-D;

erosion of the resulting mask (similar to the preprocessing in [24]) result in a more accurately detected head image, later used to extract head & skin information. We also use their pose estimation, to later search the suitable (same pose-sign) affinity space. For a more accurate facial landmarks localization we prefer using [32], with the head image as input. **Note, that HQ images are similarly processed for pose estimation, features and head extraction, to construct the HQ spaces.**

B. NN Search Using the Affinity Measure

We conduct Nearest Neighbor searches through suitable (same pose-sign) affinity spaces. The Demon affinity measure is used as a **matching criterion** to find the HQ dataset example which best matches a given LQ query. This is allowed due to the measure's robustness to image quality degradation. *Note, that this search does not require knowledge of the connections between dataset images, or paths' visual validities*. Throughout the search, **illumination adjustment** (using histogram equalization) is performed prior to distance calculation.

Performing registration on only one image channel reduces running time: using a Matlab code and the computing resources described in Sec. IV, for T=200 iterations, each registration takes about 1 sec. Using *parallel computing*, performing NN over a 20-examples space takes about 6.3 sec. Relying on [24], we assume that for a certain pose, only a limited variety of facial expressions is possible, which can be spanned using about 20-30 same pose-sign examples. For larger pose variations, due to inaccuracies in pose estimation, more accurately differentiating between poses will require better preprocessing; Following that, tens of examples will be needed to handle each pose, at the same computational cost.

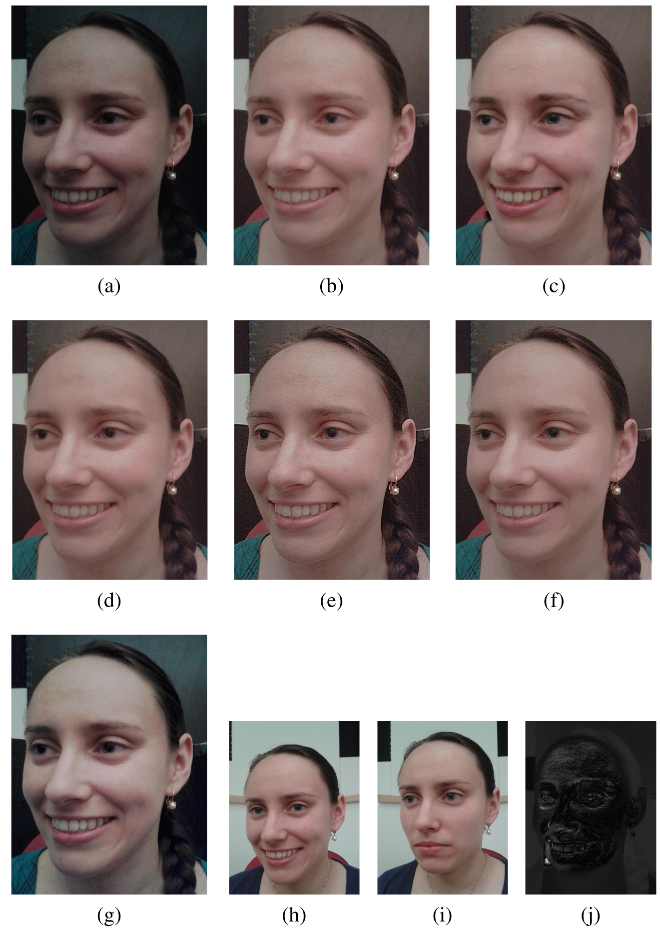

Fig. 10. Quality enhancement example (1020X768): identity #2, right pose. (a) LQ input image, NIQE score=1.2032. (b) Brightened input using NRDC [33], NIQE score=1.1058. (c) Proposed method, NIQE score=1.0187. (d) BM3D Denoising of (b) using [7], NIQE score=1.6097. (e) Deblurring of (b) using [36], NIQE score=1.1346. (f) Deblurring of (b) using [37], NIQE score=1.2835. (g) Tone enhancement of (a) using [38], NIQE score=1.227. (h) HQ example for mouth, head & illum. info. (i) HQ example for eyes info. (j) Difference image: (b) to (c).

C. Inferring Data From Highly-Correlated Regions

We use structure and context correlations between semantically meaningful face regions to infer further suitable data. To select the proper *right* eye, based on the *left* eyes space, we make the reasonable assumption of gaze, closure and shape consistency between both eyes (ignoring cross-eye and winking). This allows to extract the suitable right eye from *the same HQ image from which the left eye example was taken*. Thus, avoiding the need to construct a right eyes space.

Another use of facial semantic structural constrains regards head and skin information. In general, the shape of middle-low facial regions (cheeks, chin, facial lower contour and even nose) depends on the mouth expression, but not on the

Fig. 11. Quality enhancement example (1050X658): identity #1, frontal pose. (a) LQ input image, NIQE score=1.4501. (b) Brightened input using NRDC [33], NIQE score=1.2203. (c) Proposed method, NIQE score=1.1858. (d) BM3D Denoising of (b) using [7], NIQE score=1.9304. (e) Deblurring of (b) using [36], NIQE score=1.2937. (f) Deblurring of (b) using [37], NIQE score=1.4815. (g) Tone enhancement of (a) using [38], NIQE score=1.2499. (h) Difference image: (b) to (c). (i) No head reg. (j) No blending.

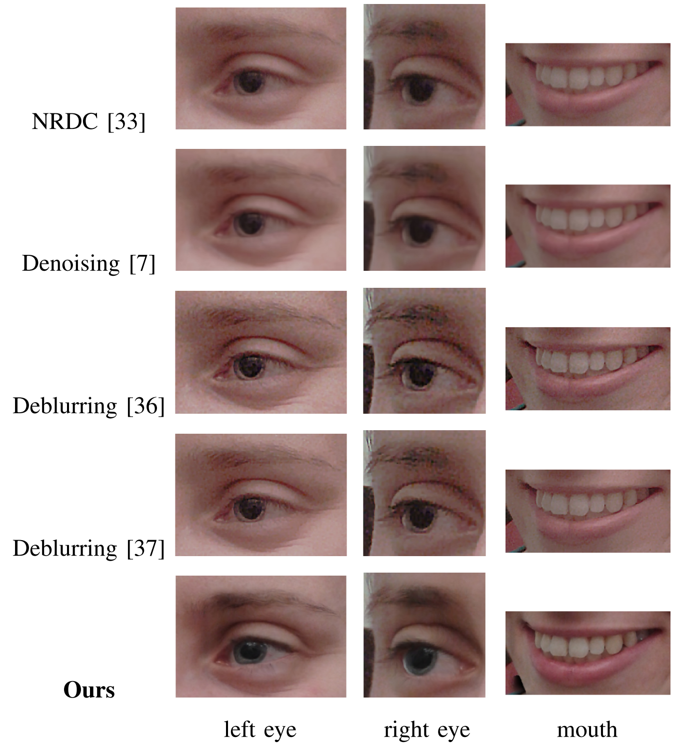

Fig. 12. Quality enhancement example: identity #2, right pose, close up.

eye expression. Thus it is only reasonable to use *the same HQ image from which the mouth was taken* to also extract head & skin information. The HQ example image is used as reference for illumination adjustment and for extracting the head/skin.

D. Embedding High-Quality Image Details

The input image and input head/skin undergo illumination adjustment to the brighter illumination of the selected HQ example image and HQ head/skin, respectively, using the NRDC algorithm [33]. Due to its randomized nature (using Generalized PatchMatch), we choose out of several repetitions the best illuminated HQ example image, *in the sense of its NIQE score* [34] (see Sec. IV). The example head/skin undergoes affine registration to best fit the input. As to the facial features, input features undergo illumination adjustment to the brighter HQ illumination. Then, example features undergo fine non-rigid registration to best fit the input feature structure. Finally, we embed the HQ head/skin and facial features into the brightened input using blending [35], to produce a seamless, smooth appearance.

IV. Experimental Results

We now show experimental results for our prior-based quality enhancement algorithm.[1]

A. Experimental Setup

Input facial images were taken using a SAMSUNG GT-S7580L cellular frontal camera (2560X1536 resolution) in a dark environment. We assume no known information regarding the camera's specifications and built-in image processing algorithms. Thus, this real life scenario of dark environment cellular shooting demonstrates well common flaws, such as noise post-processed by unknown (possibly nonlinear) filtering, slight motion blur, resolution reduction of unknown parameters, and changes of contrast and color (for recent works on tone and contrast enhancement, see [38], [40], [41]). Prior images were taken using the same cellular camera in an indoor well-illuminated environment. Prior and input images were downsampled by a factor of 2 before processing. Final results were cropped to focus on the face region. Using an unoptimized Matlab code with Matlab/C++ code segments, on a Windows 7 OS, Intel i7-4770 CPU at 3.4 GHz with 16GB RAM, the running time using a single NRDC-NIQE iteration was 3 minutes; running time was 4 minutes when using 5 iterations.

B. Visual and Quantitative Results

We demonstrate our results for multiple identities, poses and expressions. Figs. 10 to 16 display visual and quantitative

[1]Further results can be found at: https://visl.technion.ac.il/research/etyh/

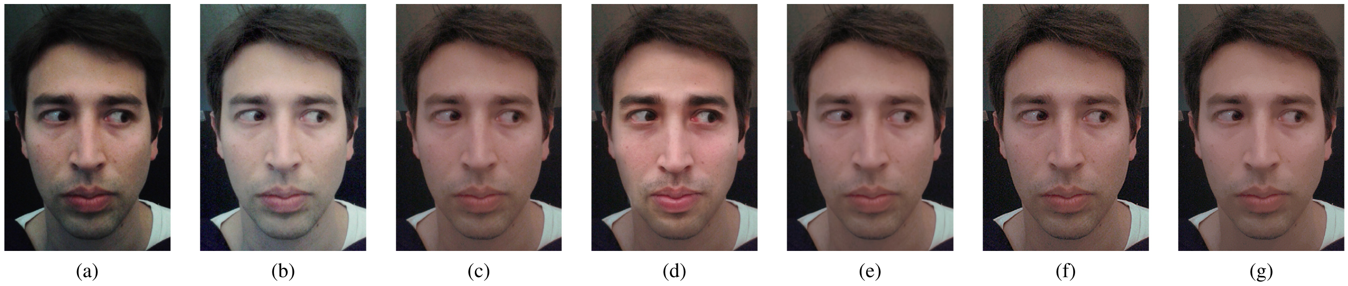

Fig. 13. Quality enhancement example (950X640): identity #3, frontal pose. (a) LQ input image, NIQE score=1.2442. (b) Tone enhancement of (a) using [38], NIQE score=1.1663. (c) Brightened input using NRDC [33], NIQE score=1.2274. (d) Proposed method, NIQE score=1.1107. (e) BM3D Denoising of (c) using [7], NIQE score=1.7819. (f) Deblurring of (c) using [36], NIQE score=1.2198. (g) Deblurring of (c) using [37], NIQE score=1.4076.

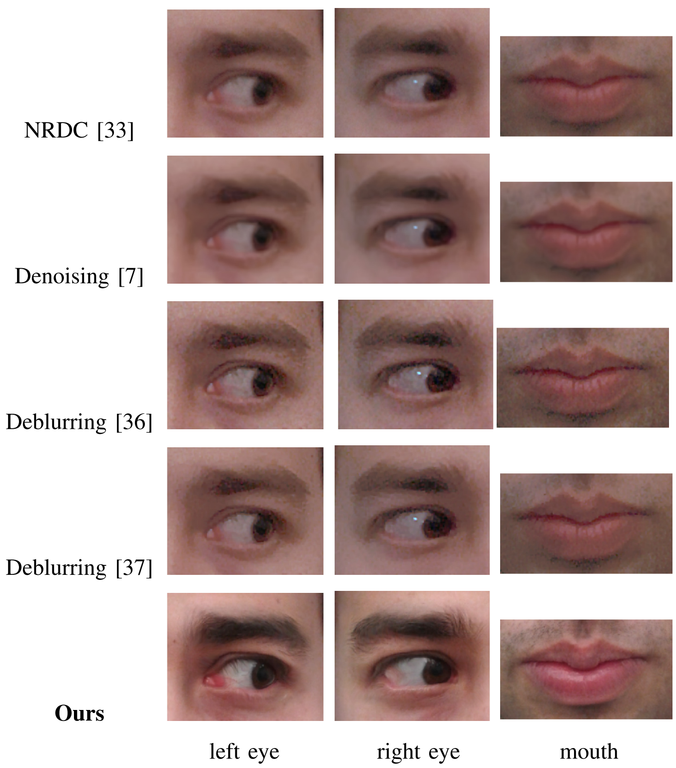

Fig. 14. Quality enhancement example: identity #3, frontal pose, close up.

comparisons of our results to multiple different methods. We compare our results to the degraded input image; tone and contrast enhancement using entropy maximization and quantization resolution upconversion [38]; a prior-based brightened version, using NRDC illumination adjustment [33]; and three state-of-the-art quality enhancement methods (using their default parameters): BM3D color denoising [7] (assuming AWG noise of std=10); blind deblurring, using a coupled adaptive sparse prior [36]; and blind deblurring using a dark channel prior [37]; All performed on the brightened image.

Figs. 19 to 22 compare our results visually and quantitatively to a state-of-the-art super-resolution algorithm using sparse representation [39] (using default parameters). First, we downsample and upsample the input (by a factor of 4, using bicubic interpolation) to get the new dark input ("dark LQ image"). This is used to find best matching patches and hence to adjust the illumination to generate the "initial brightened image". We now again reduce the resolution by a factor of 4, and compare the following four methods:

1) The "proposed method": bicubic interpolation ("bicubic interpolated brightened image") is used as background for HQ details embedding using our method.
2) "SR using default 1024-dictionary": SR using a pre-trained dictionary of size 1024 of natural images.
3) "SR using default 512-dictionary": SR using a pre-trained dictionary of size 512 of natural images.
4) "SR using specifically-trained 512-dictionary": SR using a specifically-trained dictionary of size 512 of facial images of the same identity and pose as the input image.

Since **no ground truth images are available**, a *no-reference* (NR) image quality assessment (IQA) measure is required. We use the NR blind-model IQA score NIQE [34] to quantitatively compare the methods. It is a "distortion unaware" measure: no learning was performed on distorted images of specific distortions. Therefore, rather than tuning to specific distortions, it measures deviations from natural image statistics of images considered to be of "good quality". It thus better suits unconstrained environments, such as our complex and blind degradation model. Having examined several NR IQA measures [42]–[44], we found NIQE the most reliable one. Only NIQE was consistent, in the sense that HQ images should get better scores than LQ or processed ones; and well-illuminated images - better scores than dark ones.

We display the NIQE scores of the processed images, normalized to the NIQE score of the HQ example image. As the NIQE score *decreases* as quality *increases*, **the closer the normalized score is to 1 - the better the quality**. It can be seen for all examples, **both visually and quantitatively, that our algorithm yields better results than all other methods.** Our results have a vivid and natural appearance; New HQ details and fine textures are embedded into the image and noise is removed, while preserving pose, expression and identity.

Table I and Fig. 17 show a comprehensive quantitative quality analysis, using the NIQE score, of **17 examples** of different identities, poses and expressions, comparing our method; the degraded input; the brightened image; BM3D denoising; and the two deblurring methods. Recall that the closer the normalized score is to 1 - the better the quality.

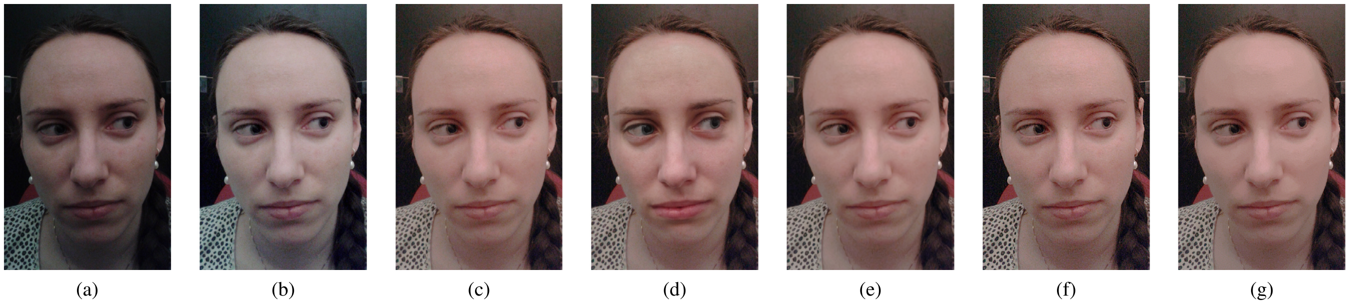

Fig. 15. Quality enhancement example (1130X710): identity #2, frontal pose. (a) LQ input image, NIQE score=1.3085. (b) Tone enhancement of (a) using [38], NIQE score=1.2987. (c) Brightened input using NRDC [33], NIQE score=1.1315. (d) Proposed method, NIQE score=1.091. (e) BM3D Denoising of (c) using [7], NIQE score=1.6522. (f) Deblurring of (c) using [36], NIQE score=1.1118. (g) Deblurring of (c) using [37], NIQE score=1.1753.

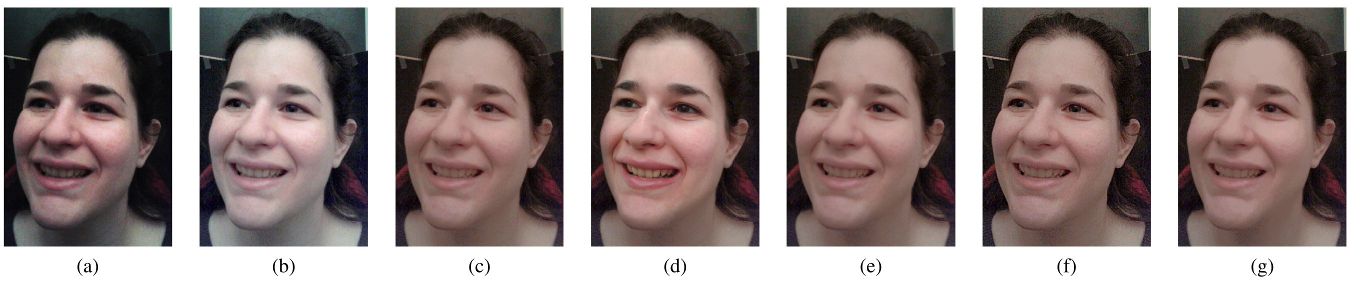

Fig. 16. Quality enhancement example (950X668): identity #1, right pose. (a) LQ input image, NIQE score=1.4289. (b) Tone enhancement of (a) using [38], NIQE score=1.3957. (c) Brightened input using NRDC [33], NIQE score=1.2712. (d) Proposed method, NIQE score=1.1921. (e) BM3D Denoising of (c) using [7], NIQE score=2.0071. (f) Deblurring of (c) using [36], NIQE score=1.2687. (g) Deblurring of (c) using [37], NIQE score=1.5122.

TABLE I

NORMALIZED NIQE QUALITY ASSESSMENT FOR 17 EXAMPLES: OUR METHOD COMPARED TO DENOISING AND DEBLURRING METHODS

	Proposed method	**Brightened image [33]**	**Deblurring [36] of brightened image**	**Input image**	**Deblurring [37] of brightened image**	**BM3D Denoising [7] of brightened image**
Average score	**1.1618**	1.2362	1.2927	1.3708	1.5316	1.8922
Relative imp. over input	**+15.24%**	+9.82%	+5.7%	-	-11.73%	-38.04%

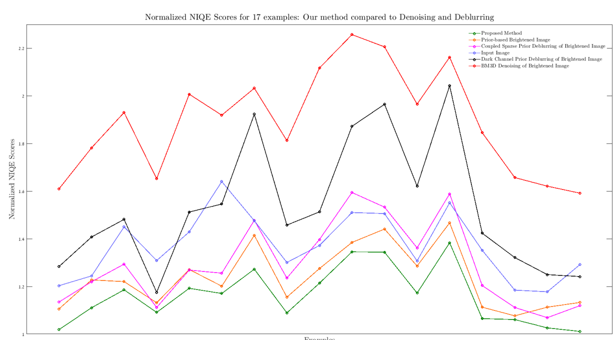

Fig. 17. Normalized NIQE scores for 17 examples: Our method compared to Denoising and Deblurring methods. Recall: the closer the normalized score is to 1 - the better the quality. **Our method outperforms all other methods.**

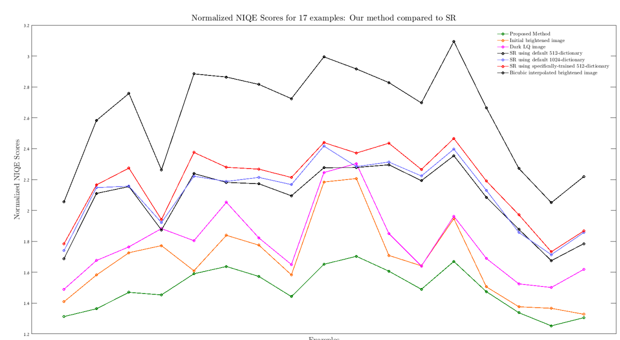

Fig. 18. Normalized NIQE scores for 17 examples: Our method compared to SR methods. Recall: the closer the normalized score is to 1 - the better the quality. **Our method outperforms all other methods.**

For all examples we can see that **our method outperforms all other methods**. One should note though that these methods were not originally designed to handle such complex, blind quality degradation, but standard synthetic degradations (e.g., BM3D works well for AWGN with known noise variance).

Table II and Fig. 18 show a comprehensive quantitative quality analysis, using the NIQE score, of **17 examples** of different identities, poses and expressions, comparing our method and the SR methods. For all examples we can see that **our method significantly outperforms the SR methods**. We can see no visual differences when using different dictionaries, even though one can expect larger dictionaries and specifically-trained dictionaries to perform better. We assume that given our complex, blind degradation model, rather than a synthetic resolution reduction model, the

TABLE II

NORMALIZED NIQE QUALITY ASSESSMENT FOR 17 EXAMPLES: OUR METHOD COMPARED TO SR METHOD [39]

	Proposed method	**Initial Brightened image [33]**	**Dark LQ image**	**SR using default 512-dictionary**	**SR using default 1024-dictionary**	**SR using specifically-trained 512-dictionary**	**Bicubic interpolated brightened image**
Average score	**1.4901**	1.68	1.7928	2.0785	2.1148	2.179	2.6285
Relative imp. over input	**+ 16.88 %**	+ 6.29 %	-	- 15.94 %	- 17.96 %	- 21.54 %	- 46.62 %

(a) (b) (c) (d) (e) (f) (g)

Fig. 19. Comparison to super-resolution (1020X768): identity #2, right pose. (a) Dark LQ image, NIQE score=1.4893. (b) Initial brightened image, NIQE score=1.4087. (c) Bicubic interpolated brightened image, NIQE score=2.0556. (d) Proposed method, NIQE score=1.3128. (e) SR [39] using default 512-dictionary, NIQE score=1.6877. (f) SR [39] using default 1024-dictionary, NIQE score=1.7395. (g) SR [39] using specifically-trained 512-dictionary, NIQE score=1.7842.

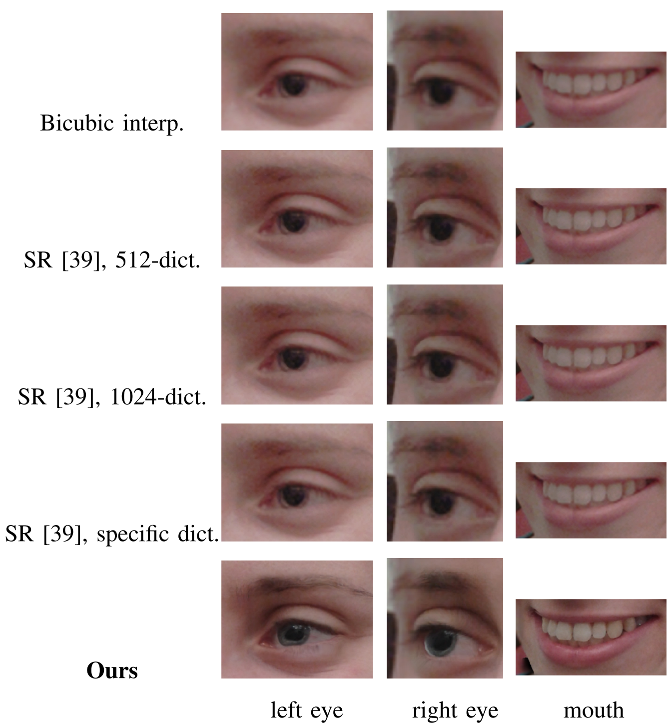

Fig. 20. Comparison to super-resolution: identity #2, right pose, close up.

algorithm's ability to handle it is quite limited, regardless of the dictionary used. This also explains why quantitative results also show no advantage using larger or specifically-trained dictionaries.

Figs. 12, 14, 20, 22 show a close-up comparison of significant facial regions (that attract human attention and convey facial expression) for different methods. Difference images between the brightened images and our results (Figs. 10, 11) show how using personal priors embeds HQ image details and fine textures, e.g. in the eyes, eyebrows, mouth and skin. Fig. 10 also displays the HQ example images used to extract prior information. One HQ image is used to extract mouth and head/skin information and for background illumination adjustment; It is similar in pose and mouth expression to the input, but is different in eye expression, background, hair and clothes. Another HQ image is used to extract eyes information.

In Figs. 11, 23 we discuss the effect of errors or omission of certain stages in the algorithm. Fig. 11 shows the necessity of the head registration and blending stages for visually reasonable results. Fig. 23 shows the effect of erroneous example facial feature selection. The Demon measure allows accurate selection. But what if the selection process resulted in errors? For example, due to insufficient expression variations in the dataset; or when skipping the illumination adjustment phase (see Fig. 5). The registration's **robustness to moderate non-rigid variations** allows it to overcome moderate selection errors, such that the interpolated features are of quite desirable shape, but somewhat distorted. However, features interpolated using very wrongly selected examples display wrong and distorted expressions.

C. Future Work

The concept of non-rigid registration-based models for image restoration, suggested in this work for faces, can be further explored in various aspects. We can consider the processing of more abstract non-facial data, such as other natural non-rigid structures, which evolve over time, or exhibit various structure variations (different "principal modes"). Other degradation scenarios can also be explored. We can also investigate this concept within a generalized framework. For example, the mathematical formulation suggested in Sec. I; or using

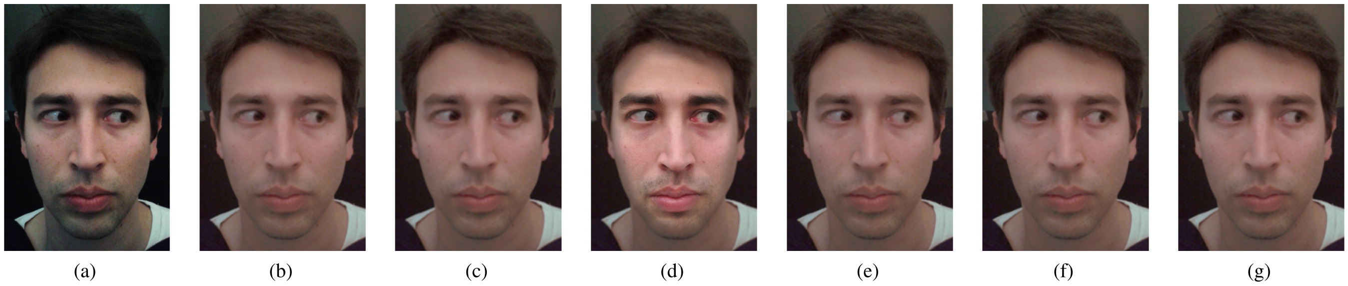

Fig. 21. Comparison to super-resolution (950X640): identity #3, frontal pose. (a) Dark LQ image, NIQE score=1.676. (b) Initial brightened image, NIQE score=1.5818. (c) Bicubic interpolated brightened image, NIQE score=2.5831. (d) Proposed method, NIQE score=1.3634. (e) SR [39] using default 512-dictionary, NIQE score=2.1103. (f) SR [39] using default 1024-dictionary, NIQE score=2.1481. (g) SR [39] using specifically-trained 512-dictionary, NIQE score=2.1656.

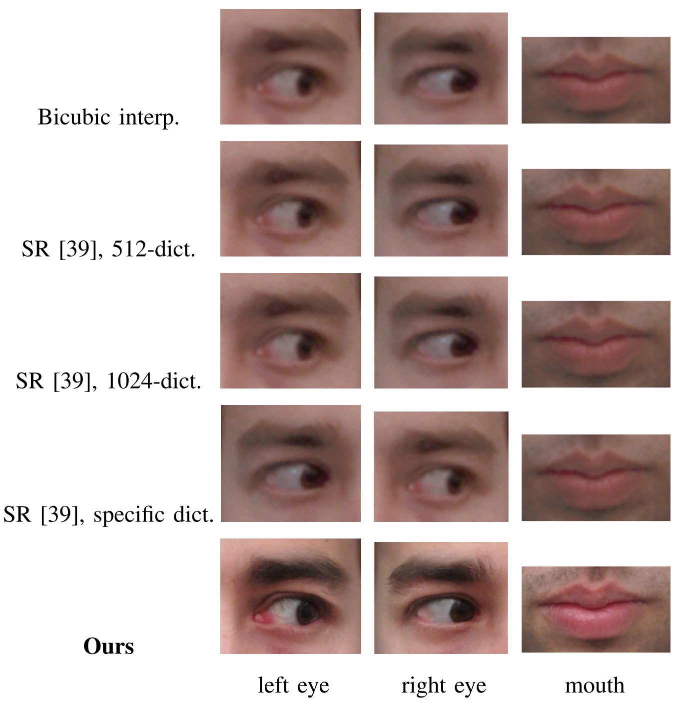

Fig. 22. Comparison to super-resolution: identity #3, frontal pose, close up.

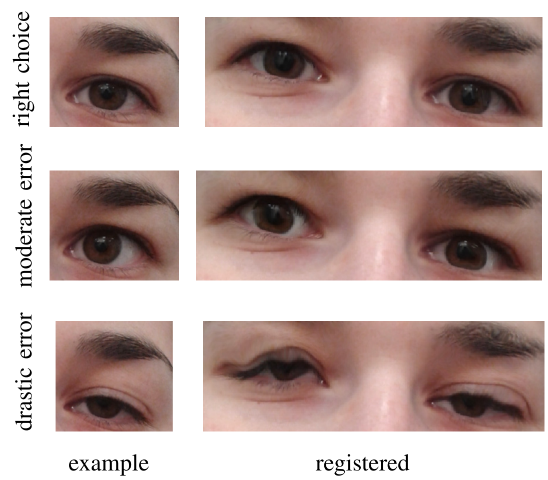

Fig. 23. The Demon measure **allows accurate selection** of example features. **But what if the selection process resulted in errors?** The registration's **robustness to moderate non-rigid variations** allows it to **overcome moderate selection errors**, resulting in features of quite desirable shape. However, more drastic errors result in distorted features.

a subset of the most suitable principal modes and visually valid interpolations as priors or constraints. Another interesting path is image quality assessment especially designed for faces.

Fig. 24 demonstrates some minor image artifacts resulting from drawbacks of our work. Note, that despite these artifacts, our results still display a natural, vivid appearance and a significant quality enhancement over the other methods, both visually and quantitatively (see Figs. 17,18).

One drawback is the inherent limitations of using the simple blending method [35] and the (rather robust) illumination adjustment method [33]. When the illumination adjustment of the input is not similar enough to the example illumination, noticeable color transitions in the facial contour may appear (second and third rows). Underlying information artifacts can also result from blending. The first row shows a dark artifact in the right eye: when the registered eye is blended into the head, some of the underlying "wrong" eye (left gaze) still shows. In the second row, some of the underlying eyebrows still appear, due to the different "eyebrows expressions", and hence their locations. These drawbacks can be addressed by improving the illumination adjustment and blending phases.

Another artifact which rarely occurred in our experiments is an inaccurate (yet naturally-looking) expression, due to inaccurate example matching. This will occur if no dataset example is structurally-similar enough to the query. The third row shows a slight change of eye expression (closure) of the right eye in our result, compared to the degraded input (recall Fig. 23 demonstrating how Demon handles possible selection errors).

This drawback can be addressed by increasing the variety of the HQ examples dataset. Other non-rigid registration methods, especially fluid-based, diffusion-based, or optimal transport based methods, may also be considered. In addition, more sophisticated and time-efficient searches can be investigated to improve matching accuracy and running times. For example, using some initial "projection" of the query onto

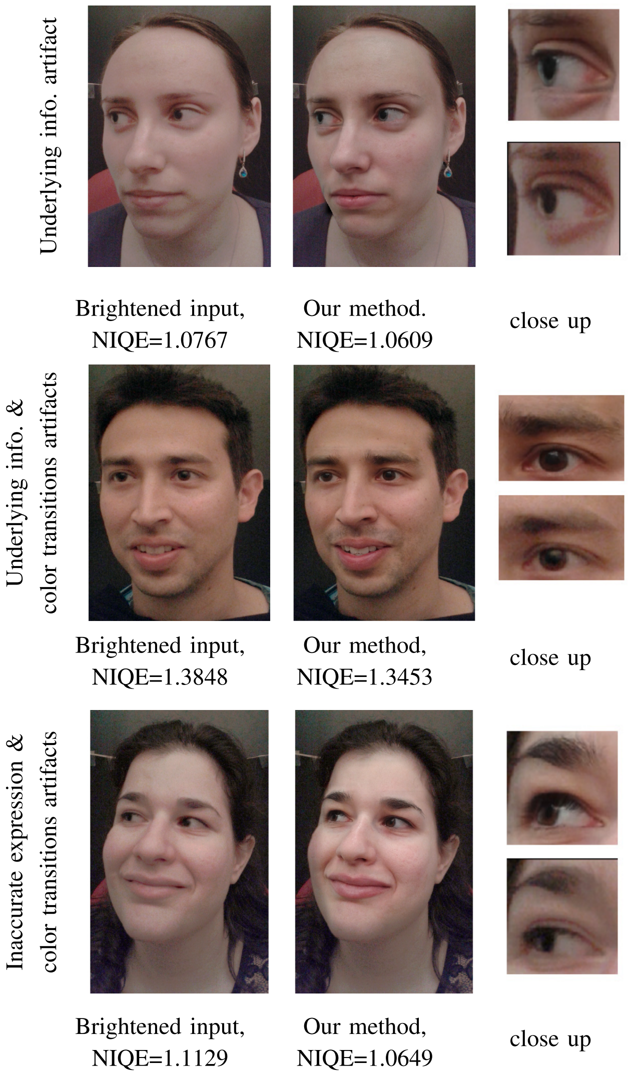

Fig. 24. Minor artifacts resulting from algorithm drawbacks. 1^{st} row: Wrong underlying information artifact due to blending. 2^{nd} row: Wrong underlying information and color transitions artifacts due to blending. 3^{rd} row: Inaccurate expression artifact due to limited dataset variety; color transitions artifact due to blending. Note that despite these artifacts, our results still outperform other methods visually and quantitatively.

the space, to first get a rough notion of the relevant subset, before performing a more comprehensive search.

V. Conclusion

In this work we propose a new way to solve a general and difficult blind inverse problem, including multiple degradations such as noise, resolution reduction, contrast and color changes. We present a novel concept for quality enhancement, combining semantic non-rigid patches of problem-specific priors and non-rigid registration. Our results demonstrate significant quality enhancement, both visually and quantitatively, for the problem of dark cellular facial images, compared to state-of-the-art quality enhancement methods.

The blind model assumption allows a very general correction mechanism which is not device and scenario dependent. Given today's easily available photography devices, our model assumes that HQ personal priors are available. We try to

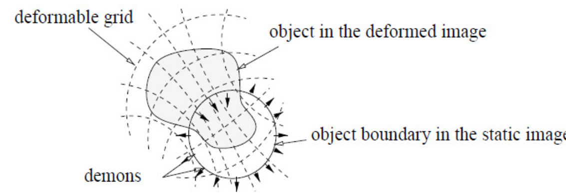

Fig. 25. Demon diffusion process [26]: the "moving" object, represented by a deformable grid, diffuses through a semi-permeable membrane representing the boundaries of the "static" object, by the action of Demon effectors.

overcome the classical processing limits by using non-rigid semantic patches and a registration algorithm, which is robust to low-to-moderate quality degradations, and can infer a HQ solution based on the priors.

Our building blocks are facial features of coherent structure and context with adaptive size and location. A new affinity measure is defined based on the non-rigid, diffusion-based Demon registration. We use it to construct data-driven, HQ facial features spaces, representing various expression variations. Its robustness to quality degradations and non-rigid variations allows accurate matches of LQ features to HQ examples. This enables significant quality enhancement, relying on only tens of personal priors, maintaining well the person's features and expressions. In a future work we consider processing of more abstract non-facial data within a generalized framework.

Appendix A
Demon Registration

Demon registration [18], [26] is a diffusion-based algorithm, approximating fluid registration, using edge-based forces.

Fig. 25 (taken from [26]) shows the Demon diffusion process. An object in the deforming image ("the moving image") is represented by a deformable grid, whose nodes are labeled "inside" or "outside"; their inner relations correspond to object rigidity. The boundaries of an object in the other image ("the static image") are represented by a semi-permeable membrane, along which Demon effectors are situated. The deformable grid gradually diffuses into the static object through its boundaries by the action of Demons. Diffusion is guided by the concept of maximal common polarity at each side of the membrane: Demon effectors act to locally "push" nodes labeled "inside" through the membrane interface into the static object, and vice versa. To this end, Demons might use spatial location, direction, pixel intensity or other information.

The final transformation results from iteratively evolving a family of transforms under two types of forces: "internal" forces, reflecting inner relations between neighboring image points, corresponding to image rigidity; and "external" forces, reflecting interaction between the static and moving images.

Fig. 26 illustrates this process, showing intermediate steps in the diffusion of an object into a same-shape translated object, until perfect registration is achieved. Fig. 27 demonstrates the difficulty in deforming a high-curvature shape into a low-curvature shape *in a given time*, and vice versa.

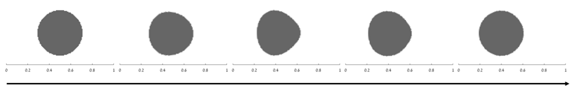

Fig. 26. Intermediate diffusion steps for object translation. From left to right: original image and intermediate images for 200, 400, 500 and 700 iterations.

(a) (b)

Fig. 27. Deformation of a circle to a square, and vice versa (for 200 iterations). Left: source image. Middle: deformed image. Right: target image. (a) From circle to square. (b) From square to circle.

Thirion [18] showed the translation of this concept into a simple gradient-based displacement field $\vec{u}$, which estimates the displacement of a pixel in the *moving image m*, required to match the corresponding point in the *static image s*.

Denoting pixel intensity as a function of time: $i(x(t), y(t), z(t), t)$, differentiating the instantaneous optical flow equation gives:

$$\frac{\partial i}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial i}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial i}{\partial z}\frac{\partial z}{\partial t} = -\frac{\partial i}{\partial t} \tag{9}$$

Considering that the evolution in one time unit is the difference between images: $\frac{\partial i}{\partial t} = s - m$, and that $\vec{u} = (\frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt})$ is the instantaneous velocity from m to s, we get:

$$\vec{u} \cdot \vec{\nabla} s = m - s \tag{10}$$

where $\vec{\nabla}$ denotes image gradient. Defining $\vec{\nabla} s$ as the internal edge-based force, and $(m - s)$ as the external force, $\vec{u}$ is computed locally as the shortest translation of a point of m onto the hyperplane approximating s:

$$\vec{u} = \frac{(m - s)\vec{\nabla} s}{|\vec{\nabla} s|^2} \tag{11}$$

Unfortunately, small intensity variations can result in infinite Demon forces, and thus we need to stabilize the former equation:

$$\vec{u} = \frac{(m - s)\vec{\nabla} s}{|\vec{\nabla} s|^2 + (m - s)^2} \tag{12}$$

To improve stability and convergence speed, Wang *et al*. [27] added an "active force". Diffusion was considered a bi-directional process, where Demon effectors also produce an internal gradient-based force of m, that diffuses s into m. Cachier *et al*. [28] suggested adding a normalization factor α to account for the adaptive force strength adjustment, yielding the following displacement field:

$$\vec{u} = (m - s) \times \left(\frac{\vec{\nabla} s}{|\vec{\nabla} s|^2 + \alpha^2 (s - m)^2} + \frac{\vec{\nabla} m}{|\vec{\nabla} m|^2 + \alpha^2 (s - m)^2} \right) \tag{13}$$

Wang *et al*. introduces a simple, iterative algorithm as follows:

1) Calculation of the disp. field using Eq. (13).
2) Regularization of the disp. field using Gaussian smoothing, to suppress noise and preserve geometric continuity.

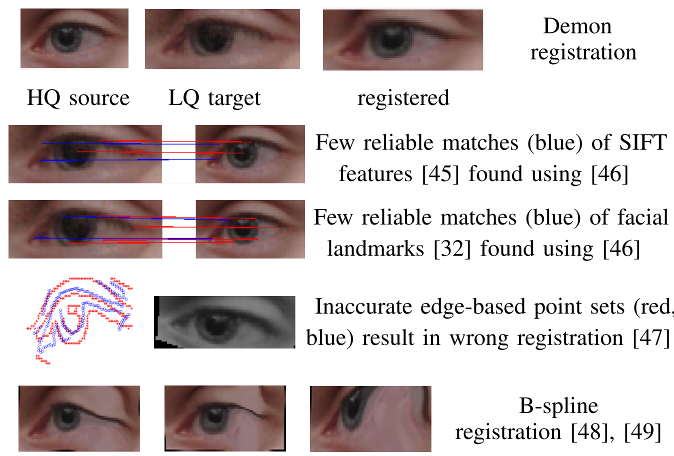

Fig. 28. Comparison of Demon registration to other methods. 1^{st} row: Demon registration finely adjusts to the desired structure, while improving patch quality. 2^{nd} and 3^{rd} rows: Compared to that, Point Matching techniques [46] have to rely on few or geometrically inconsistent features, and therefore find few reliable (blue) matches. 4^{th} row: Dense yet inaccurate point sets, based on edge maps, are not properly aligned, resulting in wrong registration [47]. 5^{th} row: Free-form deformation based on B-splines [48], [49] can also result in distortions.

3) Adding the regularized disp. field to the total disp. field.
4) Image deformation according to the total disp. field.

It was proved that the Demon algorithm can be seen as an approximation of a second order gradient descent of a SSD criterion, which can be used to compare different non-rigid registration methods [28]. But, as opposed to our work, it was not used as an affinity criterion or to evaluate the success of image deformation. Demon registration was so far usually used for medical image registration, e.g. [29].

We do not exclude other registration methods which may be used alternatively for the same concept. However, point set registration techniques are less suitable for our needs than intensity-based methods, as the first rely on geometric landmarks, which are less invariant under non-rigid structure variations [28] and complex, blind quality degradations. Fig. 28 compares registering facial features of different quality and similar non-rigid structures using Demon registration and other point set based or non-rigid registration techniques. The best matching example should generate a (non-rigidly) registered image which is both naturally-looking and of accurate expression. Fig. 29 shows the advantage of using the Demon measure, compared to other common feature-based matching criteria. We compare it to two selection criteria based on minimal error of projective transformation, based on SIFT features [45] and RANSAC [50] outliers removal, or on SURF features [51] and MSAC [52] (a variant of RANSAC).

Appendix B
Demon Registration Behavior Analysis Under Geometric Variations

We explore the behavior of the time-limited (200 iterations) Demon registration and measure between semantically-related non-rigid image structures under geometric variations. Specifically, we demonstrate the behavior for different scales (Fig. 30), translations (Fig. 31) and

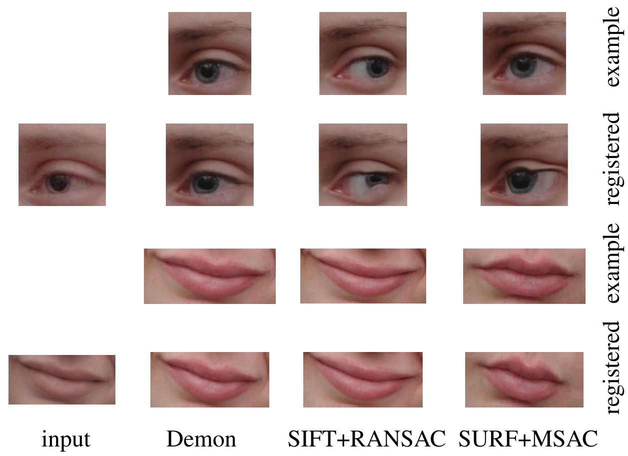

Fig. 29. Different matching criteria for HQ example selection: Demon measure; minimal error projective transformation using SIFT+RANSAC; and minimal error projective transformation using SURF+MSAC. Relying on features results in wrongly-selected examples; In many cases, no or too few feature points could be matched; or the transformation matrix was nearly singular. Thus the non-rigid (Demon) registrations are distorted or have undesired structures.

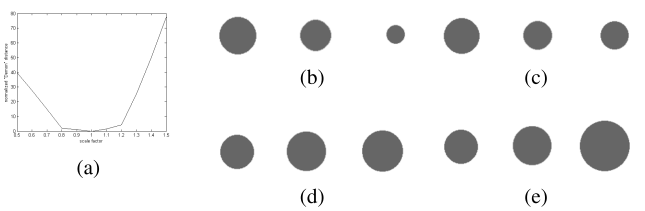

Fig. 30. Demon deformation for object scaling at different factors. For each scale factor (b)-(e): Left: source image. Middle: deformed image. Right: target image, that is, the source image scaled. (a) Demon distance vs. scale factor. (b) Scale factor=0.5. (c) Scale factor=0.8. (d) Scale factor=1.2. (e) Scale factor=1.5.

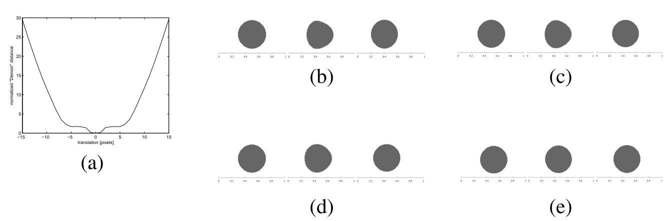

Fig. 31. Demon deformation for different object translations. For each translation (b)-(e): Left: source image. Middle: deformed image. Right: target image, that is, the source image translated. (a) Demon distance vs. translation. (b) Translation=−15 pixels. (c) Translation=−10 pixels. (d) Translation=−8 pixels. (e) Translation=−5 pixels.

rotations (Fig. 32). We also explore this behavior for a common non-rigid facial expression deformation: a change in eye gaze (Fig. 33). Note, that changing scale involves the dis/appearance of mass.

These experiments all illustrate the same behavior: for moderate variations the *time-limited* Demon deformation succeeds: the deformed image gets as close as possible to the target image (practically identical) (Figs. 30c, 30d, 31e, 32d, 32e, 33d, 33e); and the Demon measure *moderately* increases with variation. But there exists a breaking point where the *time-limited* deformation starts to fail: the deformed

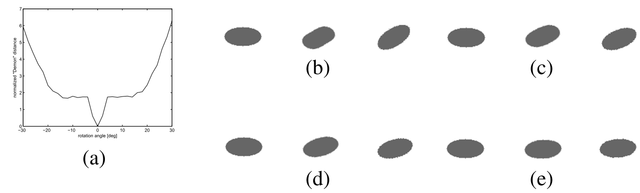

Fig. 32. Demon deformation for object rotation at different angles. For each rotation angle (b)-(e): Left: source image. Middle: deformed image. Right: target image, that is, the source image rotated. (a) Demon distance vs. rotation angle. (b) Rotation angle=30°. (c) Rotation angle=20°. (d) Rotation angle=14°. (e) Rotation angle=4°.

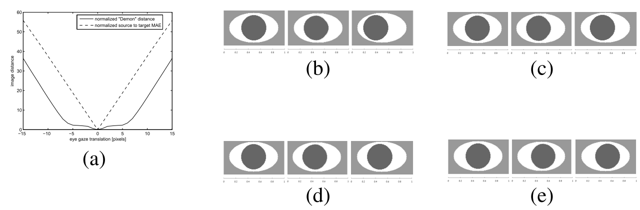

Fig. 33. Demon deformation for different eye gaze translations. For each translation (b)-(e): Left: source image, depicting a central gaze. Middle: deformed image. Right: target image, depicting a gaze change. (a) Demon distance and source to target MAE vs. eye gaze translation. (b) Eye gaze=−15 pixels. (c) Eye gaze=−7 pixels. (d) Eye gaze=−5 pixels. (e) Eye gaze=5 pixels.

image is too different from the target image, or distorted (Figs. 30b, 30e, 31b to 31d, 32b, 32c, 33b, 33c); and the measure starts to *drastically* increase. Compare this breaking-point behavior to the *linear* behavior of the MAE between *source* and target images, not reflecting the deformation's success / failure (Fig. 33a).

ACKNOWLEDGMENT

The authors would like to thank Yossi Bar Erez, Raz Nossek and Alona Baruhov for their assistance.

REFERENCES

[1] B. K. Horn and B. G. Schunck, "Determining optical flow," in *Proc. Tech. Symp. East Int. Soc. Opt. Photon.*, 1981, pp. 319–331.
[2] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 4, no. 3, pp. 367–383, Mar. 1992.
[3] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, 1992.
[4] J. Huang and D. Mumford, "Statistics of natural images and models," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1. Jun. 1999, p. 547.
[5] E. P. Simoncelli, "Bayesian denoising of visual images in the wavelet domain," in *Bayesian Inference Wavelet-Based Models*. New York, NY, USA: Springer, 1999, pp. 291–308.
[6] A. Buades, B. Coll, and J.-M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 490–530, 2005.
[7] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Color image denoising via sparse 3D collaborative filtering with grouping constraint in luminance-chrominance space," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, vol. 1. Oct. 2007, p. I–313.
[8] S. Roth and M. J. Black, "Fields of experts: A framework for learning image priors," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2. Oct. 2005, pp. 860–867.
[9] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[10] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 9, pp. 1167–1183, Sep. 2002.

[11] A. Levin and B. Nadler, "Natural image denoising: Optimality and inherent bounds," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 2833–2840.

[12] S. Baker and T. Kanade, "Hallucinating faces," in *Proc. 4th IEEE Int. Conf. Autom. Face Gesture Recognit.*, Dec. 2000, pp. 83–88.

[13] O. Bryt and M. Elad, "Compression of facial images using the k-SVD algorithm," *J. Vis. Commun. Image Represent.*, vol. 19, no. 4, pp. 270–282, 2008.

[14] J. Pan, Z. Hu, Z. Su, and M.-H. Yang, "Deblurring face images with exemplars," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 47–62.

[15] Y. Chen, W. Yu, and T. Pock, "On learning optimized reaction diffusion processes for effective image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 5261–5269.

[16] H. Talebi and P. Milanfar, "Asymptotic performance of global denoising," *SIAM J. Imag. Sci.*, vol. 9, no. 2, pp. 665–683, 2016.

[17] Y. Romano and M. Elad, "Boosting of image denoising algorithms," *SIAM J. Imag. Sci.*, vol. 8, no. 2, pp. 1187–1219, 2015.

[18] J.-P. Thirion, "Fast non-rigid matching of 3D medical images," Doctoral dissertation, INRIA - Institut National de Recherche en Informatique et en Automatique, France, 1995.

[19] D. Capel and A. Zisserman, "Super-resolution from multiple views using learnt image models," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2. Jun. 2001, p. II–627.

[20] K. Jia and S. Gong, "Generalized face super-resolution," *IEEE Trans. Image Process.*, vol. 17, no. 6, pp. 873–886, Jun. 2008.

[21] K.-C. Lee, J. Ho, M.-H. Yang, and D. Kriegman, "Video-based face recognition using probabilistic appearance manifolds," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1. Jun. 2003, p. I–313.

[22] J. Yu, B. Bhanu, Y. Xu, and A. K. Roy-Chowdhury, "Super-resolved facial texture under changing pose and illumination," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, vol. 3. Oct. 2007, p. III–553.

[23] Y. Shih, V. Kwatra, T. Chinen, H. Fang, and S. Ioffe, "Joint noise level estimation from personal photo collections," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 2896–2903.

[24] N. Joshi, W. Matusik, E. H. Adelson, and D. J. Kriegman, "Personal photo enhancement using example images," *ACM Trans. Graph.*, vol. 29, no. 2, p. 12, 2010.

[25] Y. R. Loke, P. Tan, and A. A. Kassim, "Face hallucination on personal photo albums," in *Proc. Comput. Vis.-ACCV Workshops*, 2013, pp. 284–295.

[26] J.-P. Thirion, "Image matching as a diffusion process: An analogy with Maxwell's demons," *Med. Image Anal.*, vol. 2, no. 3, pp. 243–260, 1998.

[27] H. Wang *et al.*, "Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy," *Phys. Med. Biol.*, vol. 50, no. 12, p. 2887, 2005.

[28] P. Cachier, X. Pennec, and N. Ayache, "Fast non rigid matching by gradient descent: Study and improvements of the 'demons' algorithm," Doctoral dissertation, INRIA - Institut National de Recherche en Informatique et en Automatique, France, 1999.

[29] D.-J. Kroon and C. H. Slump, "Mri modalitiy transformation in demon registration," in *Proc. IEEE Int. Symp. Biomed. Imag. Nano Macro (ISBI)*, Jul. 2009, pp. 963–966.

[30] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 2879–2886.

[31] A. Levin, D. Lischinski, and Y. Weiss, "A closed-form solution to natural image matting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 228–242, Feb. 2008.

[32] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, "Incremental face alignment in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1859–1866.

[33] Y. HaCohen, E. Shechtman, D. B. Goldman, and D. Lischinski, "Non-rigid dense correspondence with applications for image enhancement," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 70-1–70-10, 2011.

[34] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.

[35] P. J. Burt and E. H. Adelson, "A multiresolution spline with application to image mosaics," *ACM Trans. Graph.*, vol. 2, no. 4, pp. 217–236, 1983.

[36] H. Zhang, D. Wipf, and Y. Zhang, "Multi-image blind deblurring using a coupled adaptive sparse prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1051–1058.

[37] J. Pan, D. Sun, M.-H. Yang, and H. Pfister, "Blind image deblurring using dark channel prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1628–1636.

[38] Y. Niu, X. Wu, and G. Shi, "Image enhancement by entropy maximization and quantization resolution upconversion," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4815–4828, Oct. 2016. [Online]. Available: http://dx.doi.org/10.1109/TIP.2016.2598485

[39] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[40] K. Gu, G. T. Zhai, and M. Lin, "The analysis of image contrast: From quality assessment to automatic enhancement," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 284–297, Jan. 2015.

[41] K. Gu, G. Zhai, X. Yang, W. Zhang, and C. W. Chen, "Automatic contrast enhancement technology with saliency preservation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 9, pp. 1480–1494, Sep. 2014.

[42] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Using free energy principle for blind image quality assessment," *IEEE Trans. Multimedia*, vol. 17, no. 1, pp. 50–63, Jan. 2015.

[43] K. Gu, G. Zhai, W. Lin, X. Yang, and W. Zhang, "No-reference image sharpness assessment in autoregressive parameter space," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3218–3231, Oct. 2015.

[44] K. Gu, W. Lin, G. Zhai, X. Yang, W. Zhang, and C. W. Chen, "No-reference quality metric of contrast-distorted images based on information maximization," *IEEE Trans. Cybern.*, to be published. [Online]. Available: http://ieeexplore.ieee.org/abstract/document/7492198/

[45] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[46] J. Ma, J. Zhao, J. Tian, A. L. Yuille, and Z. Tu, "Robust point matching via vector field consensus," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1706–1721, Apr. 2014.

[47] A. Myronenko and X. Song, "Point set registration: Coherent point drift," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 12, pp. 2262–2275, Dec. 2010.

[48] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: Application to breast MR images," *IEEE Trans. Med. Imag.*, vol. 18, no. 8, pp. 712–721, Aug. 1999.

[49] D.-J. Kroon. *B-Spline Registration Implementation* [Online]. Available: http://www.mathworks.com/matlabcentral/fileexchange/20057-b-splinegrid– image-and-point-based-registration.

[50] M. A. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[51] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 404–417.

[52] P. H. S. Torr and A. Zisserman, "MLESAC: A new robust estimator with application to estimating image geometry," *Comput. Vis. Image Understand.*, vol. 78, no. 1, pp. 138–156, 2000.

Ester Hait received the B.Sc. (*Cum Laude*) and M.Sc. (*Cum Laude*) degrees in electrical engineering from the Technion–Israel Institute of Technology, Haifa, Israel, in 2014 and 2016, respectively. Her research interests include image processing and computer vision.

Guy Gilboa received the Ph.D. degree from the Electrical Engineering Department, Technion–Israel Institute of Technology, in 2004. He was a Post-Doctoral Fellow with UCLA and had various development and research roles with Microsoft and Philips Healthcare. Since 2013, he has been a Faculty Member with the Electrical Engineering Department, Technion–Israel Institute of Technology. He has authored some highly cited papers on topics, such as image sharpening and denoising, nonlocal operators theory, and texture analysis. He received several prizes, including the Eshkol Prize by the Ministry of Science, the Vatat Scholarship, and the Gutwirth Prize. He serves on the editorial boards of the journals IEEE SPL, JMIV, and JVCIR.