




Restoration and Enhancement of Historical Stereo Photos

Marco Fanfani ¹ , Carlo Colombo ¹  and Fabio Bellavia ^{2,*} 

¹ Dept. of Information Engineering, Università degli Studi di Firenze, Italy; {marco.fanfani,carlo.colombo}@unifi.it

² Dept. of Math and Computer Science, Università degli Studi di Palermo, Italy; fabio.bellavia@unipa.it

* Correspondence: fabio.bellavia@unipa.it; Tel.: +39-0912-389-1124

† This paper is an extended version of our paper published in the 25th International Conference on Pattern Recognition Workshop (ICPRW 2020) on Fine Art Pattern Extraction and Recognition (FAPER 2020).

Abstract: Restoration of digital visual media acquired from repositories of historical photographic and cinematographic material is of key importance for the preservation, study and transmission of the legacy of past cultures to the coming generations. In this paper, a fully automatic approach to the digital restoration of historical stereo photographs is proposed, referred to as Stacked Median Restoration plus (SMR+). The approach exploits the content redundancy in stereo pairs for detecting and fixing scratches, dust, dirt spots and many other defects in the original images, as well as improving contrast and illumination. This is done by estimating the optical flow between the images, and using it to register one view onto the other both geometrically and photometrically. Restoration is then accomplished in three steps: (1) image fusion according to the stacked median operator, (2) low-resolution detail enhancement by guided supersampling, and (3) iterative visual consistency checking and refinement. Each step implements an original algorithm specifically designed for this work. The restored image is fully consistent with the original content, thus improving over the methods based on image hallucination. Comparative results on three different datasets of historical stereograms show the effectiveness of the proposed approach, and its superiority over single-image denoising and super-resolution methods. Results also show that the performance of the state-of-the-art single-image deep restoration network Bringing Old Photo Back to Life (BOPBtL) can be strongly improved when the input image is pre-processed by SMR+.

Keywords: Image denoising; Image restoration; Image enhancement; Stereo matching; Optical flow; Gradient filtering; Stacked median; Guided supersampling; Historical photos

Citation: Fanfani, M.; Colombo, C.; Bellavia, F. Restoration and Enhancement of Historical Stereo Photos. *J. Imaging* **2021**, *1*, 0. <https://doi.org/>

Received:

Accepted:

Published:

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Submitted to *J. Imaging* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Photographic material of the XIX and XX centuries is an invaluable source of information for historians of art, architecture and sociology, as it allows them to track the changes occurred over the decades to a community and its living environment. Unfortunately, due to the effect of time and bad preservation conditions, most of the survived photographic heritage is partially damaged, and needs restoration, both at the physical (cardboard support, glass negatives, films, etc.) and digital (the image content acquired through scanners) levels. Dirt, scratches, discoloration and other signs of aging strongly reduce the visual quality of photos [1]. A similar situation also holds for the cinematographic material [2].

Digital restoration of both still images and videos has attracted considerable interest from the research community in the early 2000s. This has led to the development of several tools that improve the visual quality. Some approaches rely on the instantiation of noise models, which can either be fixed a-priori or derived from the input images [3–5]. Other approaches detect damaged areas of the image and correct them according to inpainting techniques [6]. Self-correlation inside the image, or across different frames in videos, is often exploited in this context, under the assumption that zero-mean additive noise cancels out as the available number of image data samples increases [7–9]. A

38 similar idea is exploited by super-resolution techniques, that enhance image quality by
39 pixel interpolation [10,11]. In recent years, the algorithmic methods above have been
40 sided by methods based on deep learning, that can infer the image formation model or
41 the scene content [12] from a training set in order to inject this information into the final
42 output—a process called image hallucination [13–15]. Although the final image may
43 often alter the original image data content, and hence cannot be fully trusted (e.g., in the
44 medical diagnosis domain), the hallucination methods can give visually pleasing results
45 (see Fig. 1).

46 Stereoscopy has accompanied photography since its very birth in the nineteenth
47 century, with ups and downs in popularity through time. Notwithstanding the lesser
48 spread of stereo photography with respect to standard (monocular) photography, many
49 digital archives with thousands of stereo images exist, some of which are freely available
50 on the web. Stereo photos have a richer content than standard ones, as they present two
51 different views of the same scene, thus explicitly introducing content redundancy and
52 implicitly embedding information about scene depth. This characteristic can be exploited
53 also in digital noise removal, enhancement and restoration, since a damaged area in one
54 image can be reconstructed from the other image, provided that the correspondences
55 between the two images are known. At a first glance, the above mentioned approach
56 looks similar to that of video restoration from multiple video frames, in which the scene is
57 acquired in subsequent time instants from slightly changed viewpoints. However, stereo
58 images have their own peculiarities, and actually introduce in the restoration process
59 more complications than video frames, which in movies typically exhibit an almost static
60 and undeformed background, differently from stereo pairs. As a matter of fact, although
61 several advances have been recently made in stereo matching and dense optical flow
62 estimation [17], the problem is hard and far to be fully solved, especially in the case of
63 very noisy and altered images such as those generated by early photographic stereo
64 material. To the best of the authors' knowledge, stereo photo characteristics have been
65 employed only for the super-resolution enhancement or deblurring of modern, clean
66 photos [18–20]. On the other hand, the image analysis and computer vision approaches
67 developed so far for historical stereo photos mainly aimed at achieving (usually in a
68 manual way) better visualizations or 3D scene reconstructions [21–23], with no attempt
69 at restoring the quality of the raw stereo pairs.

70 This paper proposes a new approach to clean-up and restore the true scene content
71 in degraded historical stereo photographs, named Stacked Median Restoration Plus
72 (SMR+) extending our previous work [24], and working in a fully automatic way. With
73 respect to existent single image methods, damaged image areas with scratches or dust
74 can be better detected and fixed, thanks to the availability of more sampled data points
75 for denoising. In addition, the correct illumination can be restored or enhanced in a way
76 akin to that of High Dynamic Range Imaging, where the images of the same scene taken
77 at different exposure levels are used in order to enhance details and colors [25]. For
78 this scope the optical flow, estimated with a recent state-of-the-art deep network [17], is
79 used to synthesize corresponding scene viewpoints in the stereo pair, while denoising
80 and restoration are carried out using novel yet non-deep image processing approaches.
81 The entire process is superseded by scene content consistency validation, used to check
82 critical stereo matching mispredictions that were left unresolved by the network. Our
83 approach aims to obtain an output which is fully consistent with the original scenario
84 captured by the stereo pair, in contrast with the recent super-resolution and denoising
85 approaches based on image hallucination.

86 This paper extends our previous work [24], hereafter reported as Stacked Median
87 Restoration (SMR) under several aspects:

- 88 • With respect to SMR, the novel SMR+ is redesigned so as to better preserve finer
89 details while at the same time improving further the restoration quality. This
90 is accomplished by employing supersampling [26] at the image fusion step in
91 conjunction with a weighting scheme guided by the original restoration approach.

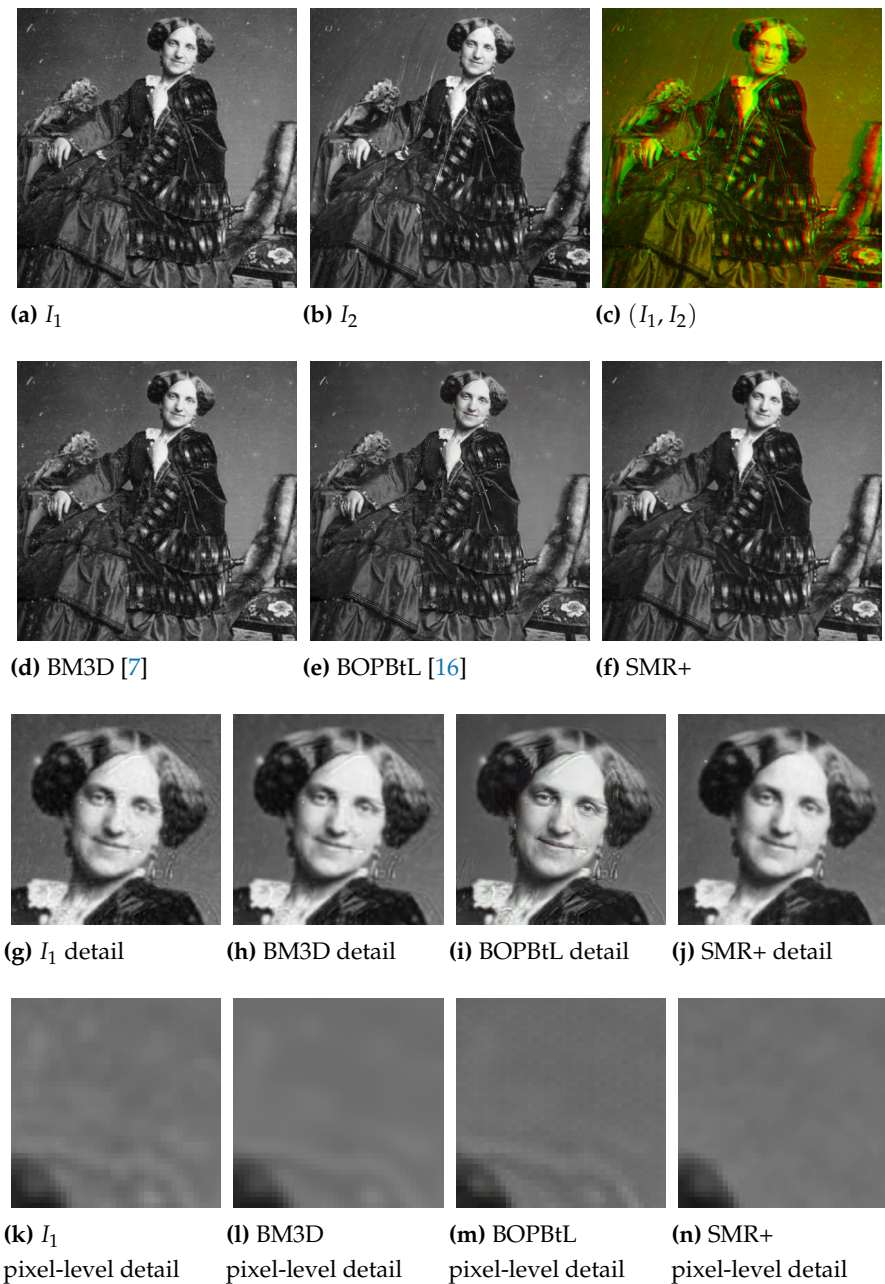


Figure 1. First row: An example of historical stereo pair of images, I_1 and I_2 , also superimposed as anaglyph. Second row: Enhancement of I_1 according to different methods, including the proposed SMR+ method. Although visually impressive, the deep super-resolution result of BOPBtL does not preserve the true input image. Third row: A detail of I_1 and the restored images according to the different methods. A closer look at BOPBtL reveals alterations with respect to the original face expression, accentuating the smile and introducing bush-like textures on the hair. Forth row: pixel level detail of I_1 and of the restored images according to the different methods. The specific image region considered is the background around the right shoulder. Notice the chessboard-like texture pattern typical of the deep network approaches, not visible at coarser scales. Best viewed in color. The reader is invited to zoom in the electronic version of the manuscript in order to better appreciate the visual differences.

- 92 • A recent state-of-the-art deep network specifically designed for old photo restoration
 93 named Bringing Old Photo Back to Life (BOPBtL) [16] is now included in the
 94 comparison, both as standalone and to serve as post-processing of SMR+.

- 95 • The collection of historical stereo photos employed as dataset is roughly doubled to
- 96 provide a more comprehensive evaluation.
- 97 • The use of renowned image quality assessment metrics is investigated and dis-
- 98 cussed for this kind of applications.

99 The rest of the paper is organized as follows: Sec. 2 introduces the proposed
100 approach. An experimental evaluation and comparison with similar approaches is
101 reported in Sec. 3. Finally, conclusions and future work are discussed in Sec. 4.

102 *Note: To ease the inspection and the comparison of the different images presented, an interactive PDF*
103 *document is provided in the [additional material](#) to allow readers to view each image at its full dimensions*
104 *and quickly switch to the other images to be compared.*

105 2. Proposed method

106 Given a pair of stereo images I_1 and I_2 , the aim of the process is to output a defect-free version
107 of one image of the pair (referred to as the reference) by exploiting the additional information
108 coming from the other image (denoted as auxiliary). For convenience, the reference is denoted
109 as I_1 (see Fig. 2a) and the auxiliary image as I_2 (see Fig. 2b), but their roles can be interchanged.
110 Images are assumed to be single channel graylevel, i.e., $I_1, I_2 : \mathbb{R}^2 \rightarrow [0, 255]$.

111 2.1. Auxiliary Image Pointwise Transfer

As first step, the recent state-of-the-art Recurrent All-Pairs Field Transforms (RAFT) deep
network [17] is used to compute the optical flow map pair $f_{\text{RAFT}}(I_1, I_2) = (m_x, m_y)$, where
 $m_x, m_y : \mathbb{R}^2 \rightarrow \mathbb{R}$ (see Fig. 2d), so that a synthesized image based on the content of I_2 and registered
onto I_1 can be obtained as

$$\tilde{I}_{2 \rightarrow 1}(x, y) = I_2(x + m_x(x, y), y + m_y(x, y)) \quad (1)$$

112 by transferring pixel intensity values from I_2 into the view given by I_1 (see Fig. 2e). Note that
113 spots of missing data can be present on $\tilde{I}_{2 \rightarrow 1}$ when no pixel in I_2 maps onto the specific image
114 area, due for instance to image occlusions. In the error free ideal case it must hold that $I_1 = \tilde{I}_{2 \rightarrow 1}$
115 for every correspondence between I_1 and I_2 . However, in real situations this may not happen, as
116 shown in Fig. 2f reporting the average absolute error between I_1 and $\tilde{I}_{2 \rightarrow 1}$ on 5×5 local window
117 patches.

118 Notice also that in the case of perfectly rectified stereo images it holds everywhere that
119 $m_y(x, y) = 0$. Under this particular setup, in which m_x is denoted as *disparity map* and is the
120 only map that needs to be estimated, several classical methods have proven to provide good
121 results while being computationally efficient [27]. However, according to our experience [22] these
122 methods are not feasible in the case of degraded historical stereo photos. First, image degradation
123 due to aging and the intrinsic image noise due to the technological limitations of the period
124 decrease the ability of these methods to find the right correspondences. Second, the output of
125 these methods is quite sensitive to the initial configuration of the parameters and, by considering
126 the variability of the historical acquisition setups, each stereo pair would require the human
127 supervision to get even a sub-optimal result. Third, the stereo alignment for the photos under
128 consideration is far from perfect due to the technological limitations of the period, hence both the
129 maps m_x and m_y are to be considered. Hence, our choice fell under the state-of-the-art RAFT, that
130 provides a sufficiently good initial estimation of the optical flow maps in most cases.

A further flow mapping pair $f_{\text{RAFT}}(I_2, I_1) = (m'_x, m'_y)$ (see Fig. 2g) can be obtained by
switching the two input images, which can be employed to synthesize a second image according to

$$\tilde{I}'_{2 \rightarrow 1}(x, y) = I_2(x - m'_x(x, y), y - m'_y(x, y)) \quad (2)$$

131 (see Fig. 2h) so that in the error free ideal case for every correspondence between I_1 and I_2 it holds
132 that $(m_x, m_y) = -(m'_x, m'_y)$ which implies that $I_1 = \tilde{I}_{2 \rightarrow 1} = \tilde{I}'_{2 \rightarrow 1}$. This usually does not happen,
133 as shown by the relative error image of Fig. 2i. Indeed, comparing the first and second rows
134 of Fig. 2, RAFT optical flow estimation is not completely accurate and does not preserve map
135 inversion when exchanging the input image order. The final synthesized image $\tilde{I}_{2 \rightarrow 1}$ (see Fig.
136 2j) is then obtained by choosing the intensity value at each pixel (x, y) as the one from $\tilde{I}_{2 \rightarrow 1}(x, y)$
137 and $\tilde{I}'_{2 \rightarrow 1}(x, y)$ that minimizes the sum of absolute errors with respect to I_1 on a small 5×5
138 local window centered on the pixel (see Fig. 3). A smaller error between the final resynthesized

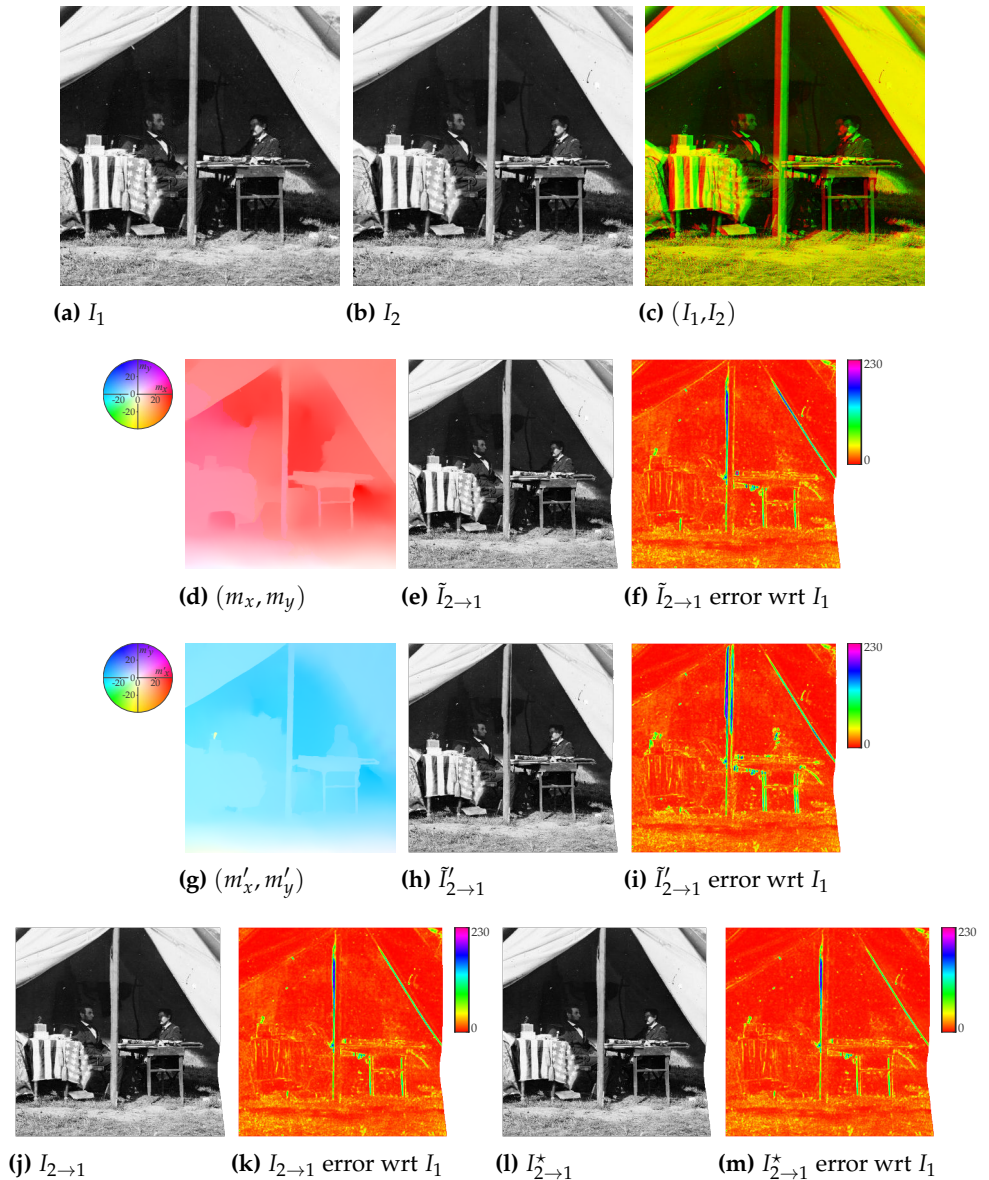


Figure 2. Auxiliary image pointwise transfer and color correction steps (see Secs. 2.1-2.2): (a) Reference image I_1 , (b) auxiliary image I_2 , (c) superimposition of I_1 and I_2 as anaglyph, (d) visual representation of the optical flow map (m_x, m_y) extracted by RAFT, (e) image $\tilde{I}_{2 \rightarrow 1}$ as resynthesis of I_2 through (m_x, m_y) and (f) its error with respect to I_1 , (g) visual representation of the optical flow map (m'_x, m'_y) extracted by RAFT after switching the input images, (h) image $\tilde{I}'_{2 \rightarrow 1}$ as resynthesis from I_2 through $-(m'_x, m'_y)$ and (i) its error with respect to I_1 , (j) final resynthesized image $I_{2 \rightarrow 1}$ considering the locally best optical flow estimation between $\tilde{I}_{2 \rightarrow 1}(x, y)$ and $\tilde{I}'_{2 \rightarrow 1}(x, y)$ and (k) its error with respect to I_1 , (l) image $I_{2 \rightarrow 1}^*$ obtained after applying GPS/LCP color correction to $I_{2 \rightarrow 1}$ using I_1 as reference and (m) the corresponding error map with respect to I_1 . Best viewed in color. The reader is invited to zoom in the electronic version of the manuscript in order to better appreciate the visual differences.

139 image $I_{2 \rightarrow 1}$ and the reference image I_1 is obtained (see Fig. 2k) with respect to the errors given by
 140 $\tilde{I}_{2 \rightarrow 1}(x, y)$ and $\tilde{I}'_{2 \rightarrow 1}(x, y)$.

141 2.2. Color Correction

Due to the technical limitations of the old photographic instrumentation, illumination conditions between the two stereo images can differ noticeably. For instance, flash lamp and, even more,

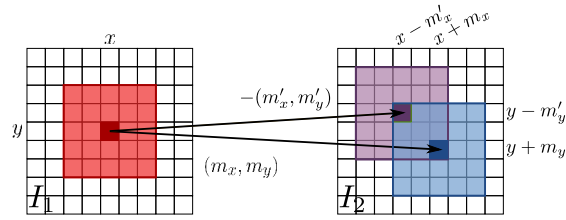


Figure 3. Illustration of the $I_{2 \rightarrow 1}$ image formation process from the two resynthesized images $\tilde{I}_{2 \rightarrow 1}(x, y)$ and $\tilde{I}'_{2 \rightarrow 1}(x, y)$, respectively driven by the optical flow estimation maps (m_x, m_y) and $-(m'_x, m'_y)$. A point (x, y) in I_1 can be mapped back to I_2 according to either Eq. 1 or Eq. 2. The best back-mapping minimizing locally the error among the two possible optical flow estimates is then chosen to form $I_{2 \rightarrow 1}$. Best viewed in color.

flash powder did not provide each time uniform and identical illumination conditions, and it was not infrequent that a single camera was moved in two different positions in order to simulate a stereo setup instead of having two synchronized cameras [22]. Moreover, discoloration of the support due to aging can be present. In order to improve the final result, the state-of-the-art color correction method named Gradient Preserving Spline with Linear Color Propagation (GPS/LCP) presented in [28] is employed to correct the illumination of $I_{2 \rightarrow 1}$ according to I_1 . Specifically, the color map $g_{\text{GPS/LCP}}(I_1, I_{2 \rightarrow 1}) = C$, with $C : \mathbb{R} \rightarrow \mathbb{R}$ is used to obtain the color corrected image $I_{2 \rightarrow 1}^*$ according to

$$I_{2 \rightarrow 1}^*(x, y) = C(I_{2 \rightarrow 1}(x, y)) \quad (3)$$

142 where in the error free ideal case it must hold that $I_1 = I_{2 \rightarrow 1}^*$ (see Fig. 2l). The GPS/LCP color
 143 correction method is able to preserve the image content and works also in the case of not perfectly
 144 aligned images. Color correction decreases the resynthesis error. This can be noted by comparing
 145 the error map of $I_{2 \rightarrow 1}^*$ (Fig. 2m) with the error map of $I_{2 \rightarrow 1}$ (Fig. 2k) — see in particular the
 146 error corresponding to the dark background above the left table. Clearly, if $I_{2 \rightarrow 1}$ presents better
 147 illumination conditions than I_1 , it is also possible to correct I_1 according to $I_{2 \rightarrow 1}$.

148 2.3. Data Fusion

Given the reference image I_1 and the synthesized one obtained from the auxiliary view $I_{2 \rightarrow 1}^*$ after the illumination post-processing, the two images are blended into a new image I_{12} according to the *stacked median* operator (see Fig. 4a)

$$I_{12} = \boxplus(I_1 \cup I_{2 \rightarrow 1}^*) \quad (4)$$

The stacked median $\boxplus(\{I\})$ for a set of images $\{I\}$ outputs a new image defined so that image intensity at pixel (x, y) is the median intensity value computed on the union of the pixels in the 3×3 local neighbourhood windows centered at (x, y) on each image of the set (see Fig. 5). Notice that the *median stacking* operation typically found as blending tool in image manipulation software corresponds to the proposed stacked median operator with degenerate 1×1 local windows. Unlike median stacking, the proposed definition does not require more than two input images and considers pixel neighborhoods, i.e. it works locally and not pointwise. Additionally, in case of missing data in $I_{2 \rightarrow 1}^*$, the stacked median acts as a standard 3×3 median filter. With this operator, dirt, scratches and other signs of photographic age or damages are effectively removed from I_{12} , but high frequency details can be lost in the process, due to the 3×3 filtering (see Fig. 4b). These are partially re-introduced by considering a blended version of the gradient magnitude

$$d_{m_{12}} = \boxplus(M(I_1) \cup M(I_{2 \rightarrow 1}^*)) \quad (5)$$

(see Fig. 4c) obtained as the stacked median of eight possible gradient magnitudes, four for each of the I_1 and $I_{2 \rightarrow 1}^*$ images, to further enhance finer details. Each gradient magnitude image in the set $M(I)$ for a generic image I is computed as

$$d_m = \left(d_x^2 + d_y^2 \right)^{\frac{1}{2}} \quad (6)$$

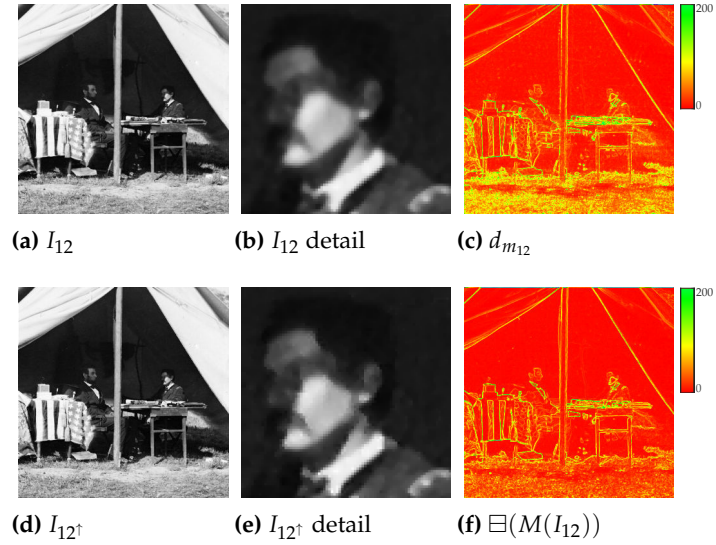


Figure 4. Data fusion step (see Sec. 2.3): (a) Stacked median I_{12} obtained from $I_1 \cup I_{2 \rightarrow 1}^*$, (b) a detail of I_{12} , (c) the stacked median $d_{m_{12}}$ of the gradient magnitudes of I_1 and $I_{2 \rightarrow 1}^*$, (d) the gradient-enhanced image I_{12}^\dagger , (e) a detail of I_{12}^\dagger , (f) the gradient magnitude $\Xi(M(I_{12}))$ of the stacked median image I_{12} . Best viewed in color. The reader is invited to zoom in the electronic version of the manuscript in order to better appreciate the visual differences.

pixelwise, where the image gradient direction pairs (d_x, d_y) are computed by the convolution of I with the following four pairs of kernel filters

$$\left\{ \left(\begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right), \left(\begin{bmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right), \left(\begin{bmatrix} 0 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 1 & 0 \end{bmatrix} \right), \left(\begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) \right\} \quad (7)$$

Notice that $d_{m_{12}} \neq \Xi(M(I_{12}))$ in the general case (compare Fig. 4c with Fig. 4f). Consider for now only a single derivative pair (d_x, d_y) of I_{12} : Each pixel intensity $I_{12}(x, y)$ is incremented by a value $v(x, y)$ satisfying

$$\left(d_x + \frac{v}{2} \right)^2 + \left(d_y + \frac{v}{2} \right)^2 = \frac{d_m^2 - d_{m_{12}}^2}{2} \quad (8)$$

This equation has a twofold solution

$$v^* = \pm (2d_x d_y - d_{m_{12}}^2)^{\frac{1}{2}} - (d_x + d_y) \quad (9)$$

In the case of two real v^* solutions, v is chosen as $v(x, y) = \arg \min_{\bar{v} \in v^*} |\bar{v}|$ in order to alter I_{12} as little as possible. In the case of complex solutions, $v(x, y)$ is set to 0. The final gradient-enhanced image is then obtained as

$$I_{12}^\dagger = I_{12} + v \quad (10)$$

(see Fig. 4d, and Fig. 4e for a detail). Since four different v values are obtained for each of the four derivative pairs of Eq. 7, their average value is actually employed.

2.4. Refinement

As already noted, the optical flow may be not perfect, causing the presence of wrong data in the image synthesis and hence in the data fusion process described in the previous step. To alleviate this issue, an iterative error-driven image correction step is introduced, where each iteration can be split into two sub-steps:

1. **Detection.** A binary correction mask is computed by considering on the error image $E = (I_1 - I_{12}^\dagger)^2$ the 11×11 local window $L(x, y)$ centered at each (x, y) . Given $L'(x, y) \subseteq L(x, y)$ as the subset of pixels with intensity values lower than the 66% percentile on $L(x, y)$, the pixel (x, y) is marked as requiring adjustment if the square root of the average intensity value on $L'(x, y)$ is higher than $t = 16$ (chosen experimentally). This results in a binary correction mask B , that is smoothed with a Gaussian kernel and then binarized again by a

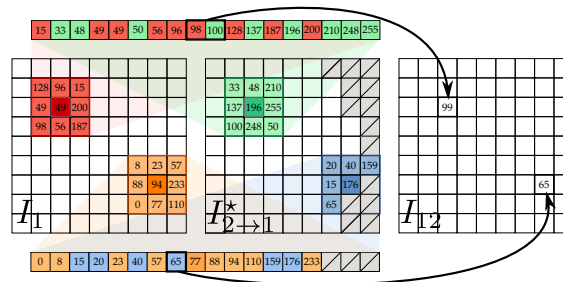


Figure 5. Application of the stacked median operator \boxplus for computing I_{12} from $I_1 \cup I_{2 \rightarrow 1}^*$. At pixel (x, y) the stacked median operator takes the union of the corresponding 3×3 local neighbourhoods for each image of the input set (in the example the union of the red and green neighbourhoods, and the union of the orange and blue ones, missing data are represented in the figure as gray ticked boxes) and assigns its median intensity value to the point (x, y) in the new image. Best viewed in color.

- 162 threshold value of 0.5. As clear from Fig. 6a, using the percentile-based subset $L'(x, y)$ is
 163 more robust than working with the whole window $L(x, y)$.
- 164 2. Adjustment. Data fusion is repeated again after updating pixels on $I_{2 \rightarrow 1}^*$ that need to be
 165 adjusted with the corresponding ones of $I_{12 \uparrow}$. Since $I_{12 \uparrow}$ is a sort of average between I_1
 166 and $I_{2 \rightarrow 1}^*$, the operation just described pushes marked pixels towards I_1 . At the end of this
 167 step, the gradient enhanced image $I_{12 \uparrow}$ is also updated accordingly and in case of no further
 168 iterations it constitutes the final output.
- 169 Iterations stop when no more pixels to be adjusted are detected in the updated $I_{12 \uparrow}$ or when the
 170 maximum number of iterations is reached (see Fig. 6). A maximum of four iterations is carried out,
 171 since it was verified experimentally that data fusion typically converges to I_1 within this number
 of steps.

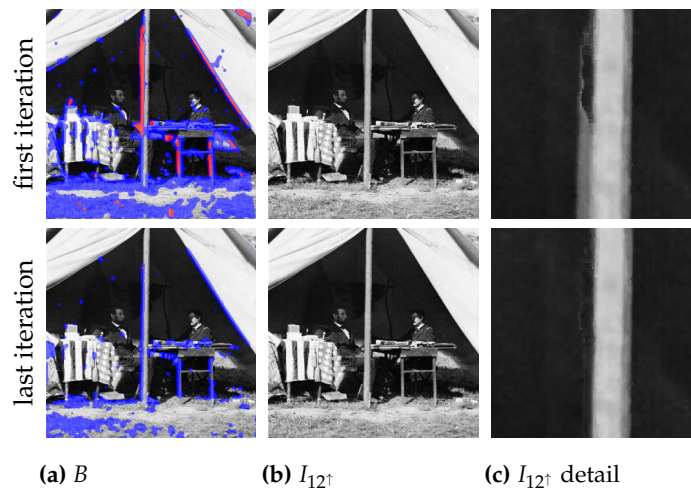


Figure 6. Refinement step (see Sec. 2.4). First (top row) and last (bottom row) iterations of the detection and adjustment sub-steps. (a) Detection mask B at the beginning of the iteration, (b) updated $I_{12 \uparrow}$ at the end of the iteration and (c) a detail of (b). Pixels to be adjusted using L' (L) are underlined in the images by saturating the red (blue) channel. By inspecting the details it can be seen that the ghosting effect is removed. Best viewed in color. The reader is invited to zoom in the electronic version of the manuscript in order to better appreciate the visual differences.

172

173 2.5. Guided Supersampling

174 Previous steps describes the original SMR implementation [24]. In order to preserve more
 175 fine details of the input images, a better image fusion is proposed hereafter, where the original

176 coarse blended image I_{12} (Eq. 4) is employed to guide a refinement on the basis of supersampling
 177 (see Fig. 7).

Let W_1 denote the image obtained by averaging $|I_1 - I_{12}|$ on a 3×3 window, and similarly W_2 the one obtained with $|I_2 - I_{12}|$. The weight mask W is computed as $W_1 / (W_1 + W_2)$ pixelwise, followed by the convolution with a Gaussian with a standard deviation of 4 pixels (see Fig. 7d). A value of W close to 0 (1) for a given pixel implies that the local neighborhood of that pixel in I_1 (I_2) is very likely less noisy and more artefact-free than I_2 (I_1). The mask W is used to define a *weighted stacked median*

$$H_{12} = \Xi_W(I_1^{\times 2}, I_2^{\star \times 2}) \quad (11)$$

178 where the superscript $\times 2$ indicates the bicubic rescaling by a factor two for supersampling (see Fig.
 179 7e). Explicitly, the weighted stacked median at (x, y) is obtained as the median of the intensities of
 180 $V_1(x, y) \cup V_2(x, y)$, where $V_1(x, y) \subseteq I_1$ is the subset of the pixels in the 3×3 local neighbourhood
 181 of (x, y) containing the $\lfloor w \times \min(1 - W(x, y), w') \rfloor$ intensity values closest to $I_{12}^{\times 2}(x, y)$, with
 182 $w = 3^2 \times 2$ and $w' = (3^2 + 1) / (2 \times 3^2 + 1)$, and likewise for $V_2(x, y) \subseteq I_2$ containing the pixels
 183 with the $\lfloor w \times \min(W(x, y), w') \rfloor$ closest values. In other words, the number of considered samples
 184 for the median taken from each image is proportional to the weight $W(x, y)$. The cardinalities of
 185 the subsets V_1 and V_2 for the different weight ranges are explicitly shown in Table 1.

186 The high resolution blended image H_{12} replaces I_{12} in the next steps of the method (see
 187 Secs. 2.3-2.4), being also I_1 and I_2 replaced accordingly by $I_1^{\times 2}$ and $I_2^{\star \times 2}$. The final output is scaled
 188 down to the original input size. With respect to the original SMR implementation, the use of
 189 guided supersampling in SMR+ preserves better fine details, also improving further the restoration
 190 process (compare Fig. 7c and Fig. 7g). Notice that after each refinement sub-step (see Sec. 2.4), the
 191 coarse I_{12} image needed to guide the process is generated by the stacked median between I_1 and
 $I_{12}^{\star \times 2}$ scaled down to the original size.

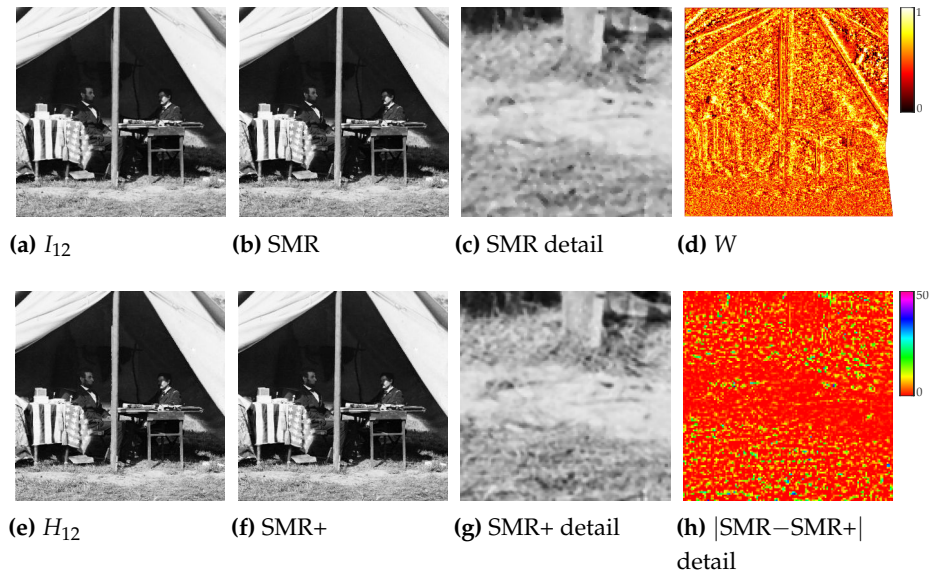


Figure 7. Guided supersampling step (see Sec. 2.5): (a) SMR stacked median I_{12} , (b) final restored image and (c) a detail of it, (d) weight mask W for the guided supersampling, (e) SMR+ weighted stacked median H_{12} , (f) final restored image, (g) a detail of it, and (h) its differences with respect to the SMR output. Best viewed in color. The reader is invited to zoom in the electronic version of the manuscript in order to better appreciate the visual differences.

192

193 3. Evaluation

194 3.1. Dataset

195 In order to evaluate the proposed approach we built a new dataset including historical stereo
 196 pairs from different sources. The left frames of the selected stereo pairs are shown as reference in
 197 Fig. 8.

198 A first set of seven stereo pairs belongs to the collection of stereograms by Anton Hautmann,
 199 one of the most active photographers in Florence between 1858 and 1862. Part of Hautmann's

Table 1: The cardinality of the sets $V_1(x, y)$ and $V_2(x, y)$ according to the weight $W(x, y)$ range (see Sec. 2.5).

$\inf W(x, y)$	0.00	0.05	0.11	0.16	0.21	0.26	0.32	0.37	0.42	0.47	0.53	0.58	0.63	0.68	0.74	0.79	0.84	0.89	0.95
$\sup W(x, y)$	0.05	0.11	0.16	0.21	0.26	0.32	0.37	0.42	0.47	0.53	0.58	0.63	0.68	0.74	0.79	0.84	0.89	0.95	1.00
$ V_1(x, y) $	9	9	9	9	9	9	9	9	9	9	8	7	6	5	4	3	2	1	0
$ V_2(x, y) $	0	1	2	3	4	5	6	7	8	9	9	9	9	9	9	9	9	9	9

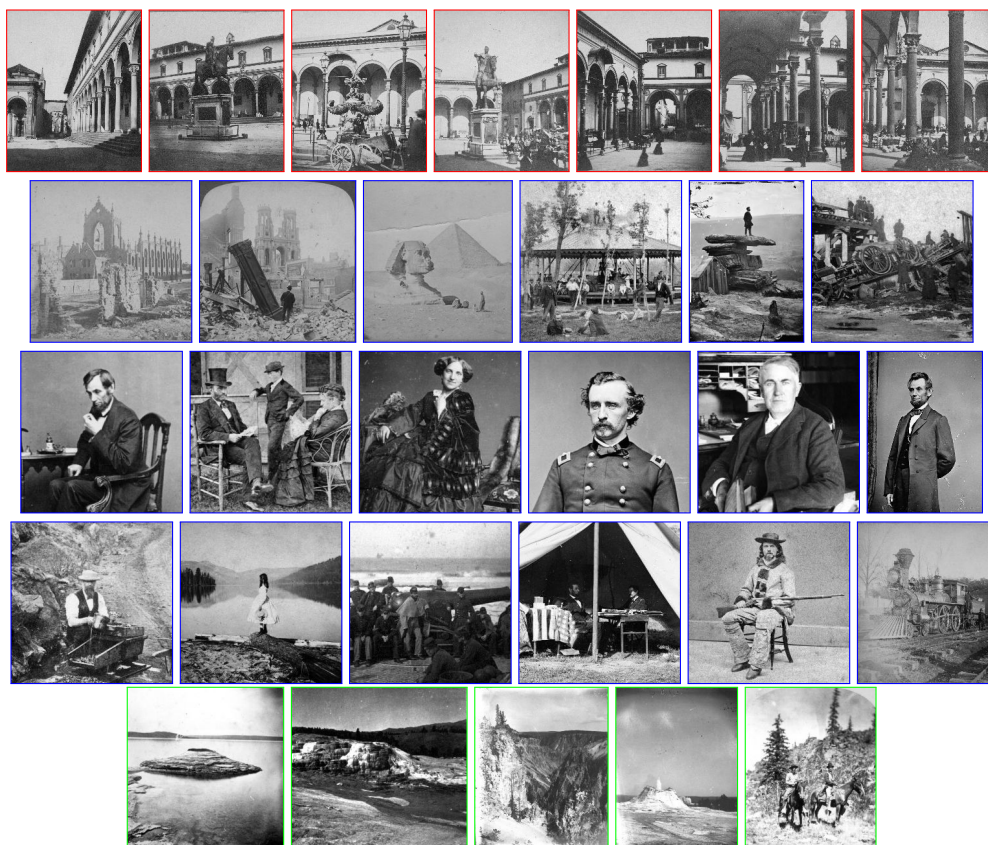


Figure 8. Left frames for some historical stereo pairs. Image frames for Hautmann’s, Stereoscopic Photos and USGS datasets are framed respectively in red, blue and green. Best viewed in color and zoomed in with the electronic version of the manuscript.

200 collection is described in [22]. The seven stereo pairs used in this work depict different viewpoints
 201 of Piazza Santissima Annunziata in Florence as it was in the middle of the XIX century. Inspecting
 202 these photos (see Fig. 8, red frames), it can be noticed that the image quality is very poor. In
 203 particular, the pairs are quite noisy, with low definition and contrast, include saturated or blurred
 204 areas and also show scratches and stains.

205 A second set includes 35 stereo pairs and increases the original set of ten images employed
 206 in [24]. These stereo pairs have been gathered from the Stereoscopic History Instagram account¹
 207 (see Fig. 8, blue frames) and contain landscape pictures of urban and natural
 208 scenes as well as individual or group portraits. This set is the most challenging one, since its
 209 images are heavily corrupted by noise and other artefacts.

210 A third set of five images was collected from the U.S. Geological Survey (USGS) Historical
 211 Stereoscopic Photos account on Flickr², and represents natural landscapes (see Fig. 8, green frames),
 212 except for the last one which includes also two horsemen with their mounts. The quality of these
 213 images is similar to that of the first set, but strong vignetting effects are also present.

¹ <https://www.instagram.com/stereoscopichistory/>

² <https://www.flickr.com/photos/usgeologicalsurvey/>

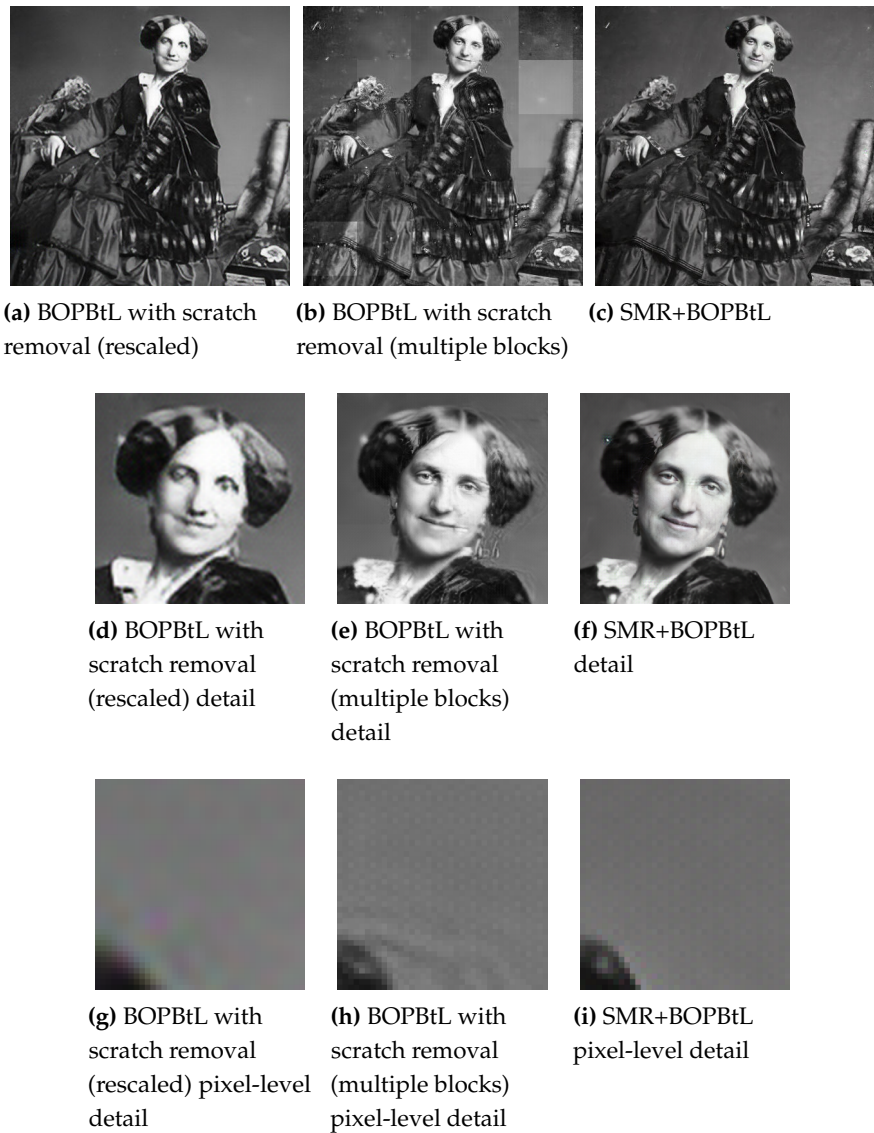


Figure 9. Results of BOPBtL with scratch removal or in combination with SMR+ on the same stereo pair of Fig. 1. Notice that the visual pleasant results of (a) are due to the frequency cutoff caused by rescaling and disappear at a larger viewing scale such in (d). Best viewed in color. The reader is invited to zoom in the electronic version of the manuscript in order to better appreciate the visual differences.

214 3.2. Compared Methods

215 The proposed SMR and SMR+ are compared against Block Matching 3D (BM3D) [7], Deep
 216 Image Prior (DIP) [13] and the recent BOPBtL [16]. BM3D and DIP are respectively a handcrafted
 217 and deep generic denoising methods, while BOPBtL is a deep network specifically designed for
 218 old photo restoration. These methods currently represent the state-of-the-art in this research area.

219 For BM3D, the legacy version was employed, since according to our preliminary experiments
 220 the new version including correlated noise suppression did not work well for our kind of images.
 221 The BM3D σ parameter, the only one present, was set to 7 and 14, values that according to our
 222 experiments gave the best visual results. In particular, $\sigma = 14$ seems to work better than $\sigma = 7$ in
 223 the case of higher resolution images. Besides applying the standard BM3D on the reference image,
 224 a modified version of this method was implemented in order for BM3D to benefit from the stereo
 225 auxiliary data. Since BM3D exploits image self-correlation to suppress noise, the modified BM3D
 226 generates auxiliary sub-images by siding two corresponding 96×96 patches from I_1 and $I_{2 \rightarrow 1}^*$,
 227 then runs the original BM3D on each sub-image and finally generates the output by collecting

228 the blocks from each sub-image corresponding to the 32×32 central I_1 patches. No difference
229 in the results with respect to the standard BM3D was observed, which plausibly implies that
230 corresponding patches for I_1 and $I_{2 \rightarrow 1}^*$ are not judged as similar to each other by BM3D.

231 In the case of DIP, the borders of the input images were cropped due to network architectural
232 constraints: These missing parts were replaced with content from the original input images.

233 Concerning BOPBtL, the scratch removal option was disabled since it caused the network to
234 crash. This is a known issue related to the high memory requirement exceeding the standard 12
235 GB GPU amount to run the network on standard image input³, and does not occur only when
236 the input image size is small. To circumvent this problem, two solutions were attempted, yet
237 without satisfying results. Specifically, in the first solution the input image was rescaled to a fixed
238 size (from 50% to 33% of its original size), but the final result lost too many details (see Fig.
239 9a). In the second solution, the input was processed in separated blocks, causing a lack of global
240 consistency in the output (see Fig. 9b). Moreover, in both solutions, the chessboard artefact effect
241 typical of many deep networks that resynthesize images looked more evident than in the original
242 BOPBtL implementation. BOPBtL was employed to post-process the output of SMR+, which was
243 denoted as SMR+BOPBtL in the results (see Fig. 9c).

244 3.3. Results

245 Figures 10-11 show some examples of the results obtained with the compared methods. For
246 a thorough visual qualitative evaluation, the reader is invited to inspect the full-resolution results
247 obtained on the whole dataset, which are included in the [additional material](#). From a direct visual
248 inspection of the results, BM3D and DIP often seem to oversmooth relevant details in the image,
249 with BM3D producing somewhat better results than DIP, which sometimes simply fails to obtain
250 a reasonable output (see Fig. 11d, row DIP). BOPBtL is able to bring out fine details, providing
251 altogether a locally adaptive smoothing and contrast enhancement of the image, with satisfactory
252 results. Nevertheless, none of the previous methods is able to detect and compensate for dust,
253 scratches and other kinds of artefacts, that conversely may be even amplified in the restoration
254 process, as one can check by locating dust spots and sketches in Fig. 10e, rows BM3D, DIP and
255 BOPBtL. This problem is mostly evident for BOPBtL, where image artefacts are heavily boosted
256 together with finer details.

257 Conversely, SMR-based methods are able to solve these issues by exploiting the additional
258 information present in the auxiliary image, with exception of very severe conditions such as the
259 stains appearing in the right skyline of Fig. 11c, for which anyways SMR-based methods still get
260 the best restoration of all. SMR-based methods also successfully enhance the image contrast, as
261 it happens for the window in the dark spot under the right arcade in Fig. 10b, rows SMR and
262 SMR+. When image degradations are even more severe than that, good results can nevertheless
263 be obtained by forcing the illumination of the auxiliary image into the reference (see Sec. 2.2), as
264 done for Fig. 10d, rows SMR, SMR+ and SMR+BOPBtL. Concerning the guided supersampling
265 introduced for SMR+, this is able not only to preserve high frequency details (see again Fig. 7), but
266 also to better clean-up the image, as one can notice by inspecting the removed scratch from Fig.
267 10c, row SMR+. Guided supersampling also alleviates spurious artefacts arising from inaccurate
268 optical flow estimation as in the case of the light pole of Fig. 10a (compare rows SMR and SMR+).
269 Only in few cases of very inaccurate optical flow estimation, SMR+ is unable to fix inconsistencies
270 and generates some spurious artefacts as in the bottom-left white scratch in Fig. 11e, rows SMR+
271 and SMR+BOPBtL. Finally, it can be noted that SMR+BOPBtL is able to take the best from both
272 the methods, i.e. the artefact removal from SMR+ and the image enhancement from BOPBtL, and
273 provides very visually striking results.

274 Table 2 reports the score obtained by the compared methods on the images discussed so far
275 according to popular no-reference quality assessment metrics. Specifically, scores are reported for
276 the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [29], the Naturalness Image
277 Quality Evaluator (NIQE) [30] and Perception based Image Quality Evaluator (PIQE) [31]. Due
278 to the lack of ground-truth clean data and of a well-defined image model for the generation of
279 synthetic images with the same characteristics of the input image under evaluation, image quality
280 measurements requiring a reference image such as the Structural Similarity Index (SSIM) [32]
281 cannot be applied. By inspection of the scores obtained, it clearly emerges that these quality metrics
282 do not reflect the human visual judgment, hence they are unsuitable for a reliable quantitative
283 evaluation in this specific application scenario. In particular, there is no agreement among

³ <https://github.com/microsoft/Bringing-Old-Photos-Back-to-Life/issues/>

284 the various metrics and in about half of the cases the input image even gets a better score
285 than the restored one, in contrast with the human visual assessment. Furthermore, SMR+ and
286 SMR+BOPbTL obtain worse scores than the original images or BOPbTL in the cases where SMR-
287 methods successfully cleaned the image by removing strong image artefacts, again in contrast with
288 human judgment (see Figs. 11b,11d). A possible explanation of this behavior is that these metrics
289 only rely on low-level, local image properties and not on high-level, semantic image characteristics.
290 Hence, they are unable to distinguish between fine image details and artefacts. Nevertheless,
291 according to Table 2 SMR+, with or without BOPbTL, shows good results under these blind quality
292 assessment metrics, implying that it is able not only to remove structural artefacts from the original
293 image, but also to maintain high quality visual details besides the semantic interpretation of the
294 scene.

295 Concerning running times, BM3D, BOPbTL and DIP require respectively 10 s, 30 s and 20 min
296 on average for processing the dataset images, while SMR and SMR+ take respectively 6 min and 9
297 min. The running environment is a Ubuntu 20.04 system running on a Intel Core i7-3770K with
298 8 GB of RAM equipped with a 12 GB NVIDIA Titan XP. BM3D is a Matlab optimized .mex file,
299 BOPbTL and DIP implementations run on Pytorch exploiting GPU acceleration, while, with the
300 exception of RAFT optical flow estimation, SMR and SMR+ are based on non-optimized Matlab
301 code running on CPU. For both SMR and SMR+ the times include the image resynthesis and color
302 correction steps that take 4.5 min altogether on average. Under these considerations, both SMR
303 and SMR+ running times are reasonable for offline applications. None of the compared methods
304 can be used for real-time applications, as in the best case corresponding to BM3D, 10 s are required
305 for processing the input image.

306 4. Conclusions and Future Work

307 This paper proposed a novel method for the fully automatic restoration of historical stereo
308 photographs. By exploiting optical flow, the auxiliary view of the stereo frame is geometrically
309 and photometrically registered onto the reference view. Restoration is then carried out by fusing
310 the data from both images according to our stacked median approach followed by gradient ad-
311 justments aimed at preserving details. Guided supersampling is also introduced and successfully
312 applied for enhancing finer details and simultaneously providing a more effective artefact removal.
313 Finally, an iterative refinement step driven by a visual consistency check is performed in order to
314 remove the artefacts due to optical flow estimation errors in the initial phase.

315 Results on several historical stereo pairs show the effectiveness of the proposed approach,
316 that is able to remove most of the image defects including dust and scratches, without excessive
317 smoothing of the image content. The approach works better than its single-image denoising
318 competitors, thanks to the ability of exploiting stereo information. As a matter of fact, single-image
319 methods have severe limitations in handling damaged areas, and usually produce more blurry
320 results. Nevertheless, experimental results show that single image BOPbTL, when cascaded with
321 our approach into SMR+BOPbTL, can achieve remarkably good performances.

322 Future work will investigate novel solutions to refine the optical flow in order to reduce
323 pixel mismatches. A further research direction will be towards consolidating the stacked median
324 approach as an image blending technique. Finally, the proposed method will be extended and
325 adapted to the digital restoration of historical films.

326 **Author Contributions:** Conceptualization, F.B.; methodology, F.B. and M.F.; software, F.B. and
327 M.F.; validation, F.B., M.F. and C.C.; formal analysis, F.B., M.F. and C.C.; investigation, F.B. and
328 M.F.; resources, F.B., M.F. and C.C.; data curation, F.B., M.F. and C.C.; writing—original draft
329 preparation, F.B.; writing—review and editing, M.F., C.C. and F.B.; visualization, F.B., M.F. and
330 C.C.; supervision, F.B. and C.C.; project administration, F.B. and C.C.; funding acquisition, F.B. All
331 authors have read and agreed to the published version of the manuscript.

332 **Funding:** This work was supported by the Italian Ministry of University and Research (MUR)
333 under the program PON Ricerca e Innovazione 2014-2020, cofunded by the European Social Fund
334 (ESF), CUP B74I18000220006, id. proposta AIM 1875400, linea di attività 2, Area Cultural Heritage.

335 Data Availability Statement:

336 Additional material including code, dataset and evaluation results are freely available online
337 at <https://drive.google.com/drive/folders/1Fmsm50bMMDSd0z4JXOhCZ3hPDIXdwMUL>.

338 **Acknowledgments:** The Titan Xp used for this research was generously donated by the NVIDIA
339 Corporation.

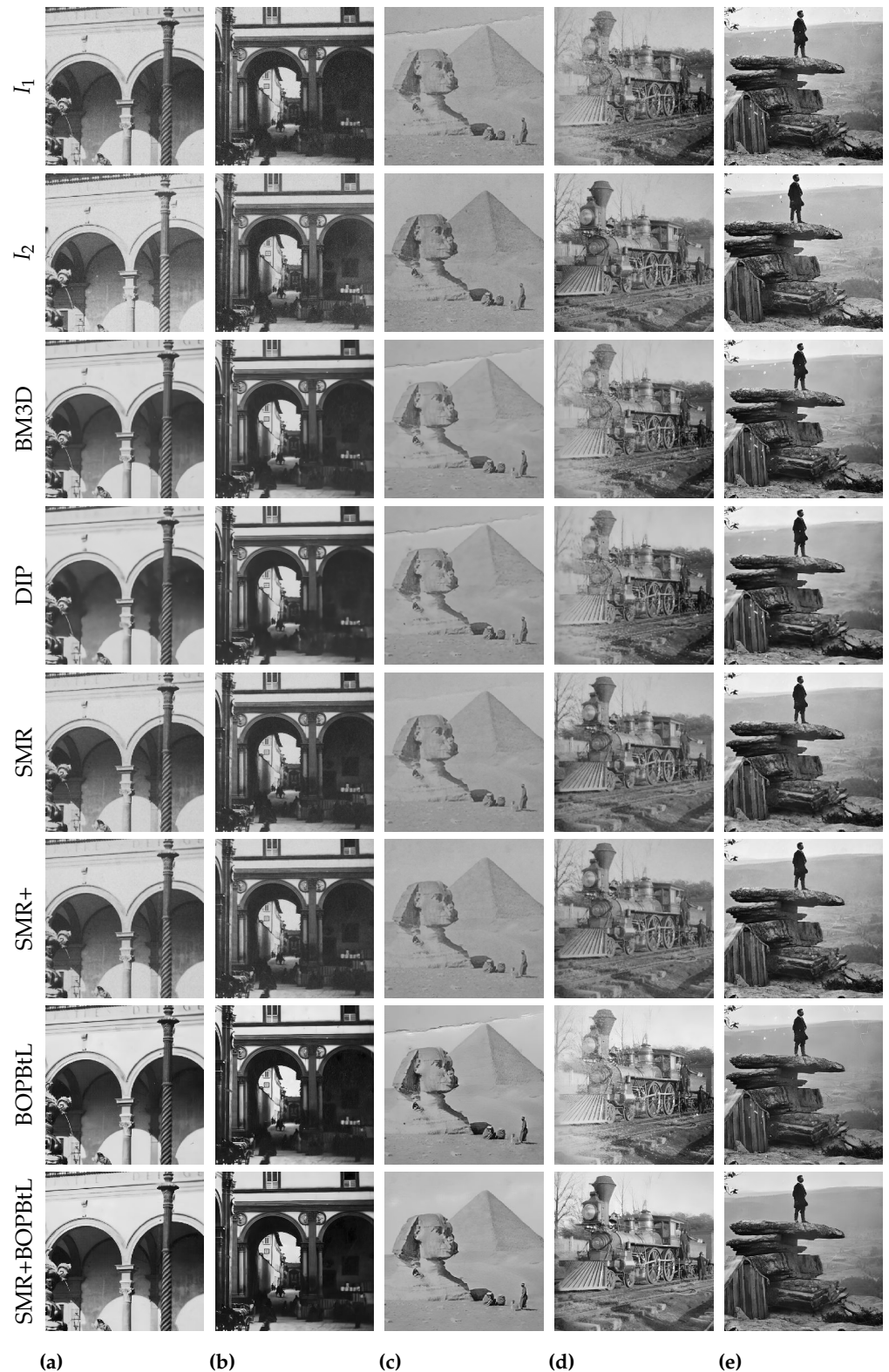


Figure 10. Qualitative visual comparison of the methods under test. Best viewed in color. The reader is invited to zoom in the electronic version of the manuscript in order to better appreciate the differences.

340 We would like to thank Drs. Costanza Caraffa and Ute Dercks at Photothek des Kunsthis-
 341 torischen Instituts in Florenz – Max-Planck-Institut for allowing the reproduction of the photos in
 342 this paper. Hautmann’s collection digital scans: ©Stefano Fancelli/KHI.



Figure 11. Qualitative visual comparison of the methods under test. Best viewed in color. The reader is invited to zoom in the electronic version of the manuscript in order to better appreciate the differences.

343 **Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the
 344 design of the study; in the collection, analyses, or interpretation of data; in the writing of the
 345 manuscript, or in the decision to publish the results.

Table 2: No-reference assessment metric results (lower values are better). Values in bold indicate the best score among the compared methods. Scores that are better in the original images than in the restored ones are underlined

		I_1	BM3D	DIP	SMR	SMR+	BOPBtL	SMR+ BOPBtL
Figs. 1,9	BRISQUE	41.89	54.34	51.47	53.11	43.46	24.15	24.20
	NIQE	4.23	5.31	5.31	5.09	3.98	4.09	4.24
	PIQE	45.97	78.93	85.33	50.60	46.35	22.55	25.90
Fig. 10a	BRISQUE	<u>10.74</u>	46.03	31.11	42.18	33.06	25.41	31.37
	NIQE	<u>2.79</u>	3.83	3.94	3.28	3.76	4.05	4.08
	PIQE	<u>25.02</u>	79.24	81.50	43.32	28.09	38.50	35.35
Fig. 10b	BRISQUE	<u>9.84</u>	48.68	35.95	41.57	29.69	14.17	34.69
	NIQE	3.16	4.07	3.92	2.92	3.34	3.65	4.01
	PIQE	29.73	78.53	78.16	37.26	23.61	29.98	34.31
Fig. 10c	BRISQUE	<u>9.26</u>	44.97	31.28	38.29	33.94	12.13	19.06
	NIQE	<u>2.79</u>	4.22	4.11	3.47	4.04	5.43	5.31
	PIQE	<u>15.80</u>	60.33	53.28	42.81	23.02	20.30	20.00
Fig. 10d	BRISQUE	14.57	31.93	22.82	36.91	25.66	15.89	10.96
	NIQE	<u>2.61</u>	3.11	3.72	3.49	3.65	3.97	3.62
	PIQE	<u>9.31</u>	43.23	52.66	38.28	24.24	10.48	11.76
Fig. 10e	BRISQUE	<u>12.85</u>	30.58	28.31	31.95	22.40	29.13	28.87
	NIQE	<u>2.17</u>	2.26	3.30	3.13	2.92	4.05	3.97
	PIQE	27.52	42.54	45.40	40.00	24.43	14.67	16.92
Fig. 11a	BRISQUE	42.58	48.03	40.26	51.88	41.23	38.48	39.21
	NIQE	<u>3.80</u>	4.77	4.97	4.66	3.93	4.57	4.75
	PIQE	26.39	74.37	79.44	45.89	36.91	13.28	14.60
Fig. 11b	BRISQUE	39.15	49.22	53.80	45.41	40.85	14.75	17.74
	NIQE	4.33	5.43	5.78	4.93	4.15	4.32	4.56
	PIQE	28.96	82.41	84.95	46.49	38.68	15.54	17.70
Fig. 11c	BRISQUE	30.43	52.90	55.07	52.86	39.59	25.54	20.06
	NIQE	<u>3.13</u>	5.22	5.53	4.25	3.20	4.59	4.36
	PIQE	<u>17.20</u>	85.95	88.53	43.98	30.33	25.39	27.83
Fig. 11d	BRISQUE	28.40	45.63	47.19	41.24	31.51	22.09	23.47
	NIQE	2.11	4.17	6.28	3.89	2.85	3.49	3.85
	PIQE	31.65	72.88	94.84	48.02	36.64	20.68	22.81
Fig. 11e	BRISQUE	40.12	38.54	37.95	20.01	22.15	38.12	22.07
	NIQE	6.27	3.49	4.08	2.84	3.06	4.60	4.42
	PIQE	58.45	51.79	48.00	19.77	13.28	13.35	11.45

References

1. Ardizzone, E.; De Polo, A.; Dindo, H.; Mazzola, G.; Nanni, C. A Dual Taxonomy for Defects in Digitized Historical Photos. 10th International Conference on Document Analysis and Recognition, 2009, pp. 1166–1170.
2. Kokaram, A.C. *Motion Picture Restoration: Digital Algorithms for Artefact Suppression in Degraded Motion Picture Film and Video*; Springer-Verlag, 1998.
3. Tegolo, D.; Isgro, F. A genetic algorithm for scratch removal in static images. International Conference on Image Analysis and Processing (ICIAP2001), 2001, pp. 507–511.
4. Stanco, F.; Tenze, L.; Ramponi, G. Virtual restoration of vintage photographic prints affected by foxing and water blotches. *Journal of Electronic Imaging* **2005**, *14*, 043008.
5. Besserer, B.; Thiré, C. Detection and Tracking Scheme for Line Scratch Removal in an Image Sequence. European Conference on Computer Vision (ECCV2004), 2004, pp. 264–275.

6. Criminisi, A.; Perez, P.; Toyama, K. Object removal by exemplar-based inpainting. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR2003)*, 2003, Vol. 2.
7. Dabov, K.; Foi, A.; Katkovnik, V.; Egiazarian, K. Image Denoising by Sparse 3D Transform-Domain Collaborative Filtering. *IEEE Transactions on Image Processing* **2007**, *16*, 2080–2095.
8. Chen, F.; Zhang, L.; Yu, H. External Patch Prior Guided Internal Clustering for Image Denoising. *IEEE International Conference on Computer Vision (ICCV2015)*, 2015, pp. 603–611.
9. Buades, A.; Lisani, J.; Miladinović, M. Patch-Based Video Denoising With Optical Flow Estimation. *IEEE Transactions on Image Processing* **2016**, *25*, 2573–2586.
10. Nasrollahi, K.; Moeslund, T.B. Super-resolution: a comprehensive survey. *Machine Vision and Applications* **2014**, *25*, 1423–1468.
11. Yang, J.; Wright, J.; Huang, T.S.; Ma, Y. Image Super-Resolution Via Sparse Representation. *IEEE Transactions on Image Processing* **2010**, *19*, 2861–2873.
12. Castellano, G.; Vessio, G. Deep learning approaches to pattern extraction and recognition in paintings and drawings: an overview. *Neural Computing and Applications* **2021**.
13. Ulyanov, D.; Vedaldi, A.; Lempitsky, V. Deep Image Prior. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR2018)*, 2018.
14. Wang, Z.; Chen, J.; Hoi, S.C.H. Deep Learning for Image Super-resolution: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2020**.
15. Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; Fu, Y. Residual Dense Network for Image Restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2020**.
16. Wan, Z.; Zhang, B.; Chen, D.; Zhang, P.; Chen, D.; Liao, J.; Wen, F. Bringing Old Photos Back to Life. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
17. Teed, Z.; Deng, J. RAFT: Recurrent All-Pairs Field Transforms for Optical Flow. *European Conference on Computer Vision (ECCV2020)*, 2020.
18. Jeon, D.S.; Baek, S.; Choi, I.; Kim, M.H. Enhancing the Spatial Resolution of Stereo Images Using a Parallax Prior. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR2018)*, 2018, pp. 1721–1730.
19. Zhou, S.; Zhang, J.; Zuo, W.; Xie, H.; Pan, J.; Ren, J.S. DAVANet: Stereo Deblurring With View Aggregation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR2019)*, 2019, pp. 10988–10997.
20. Yan, B.; Ma, C.; Bare, B.; Tan, W.; Hoi, S. Disparity-Aware Domain Adaptation in Stereo Image Restoration. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR2020)*, 2020, pp. 13176–13184.
21. Schindler, G.; Dellaert, F. 4D Cities: Analyzing, Visualizing, and Interacting with Historical Urban Photo Collections. *Journal of Multimedia* **2012**, *7*.
22. Fanfani, M.; Bellavia, F.; Bassetti, G.; Argenti, F.; Colombo, C. 3D Map Computation from Historical Stereo Photographs of Florence. *IOP Conference Series: Materials Science and Engineering* **2018**, *364*, 012044.
23. Luo, X.; Kong, Y.; Lawrence, J.; Martin-Brualla, R.; Seitz, S. KeystoneDepth: Visualizing History in 3D, 2019, [[1908.07732](#)].
24. Fanfani, M.; Colombo, C.; Bellavia, F. Restoration and Enhancement of Historical Stereo Photos through Optical Flow. *Proceedings of the ICPR Workshop on Fine Art Pattern Extraction and Recognition (FAPER)*, 2021.
25. McCann, J.J.; Rizzi, A. *The Art and Science of HDR Imaging*; Wiley & Sons, 2011.
26. Sherrod, A. *Game Graphic Programming*.
27. Chambon, S.; Crouzil, A. Similarity measures for image matching despite occlusions in stereo vision. *Pattern Recognition* **2011**, *44*, 2063–2075.
28. Bellavia, F.; Colombo, C. Dissecting and Reassembling Color Correction Algorithms for Image Stitching. *IEEE Transactions on Image Processing* **2018**, *27*, 735–748.
29. Mittal, A.; Moorthy, A.K.; Bovik, A.C. No-Reference Image Quality Assessment in the Spatial Domain. *IEEE Transactions on Image Processing* **2012**, *21*, 4695–4708.
30. Mittal, A.; Soundararajan, R.; Bovik, A.C. Making a “Completely Blind” Image Quality Analyzer. *IEEE Signal Processing Letters* **2013**, *20*, 209–212.
31. Venkatanath, N.; Praneeth, D.; Chandrasekhar, B.; Channappayya, S.; Medasani, S. Blind image quality evaluation using perception based features. *21st National Conference on Communications (NCC)*, 2015, pp. 1–6.
32. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* **2004**, *13*, 600–612.