

(51) International Patent Classification:
G06T 3/40 (2006.01)(21) International Application Number:
PCT/EP2014/051117(22) International Filing Date:
21 January 2014 (21.01.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
13305084.9 24 January 2013 (24.01.2013) EP
13305956.8 5 July 2013 (05.07.2013) EP(71) Applicant: THOMSON LICENSING [FR/FR]; 1-5 rue
Jeanne d'Arc, F-92130 Issy-les-Moulineaux (FR).(72) Inventors: TÜRKAN, Mehmet; 13 rue du Bois Perrin, F-
35700 Rennes (FR). THOREAU, Dominique; 39 rue du
Réage, F-35510 Cesson Sevigne (FR). GUILLOTEL,
Philippe; 15 rue Auguste Renoir, F-35770 Vern sur Seiche
(FR).(74) Agent: KÖNIG, Uwe; Karl-Wiechert-Allee 74, 30625
Hannover (DE).(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,
BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM,
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR,
KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME,
MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ,
OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA,
SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM,
TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM,
ZW.(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ,
UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ,
TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,
EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,
MC, MK, MT, NL, NO, PL, PT, QA, RO, RS, SE, SI, SK, SM,
TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
KM, ML, MR, NE, SN, TD, TG).**Published:**

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the
claims and to be republished in the event of receipt of
amendments (Rule 48.2(h))

(54) Title: METHOD AND APPARATUS FOR PERFORMING SUPER-RESOLUTION OF SINGLE IMAGES

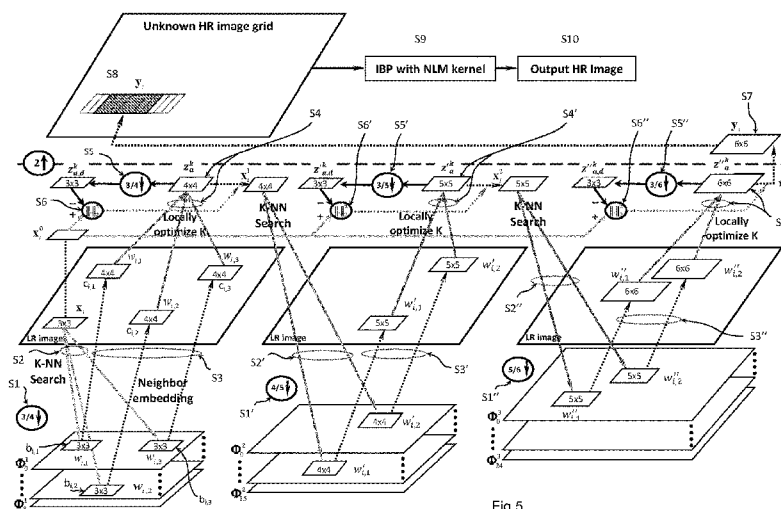


Fig. 5

(57) **Abstract:** A method for performing super-resolution of a single image comprises steps of downsampling (S₁) a LR input image by a first factor, separating the LR input image into overlapping blocks (x_i), and for each of the blocks (x_i) searching similar blocks in the downsampled LR images, determining blocks (c_{i,1}, c_{i,2}, c_{i,3}) at corresponding positions in the LR input image, locally optimizing (S₄) the blocks by linear approximation to obtain locally optimized patches (z^k_a), downsampling (S₅) the locally optimized patches (z^k_a) by the first factor to obtain a downsampled locally optimized patch (z^k_{a,d}), comparing the downsampled locally optimized patch (z^k_{a,d}) with the input block (xⁱ₀) and optimizing the number and/or weights (w_{i,1}, w_{i,2}, w_{i,3}) of the determined blocks (c_{i,1}, c_{i,2}, c_{i,3}), wherein a locally optimized patch (xⁱ₁) is obtained, and repeating the above steps with different sampling factors, wherein for each repetition the locally optimized patch (x¹_i, x²_i, x³_i) of the previous iteration is used as input block. The locally optimized patch (x³_i) of the last iteration are accumulated and averaged (S₈), and IBP (S₉) is performed to obtain a HR output image (S₁₀).

METHOD AND APPARATUS FOR PERFORMING SUPER-RESOLUTION OF SINGLE IMAGES

Field of the invention

5

This invention relates to a method for performing super-resolution of single images, and to an apparatus for performing super-resolution of single images.

Background

10

Super-resolution (SR) is an inverse problem of recovering a high-resolution (HR) image from one or more low-resolution (LR) images. HR means high pixel density in an image, hence an HR image provides more clear and detailed contents to the viewer. A number of real-world applications utilizes SR algorithms, as SR provides
15 a relatively low-cost software solution to this inverse problem when compared to high-cost hardware enhancements in optics and sensors technology.

In order to recover the unknown HR image, conventional SR methods (e.g., [1,2, 3]) fuse a set of observed LR images at sub-pixel shifts. If there are enough LR
20 images, the missing high-frequency information can successfully be recovered in the HR image. However, an insufficient number of LR observations, registration (motion estimation) errors and unknown point spread function practically limit the applicability of these classical multi-image SR methods to small upscaling ratios which is typically less than two [4].

25

Exemplar-based methods have been developed to overcome limitations of classical multi-image SR. These methods assume that the missing high-frequency information is explicitly available in other natural images and it can successfully be retrieved from exemplar LR and HR patch pairs. In [5], LR and HR patch pairs are
30 collected from other natural images, and low- and high-frequency relations of these patches are learned via a Markov network. This method has later been simplified in [6] with a sequence of estimations of HR patches by a nearest neighbor (NN) search from the database of exemplars, i.e. examples. A manifold

learning based SR approach has been proposed in [7]. In this method, each LR patch is first reconstructed from its K nearest neighbors (K -NN) taken from the exemplar database. The HR patch is then *hallucinated* (or estimated) with the corresponding HR exemplar patches by considering geometric similarities of LR and HR patch spaces (manifolds). A sparse representations based SR method has been proposed in [8]. This method jointly learns two connected dictionaries by enforcing the similarity between sparse representations of exemplar LR and HR patch pairs. In this way, the sparse representation of a given LR patch over the learned LR dictionary can be applied with the learned HR dictionary to hallucinate the unknown HR patch.

A unified framework combining the classical multi-image and exemplar-based methods has been proposed in [9] by considering SR from a single LR image. This method first searches image patches in downscaled versions of a given LR image. When a NN is found for the current LR patch, its *parent* in the LR image is regarded as the HR patch and copied to the appropriate location in the appropriate scale of the HR image. The obtained different scales of HR images have finally been generated coarse-to-fine until the desired resolution is reached. Another approach of SR from a single LR image has been described in [10] by considering a *local self-similarity* assumption on natural images. The given LR image is initially interpolated using a custom designed linear operator. After carrying out a strictly-localized NN search in the low-frequency band, the missing high-frequency information of the interpolated image is retrieved from the high-frequency band of the LR image. A similar SR technique has been proposed in [11], exploiting local cross-scale self-similarities to extrapolate the high-frequency band of the given LR image. This algorithm additionally imposes a total variation [12] regularization in its stages for a robust resolution enhancement.

Summary of the Invention

The present invention provides an improved self-contained and exemplar-based framework for single-image SR relying on multi-scale and locally-optimized neighbor embeddings of LR and HR patches. Following the recurrence property of similar patches across different scales of an image, it uses only the observed LR

image and its appropriately downscaled versions for SR. Given a LR patch, its size is gradually increased by relying on local geometric similarities of low- and high-resolution patch spaces under small scaling factors. In each step of the algorithm, the local geometry of the given patch and its K-NN patches is
 5 characterized using locally linear embedding (LLE) [13]. In order to assure the local image compatibility, the parameter K is locally optimized. The invention allows averaging overlaps between adjacent patches, which improves the smoothness of the target HR image. Finally, the global consistency between the output HR image and the input LR image is enforced by using an iterative back
 10 projection (IBP) [2] with a non-local means (NLM) [14] kernel. The proposed algorithm can easily be ported and parallelized on GPU processors. Moreover, an advantage confirmed by experimental results is that it produces more natural looking textures together with cleaner and sharper edges in the resulting HR images, compared to other known SR methods.

15 In one embodiment, the present invention provides a method for performing super-resolution of single images. In another embodiment, the present invention provides an apparatus for performing super-resolution of single images.

20 In one embodiment, the invention provides a computer readable medium having executable instructions to cause a computer to perform said method for performing super-resolution of single images.

In principle, a method for performing super-resolution of a single image comprises
 25 steps of downsampling S_1 a LR input image by a first factor, separating the LR input image into overlapping blocks x_i , and for each of the blocks x_i searching similar blocks in the downsampled LR image, determining blocks $c_{i,1}, c_{i,2}, c_{i,3}$ at corresponding positions in the LR input image, locally optimizing S_4 the blocks by linear approximation to obtain locally optimized patches z_a^k , downsampling S_5 the
 30 locally optimized patches z_a^k by the first factor to obtain a downsampled locally optimized patch $z_{a,d}^k$, comparing the downsampled locally optimized patch $z_{a,d}^k$ with the input block x_0^i and optimizing the number and/or weights $w_{i,1}, w_{i,2}, w_{i,3}$ of the determined blocks $c_{i,1}, c_{i,2}, c_{i,3}$, wherein a locally optimized patch x_i^1 is

obtained, and repeating the above steps with different sampling factors, wherein for each repetition the locally optimized patch x_i^1, x_i^2, x_i^3 of the previous iteration is used as input block. The locally optimized patch x_i^3 of the last iteration are accumulated and averaged S_8 , and iterative back projection S_9 is performed to
5 obtain a HR output image S_{10} .

In one embodiment, the full (or at least a more sophisticated) super-resolution processing is done only for blocks with high texture, while a simplified super-resolution processing is applied for blocks with low texture. In one embodiment,
10 the super-resolution processing is done only for blocks with high texture, while it is skipped for blocks with low texture.

Further objects, features and advantages of the invention will become apparent from a consideration of the following description and the appended claims when
15 taken in connection with the accompanying drawings.

Brief description of the drawings

Exemplary embodiments of the invention are described with reference to the
20 accompanying drawings, which show in

- Fig.1 how exemplar LR and HR patch pairs are collected directly from the input LR image and its appropriately downscaled version;
Fig.2 a flow-chart (part 1) of the method in a first embodiment of the invention;
25 Fig.3 a flow-chart (part 2) of the method in the first embodiment of the invention;
Fig.4 a flow-chart (part 3) of the method in the first embodiment of the invention;
Fig.5 an overview of the method in one embodiment of the invention;
Fig.6 an apparatus according to one embodiment of the invention;
Fig.7 a comparison of experimental results for fine-detailed texture areas;
30 Fig.8 a comparison of experimental results for reducing jaggy and ringing artifacts;
Fig.9 experimental results showing sharper and cleaner edges as compared to results from [11];

Fig.10 experimental results showing sharper and cleaner edges as compared to results of bi-cubic interpolation;

Fig.11 a flow-chart of a method in one embodiment of the invention where highly textured blocks and lowly textured blocks are treated different; and

5 Fig.12 an apparatus according to one embodiment of the invention, where highly textured blocks and lowly textured blocks are treated different.

Detailed description of the invention

10 In the following, the underlying ideas and main steps of the proposed SR algorithm are described, and the obtained experimental results are demonstrated.

First, single-image super-resolution is explained.

The aim of single-image SR is to recover a HR image which is consistent with the given LR image according to an image generation model. In general, the
15 observed LR image $\mathbf{X}$ of size $N \times M$ is assumed to be a blurred and downsampled version of the unknown HR image $\mathbf{Y}$ (of size $SN \times SM$), hence the image generation model can be written as

$$\mathbf{X} = (\mathbf{Y} * \mathbf{H}) \downarrow S \quad (1)$$

20 where $\mathbf{H}$ is a low-pass (blur) kernel, and $*$ and $\downarrow S$ represent convolution and downscaling (with a factor of S) operators respectively. The inverse problem, estimating $\mathbf{Y}$ for a given $\mathbf{X}$, is highly ill-posed since there might be multiple HR solutions (under different initial assumptions) which produce the same LR observation $\mathbf{X}$. In order to regularize the problem, a global energy function
25 incorporating a prior on $\mathbf{Y}$ is generally formulated and minimized as

$$\operatorname{argmin}_y \|\mathbf{X} - (\mathbf{Y} * \mathbf{H}) \downarrow s\|_2^2 + \gamma R(\mathbf{Y}) \quad (2)$$

where $\|\mathbf{X} - (\mathbf{Y} * \mathbf{H}) \downarrow s\|_2^2$ is the data term with regard to eq. (1), $R(\mathbf{Y})$ is an applicable regularization term based on a prior knowledge on $\mathbf{Y}$ and γ is a parameter controlling the contribution of the prior on the solution. As most of the
30 natural images vary smoothly with a limited local activity, $R(\mathbf{Y})$ is usually chosen to penalize the variation of $\mathbf{Y}$.

According to the invention, the above SR optimization is solved in two successive steps; first, by calculating an initial estimate Y_0 of the HR image to satisfy a patch-based local constraint through a prior $R(Y)$ as

$$Y_0 = \arg \min_y R(Y) \quad (3)$$

- 5 and then, by minimizing this data term iteratively by projecting Y_0 onto the solution space of (1) in order to satisfy the image generation model exactly, i.e.,

$$\arg \min_Y \|Y - Y_0\|_2^2 \text{ subject to } (Y * H) \downarrow s = X \quad (4)$$

Next, local priors from optimized neighbor embeddings are described.

- 10 Consider a lexical ordering of image pixels and assume image patches as column vectors. Denote an LR patch (of size $n \times n$) in X centered at pixel i as x_i , and its spatially corresponding unknown HR patch (of size $S_n \times S_n$) in Y as y_i , $i = 1 \dots NM$. To recover the HR patch y_i (for all i), the LR patch x_i (for all i) is iteratively upscaled by a small factor at each step t of the algorithm until the desired scaling
- 15 factor S has been reached. In the following, details of the underlying ideas of the proposed local prior are given.

- First, exemplar patch pairs from a single image are considered. The underlying main idea is based on the assumption that small image patches are very likely to
- 20 repeat themselves within and across different scales of an image [9]. Hence, exemplar LR and HR patch pairs are collected directly from the given LR image and its appropriately downscaled version by matching LR patches (located in the downscaled image) and their spatially corresponding HR parents (suited in the input LR image). Fig.1 shows how exemplar LR and HR patch pairs are collected
- 25 directly from the input LR image 10 and its appropriately downscaled version 11. E.g. LR patch 13 and HR patch 15 are a (LR and HR) patch pair, and LR patch 14 and HR patch 15 are a patch pair. Each LR patch in the downscaled image 11 has a corresponding HR patch in the input image 10.

- A fundamental observation that has been reported in [10] is that natural images
- 30 can preserve their various important features only under small scaling factors. Therefore, the input LR image is downscaled with a sufficiently reasonable ratio in order to extract LR and HR patch pairs of greater relevance, and moreover, to maintain the local geometric similarities of LR and HR patch spaces (as explained

further below). For this purpose, the size of a given LR patch is increased gradually, e.g., by p pixels in each dimension (p is admissibly smaller than n ; e.g. $p=1,2$ or 3) at each step of the algorithm. In one embodiment, the size of a given LR patch is increased gradually by p_h pixels in horizontal dimension and by p_v pixels in vertical dimension (p_h not equal p_v) in at least one step of the algorithm. However, usually p_h equals p_v .

Now consider a given LR patch $\mathbf{x}_i$ of size $n \times n$. At each step $t > 0$, expand the size incrementally by p -pixels to obtain $\mathbf{x}_i^t$ of size $(n + tp) \times (n + tp)$ using $\mathbf{x}_i^{t-1}$ of size $(n + (t-1)p) \times (n + (t-1)p)$. To do this, collect HR exemplars of sizes $(n + tp) \times (n + tp)$ from the input LR image and their matching LR exemplars of sizes $(n + (t-1)p) \times (n + (t-1)p)$ from the downsampled LR image by a factor of $s_t = (n + (t-1)p) / (n + tp)$.

As the input LR image is downsampled by a non-integer value s_t , irregular grid points of the two images are related in order to ensure that the collected LR and HR patches are *in phase*. Because of the partial lack of translation invariance, we inevitably need to search and match exemplars in regular offsets (in both images), which will lead to a drastic reduction in the number of extracted exemplar pairs. In order to maximize the amount of the data present at a step t , multiple downsampled images Φ_q^t are created by offsetting the input LR image $\mathbf{X}$ by $0, \dots, n + (t-1)p - 1$ pixels in each dimension, and then collect exemplar patches with regular offsets (depending on s_t) from image pairs $\{\mathbf{X}, \Phi_q^t\} \forall q$ ($\forall$ means “for all”) while keeping the phases of LR and HR patch pairs always aligned together. Denote these pairs in a set as $\{z_j^{l,t}, z_j^{h,t}\} \forall t$.

Now the term “partial lack of translation invariance” is explained. As the LR image is downsampled with non-integer factors, it may happen that not all patches in the LR image and downsampled image correspond to each other because of the irregular grid relation. For example, when downsampling by $\frac{3}{4}$, patches need to be related (to collect exemplars) by shifting 4 pixels (offset) in the LR image and 3 pixels (offset) in the downsampled image. The patches in between these offsets are related to each other in sub-pixel precision. So, there is a partial lack of translation invariance, as there is no one-to-one correspondence between all

kinds of translations of patches in both images. The two images are translation invariant just with suitable offsets. This is actually related with phase information of pixels. According to the invention, the reduction of the number of exemplars is compensated by creating different downsampled images with different phases.

5

Next, optimized neighbor embeddings are described.

As a local prior, each local HR patch should resemble its neighboring (similar) patches taken from the given exemplar set. For this purpose, the regularization term is formulated with a neighbor embedding optimization such that each HR

10 patch $x_i^t \forall i$ at step t can optimally be hallucinated (i.e. estimated or modeled) as a weighted linear combination of its K -NN patches $\{z_{i,k}^{h,t}\}_{k=1}^K$ as

$$R(x_i^t) = \|x_i^t - \sum_k w_{i,k} z_{i,k}^{h,t}\|_2^2 + \alpha \|1 - \sum_k w_{i,k}\|_2^2 \quad (5)$$

where $w_{i,k}$ denotes the weight assigned for the k -th NN $z_{i,k}^{h,t}$ of x_i^t . The second term in eq. (5) represents a *sum-to-one constraint* on the weighting coefficients.

15 This constraint enforces the approximation of x_i^t to lie in the sub-space spanned by its K -NN, and the invariance to translations of the data point and its neighbors.

Similar to other manifold learning based SR methods, the invention benefits from the local geometric similarities of LR and HR patch spaces, and relates the
20 intrinsic geometric properties of local LR and HR patch neighborhoods using the neighbor embedding assumption. Here, it is also reasonable to assume a stronger neighbor embedding relation as the collected exemplar patches are of greater relevance under small scaling factors (see above in Exemplar patch pairs from a single image). Hence, begin transforming eq. (5) into an *approximate* LR one, such that

25

$$R(x_i^{t-1}) = \|x_i^{t-1} - \sum_k w_{i,k} z_{i,k}^{l,t}\|_2^2 + \beta \|1 - \sum_k w_{i,k}\|_2^2 \quad (6)$$

To characterize the intrinsic properties of the LR neighborhood, eq. (6) can optimally be minimized and solved for the weighting coefficients $w_i = [w_{i,1}, \dots, w_{i,K}]^T$
30 using LLE as $w_i = (G_i^{-1} \mathbf{1}) / (\mathbf{1}^T G_i^{-1} \mathbf{1})$ where G_i denotes the Gram matrix of K -NN patches $\{z_{i,k}^{l,t}\}$ in reference to x_i^{t-1} and $\mathbf{1}$ is the column vector of ones. Assuming

that the local LR geometry has been preserved in the HR patch space, the HR embedding x_i^t can then be estimated by

$$x_i^t = \sum_{k=1}^K w_{i,k} z_{i,k}^{h,t} \quad (7)$$

In practice, transforming eq. (5) into eq. (6) does not necessarily secure the local neighborhood relation in both LR and HR patch spaces, i.e., HR exemplar patches $\{z_{i,k}^{h,t}\}$ corresponding to the K -NN $\{z_{i,k}^{l,t}\}$ of x_i^{t-1} may not be the K -NN for x_i^t because of the noise and other reasons. As a result, the calculated weights $\{w_{i,k}\}$ would not be the optimum for hallucinating the HR patch x_i^t in (7). In other words, the local prior $R(x_i^{t-1})$ in (6) does not put any constraint on the estimated values of x_i^t except the initial neighbor embedding assumption between LR and HR patch spaces. In order to improve the localized estimation of x_i^t , we introduce a new regularization term by enforcing the hallucinated HR patch x_i^t to resemble the original LR correspondence x_i^0 according to the image generation model as

$$\hat{R}(x_i^{t-1}) = R(x_i^{t-1}) + \eta \|x_i^0 - P^c [(Px_i^t * h) \downarrow_{\hat{s}_t}]\|_2^2 \quad (8)$$

where h is a low-pass kernel, P transforms stacked (vectorized) values into patches with respective sizes, P^c is the inverse operator of P and $\hat{s}_t$ is an appropriate scaling factor, i.e., $\hat{s}_t = s_1 \times s_2 \times \dots \times s_t$. For example, if $S=2$, $p=1$, $n=3$, then three iterations are needed to upscale the LR patch for $S=2$ with $p=1$: $s_1 = 3/4$, $s_2=4/5$, $s_3=5/6$, then $\hat{s}_1 = s_1 = 3/4$, $\hat{s}_2 = s_1 \times s_2 = 3/5$ and $\hat{s}_3 = s_1 \times s_2 \times s_3 = 3/6$.

In order to minimize eq. (8), an iterative algorithm based on an optimization on the number of used NN patches is used. At any step t , the method proceeds as follows. For each given LR patch x_i^{t-1} , first a K -NN search in the given exemplar set of LR patches is made. Denote the obtained LR patches and their HR correspondences as $\{z_{i,k}^{l,t}, z_{i,k}^{h,t}\}_{k=1}^K$. Then vary the number κ of used NN patches from 1 to K and optimize $R(x_i^{t-1})$ for each distinct value of κ independently. As a result, K possible estimations $\{x_i^{t,\kappa}\}_{\kappa=1}^K$ of the HR signal x_i^t are obtained, i.e.

$$x_i^{t,\kappa} = \sum_{k=1}^{\kappa} w_{i,k}^* z_{i,k}^{h,t} \quad (9)$$

where $\{w_{i,k}^*\} = \arg \min_{\{w_{i,k}\}} R(x_i^{t-1})$ and $k = 1, \dots, \kappa$, $\forall \kappa$.

We finally obtain an estimation of the HR patch x_i^t according to the local patch compatibility criterion by

$$x_i^t = \arg \min_{x_i^{t,\kappa}} \|x_i^0 - P^c [(Px_i^{t,\kappa} * h) \downarrow_{\hat{s}_t}]\|_2^2 \quad (10)$$

The above patch-based optimization on x_i^{t-1} to obtain x_i^t is iteratively applied until the desired scaling factor S has been reached, e.g., if τ iterations are required for an upscaling of S , then $y_i = x_i^\tau \forall i$.

10 Now, it is explained how to come from local prior to global smoothness. Since the optimized neighbor embedding of LR and HR patches have been calculated for each single pixel i in the LR image, we implicitly enforce as much as possible the overlaps between adjacent patches in the target HR image. In this manner, the pixel values in the HR grid have been estimated multiple times. These estimations
15 are averaged in order to force all image patches to agree on the overlapped areas, hence to satisfy the smoothness constraint in the HR image. In other embodiments, other weighting strategies than uniform averaging can be used, but this means an increase in the computational complexity.

20 In the following, it is explained how global consistency can be improved through back projection, in particular iterative back projection (IBP).

In order to enforce the global consistency between the recovered HR image $\mathbf{Y}$ and the input LR image $\mathbf{X}$, equation (4) is solved by employing the known IBP algorithm. Denote $\mathbf{Y}_0$ as the initial HR image as recovered above. The update
25 equation for IBP is given as

$$\mathbf{Y}_m = \mathbf{Y}_{m-1} + [(\mathbf{X} - (\mathbf{Y}_{m-1} * \mathbf{H}) \downarrow_s) \uparrow_s] * \mathbf{P} \quad (11)$$

where $\mathbf{Y}_m$ is the estimated HR image at iteration m , $m = 1 \dots M$, $\uparrow_s$ represents the upscaling (with a factor of S) operator and $\mathbf{P}$ is a back-projection kernel. Typically, $\mathbf{P}$ is chosen to be a fixed isotropic kernel (isotropic kernel is a 2D filter which is
30 symmetric in all dimensions, or circularly symmetric). However, edges usually have anisotropic structures, and the content of the projected data is dynamic along the iterations of IBP. As a result, this kind of kernel often produces results with jaggy and ringing artifacts especially around edges. To guide IBP in a wise

manner, an NLM-based kernel (non-local means [14]) is used instead, because it is more adapted to the content of the data to be processed. After M iterations of IBP, we take the result Y_M as our final estimate of the HR image.

- 5 Figs.2-4 show an exemplary flow-chart according to one embodiment of the invention. Fig.5 shows an overview of one embodiment of the invention (for $\{S, p, n\} = \{2, 1, 3\}$, exemplary). Reference signs correspond in all figures. While three iterations are shown (first iteration S_1 - S_6 , second iteration S_1' - S_6' and third iteration S_1'' - S_6''), it is also possible to use more or less iterations, depending on a
 10 desired scaling factor and on a maximum scaling factor per iteration.

Some of the particularly important features of the invention are the following. The size of an LR patch is gradually increased (e.g., 1-pixel at a time or per iteration), until the required scale is reached.

- 15 The LR image is downsampled with non-integer factors and example patches are collected, taking the phase information into account.
 The K value (for K-nearest neighbor or K-NN search) is optimized by checking the compatibility of the HR patch at each step.
 Overlapping areas are simply averaged.
 20 A back-projection method with non-local means based kernel is used.
 In one embodiment, the super-resolution processing is simplified or skipped for low textured portions of the image.

- As shown in Fig.5 and according to one embodiment of the invention, super-
 25 resolution of a single image is performed by downsampling S_1 a LR input image by a first small factor (e.g. $\frac{3}{4}$) several times with different phases to obtain several different downsampled LR images Φ_0^1 - Φ_8^1 , separating the LR input image into overlapping blocks (e.g. 3x3 pixels each) x_i , for each of the blocks x_i searching one or more similar (preferably equal, or substantially equal) blocks of same size
 30 $b_{i,1}, b_{i,2}, b_{i,3}$ in one or more or each of the downsampled LR images, and determining blocks $c_{i,1}, c_{i,2}, c_{i,3}$ at corresponding positions in the LR input image (wherein the blocks in the LR image are upsampled by the reciprocal of the first small factor, e.g. $\frac{4}{3}$ so that they have a size of e.g. 4x4 pixels; they are

determined just by their position). These blocks are locally optimized S_4 by linear approximation as described above to obtain locally optimized patches z_a^k and then downsampled S_5 by the same small factor as previously (e.g. $\frac{3}{4}$) to obtain $z_{a,d}^k$.

This patch is then compared to the original patch x_0^i and optimized to minimize the
 5 SSD (sum of squared distances) by varying the number of used patches $b_{i,1}$, $b_{i,2}$, $b_{i,3}$, or the weights $w_{i,1}$, $w_{i,2}$, $w_{i,3}$ of the used patches $b_{i,1}$, $b_{i,2}$, $b_{i,3}$, or both (one or more of) the patches $b_{i,1}$, $b_{i,2}$, $b_{i,3}$ and (one or more of) their weights $w_{i,1}$, $w_{i,2}$, $w_{i,3}$.

A flow-chart of one embodiment of the invention is shown in Fig.11. In this
 10 embodiment, blocks with higher texture are treated different from blocks with lower texture. Each block is determined to be either highly textured or lowly textured. Super-resolution of a single image is performed by downsampling S_1 a LR input image by a first small factor (e.g. $\frac{3}{4}$) several times with different phases to obtain several different downsampled LR images $\Phi_0^1 - \Phi_8^1$, separating the LR
 15 input image into overlapping blocks (e.g. 3x3 pixels each) x_i , and determining highly textured blocks and lowly textured blocks respectively among the overlapping blocks x_i . The method further includes for each of the lowly textured blocks x_i performing a simplified upsampling method, and for each of the highly textured blocks x_i searching one or more similar (preferably equal, or substantially
 20 equal) blocks of same size $b_{i,1}$, $b_{i,2}$, $b_{i,3}$ in one or more of each of the downsampled LR images, and determining blocks $c_{i,1}$, $c_{i,2}$, $c_{i,3}$ at corresponding positions in the LR input image (wherein the blocks in the LR image are upsampled by the reciprocal of the first small factor, e.g. $\frac{4}{3}$ so that they have a size of e.g. 4x4 pixels; they are determined just by their position). These blocks are locally optimized S_4 by
 25 linear approximation as described above to obtain locally optimized patches z_a^k and then downsampled S_5 by the same small factor as previously (e.g. $\frac{3}{4}$) to obtain $z_{a,d}^k$. This patch is then compared to the original patch x_0^i and optimized to minimize the SSD (sum of squared distances) by varying the number of used patches $b_{i,1}$, $b_{i,2}$, $b_{i,3}$, or the weights $w_{i,1}$, $w_{i,2}$, $w_{i,3}$ of the used patches $b_{i,1}$, $b_{i,2}$, $b_{i,3}$, or
 30 both (one or more of) the patches $b_{i,1}$, $b_{i,2}$, $b_{i,3}$ and (one or more of) their weights $w_{i,1}$, $w_{i,2}$, $w_{i,3}$.

For the lowly textured blocks, any simplified upsampling method can be used that reduces the computation effort. For example, linear, bilinear, cubic or bi-cubic

interpolation can be used for the lowly textured blocks. In one embodiment, the upsampling method for lowly textured blocks is similar to the upsampling method of the highly textured blocks, except at least one of the following: for each of the lowly textured blocks x_i , searching less blocks (than for the highly textured blocks) of same size $b_{i,1}, b_{i,2}, b_{i,3}$ that are similar or substantially similar to the lowly textured block, searching in less of the downsampled LR images (than for the highly textured blocks), using simplified local optimization as compared with the local optimization for the highly textured blocks, and using less patches $b_{i,1}, b_{i,2}, b_{i,3}$ or using a limited range of weights $w_{i,1}, w_{i,2}, w_{i,3}$ for the used patches $b_{i,1}, b_{i,2}, b_{i,3}$, or both, in the optimization to minimize the SSD, or skip the optimization to minimize the SSD.

A description of a separation of the blocks into highly textured and lowly textured blocks is given below.

15

With the resulting locally optimized patch x_i^1 , the above-described procedure is then repeated at least once. However, the resulting locally optimized patch x_i^1 that serves as input has now more pixels (e.g. 4x4) than the patch x_i^0 that was used in the previous iteration, and therefore also the new LR image and its downsampled versions are larger, the downscaling factor is different (e.g. 4/5) and more downsampled versions $\Phi_0^2 - \Phi_{15}^2$ exist because more positions and thus more phases (e.g. 4x4=16 phases for a 4x4 pixel patch) of the patch are possible.

20

Finally, the patches are super-imposed and averaged S_8 , as described above, and the IBP S_9 is performed to obtain the HR output image S_{10} . It is a super-resolution version of the LR input image.

25

The above-mentioned searching one or more similar blocks of same size $b_{i,1}, b_{i,2}, b_{i,3}$ in one or more or each of the downsampled LR images can use any commonly known search algorithm, e.g. as used in motion estimation (e.g. SSD).

30

In one embodiment, the accumulating and averaging S_8 comprises, for each pixel, counting overlapping patches that contribute to the pixel, adding values of

overlapping patches that contribute to the pixel and dividing the accumulated values by the number of overlapping patches.

Further, it is important to note the following.

- 5 In each iteration (first iteration S_1 - S_6 , second iteration S_1' - S_6' and third iteration S_1'' - S_6''), there are several downscaled images (e.g. Φ_0^1 - Φ_8^1 in the first iteration S_1 - S_6), since the input image is downscaled with non-integer values. These downscaled images correspond to different phases of pixels caused by this non-integer downsampling. These provide the “neighbors” of the neighbor embedding.
- 10 In steps S_6, S_6', S_6'' a simple norm calculation is performed. It may be SSD (sum of squared distance). Optimization can be done by minimizing this norm (SSD).

- Exemplary upscaling factors shown in Fig.2-5 are 3:4, 4:5 and 5:6, i.e. one additional pixel in each dimension per iteration. However, as described above,
- 15 can be generalized to be p pixels. p is considerably smaller than the block size, e.g. not more than $\frac{1}{2}$ of the block size. E.g. it can be 1,2,3,... depending on the block size (n) and computational sources. For example, for a 3x3 low-resolution patch, three iterations are used to make it 6x6 when $p=1$. So, the patch size is increased by p pixels until the required scaling is reached. As another example,
 - 20 for a scaling factor of three, if $p=2$ and $n=3$, factors would be 3/5, 5/7, 7/9, i.e. two additional pixel in each dimension per iteration.

In step S_7 , the algorithm simply stops if the required scaling has been reached and y_i is assigned to the final result.

25

- Non-local means kernel in step S_9 is a well-known adaptive filter for denoising applications. In the above-described embodiment, the back-projection kernel was changed with the Non-local means kernel, as it gives better results than using a Gaussian filter. Back projection is generally well known, but here it is a part of the
- 30 algorithm and the kernel is NLM-based. In another embodiment however, back-projection kernel may be used instead.

Fig.6 shows, in one embodiment, an apparatus for performing super-resolution of a single image, comprising a separator SEP_1 for separating the LR input image into overlapping blocks x_i , one or more (here three) cascaded upscaling units CAU_{1-3} , a HR image assembling unit IAU, and an IBP unit IU. Each cascaded upscaling unit provides its output x_1^i, x_2^i, x_3^i as input to the next upscaling unit. Each cascaded upscaling unit CAU_{1-3} comprises one or more first downsamplers DU_1 for downsampling a LR input image by a first factor several times with different phases to obtain several different downsampled LR images $\Phi_0^1-\Phi_8^1$, a first selector SEL_1 for selecting a current block x_0^i , one or more first search units SU_1 for searching, for each current block x_0^i of all the blocks x_i , one or more similar blocks of same size $b_{i,1}, b_{i,2}, b_{i,3}$ of a current block x_0^i in one or more or each of the downsampled LR images, one or more first position determining units PDU_1 for determining blocks $c_{i,1}, c_{i,2}, c_{i,3}$ at corresponding positions in the LR input image, wherein the blocks in the LR image are upsampled in one or more up-samplers UPS by the reciprocal of the first factor, a first optimizer OPT_1 for locally optimizing the blocks by linear approximation to obtain locally optimized patches z_a^k (as in step S₄ of the method described above), a second downsampler DU_2 for downsampling the locally optimized patches z_a^k by the first factor to obtain a downsampled locally optimized patch $z_{a,d}^k$, a first comparator CU_1 for comparing the downsampled locally optimized patch $z_{a,d}^k$ with the input block x_0^i , and a first control unit CTR_1 for optimizing the number and/or weights $w_{i,1}, w_{i,2}, w_{i,3}$ of the determined blocks $c_{i,1}, c_{i,2}, c_{i,3}$, wherein the Sum-of-Square-Distance between the downsampled locally optimized patch $z_{a,d}^k$ with the input block x_0^i is minimized and wherein a locally optimized patch x_i^1 is obtained.

25

The first and second downsampler DU_1, DU_2 and up-samplers UPS use a different factor (and its reciprocal) per each of the one or more cascaded upscaling units (e.g. $3/4$ for downsampling and $4/3$ for upsampling in the first upscaling unit CAU_1 , $4/5$ for downsampling and $5/4$ for upsampling in the second upscaling unit CAU_2 , and $5/6$ for downsampling and $6/5$ for upsampling in the third upscaling unit CAU_3).

30

The last of the cascaded upscaling units CAU₃ provides its output to the HR image assembling unit IAU. The HR image assembling unit IAU, in one embodiment, accumulates and averages the locally optimized patch x_i^3 of the last cascaded upscaling unit CAU₃ and provides its output to the IBP unit IU, which
5 provides the final HR output image.

In principle, an optimal number K of patches, i.e. a number K that is optimal for a current block to be upsampled, can be calculated in various ways. In one embodiment, an optimal number K of patches is obtained by downsampling the reconstructed (or upsampled) block, and computing an error between a current
10 block and the downsampled/reconstructed block. Different numbers K of patches are tested, and the number K is selected that leads to a minimization of the error. In one embodiment, the error is the sum of absolute differences. In another embodiment, the error is the square distance or square difference.

15 In the following, a separation of the blocks into highly textured and lowly textured blocks is described.

In one embodiment, the separation of the blocks into highly textured and lowly textured blocks is done according to the blocks' energy. In one embodiment, the
20 gradient energy is used. The gradient energy can be determined, for example, as a sum of squared spatial differences of luminance values within the block, or as a sum of horizontal and vertical gradients within the block. The higher the gradient energy, the higher the texturization of the block.

25 In one embodiment, the block's energy is determined as the difference between a minimum and a maximum luminance value within the block. Then, the separation of the blocks into highly textured and lowly textured blocks is done according to the difference between a minimum and a maximum luminance value within the block. The higher the difference, the higher the texturization of the block.

30

In one embodiment, a threshold is defined for determining whether a block is highly textured or lowly textured. If the block energy is below the threshold, a simplified upscaling is used. In one embodiment, a pre-defined fixed threshold is

used. A typical value (for block energy calculated according to min-max luminance difference) would be 5, i.e. if $\text{luminance}_{\max} - \text{luminance}_{\min} > 5$, the block is considered as highly textured. In one embodiment, an adaptive threshold is used. Since the threshold selection has an impact on both processing power and image quality, it may be adaptive with respect to available processing power (with a lower threshold for more available processing power), with respect to required image quality (with a lower threshold for higher image quality), or both. Also other factors may be considered in threshold selection, e.g. image size, bit amount per pixel etc. For example, the threshold may be in a range of 3-7 instead of 5. Note that the threshold needs to be scaled according to the employed energy measure, such that a reasonable amount of blocks is classified in each of high and low texture.

Fig.11 shows a flow-chart of a method, in one embodiment of the invention, where highly textured blocks are treated different from lowly textured blocks. In this embodiment, a method for performing super-resolution of a single image comprises steps of downsampling S_1 a LR input image by a first factor several times with different phases to obtain several different downsampled LR images $\Phi_0^1 - \Phi_8^1$, separating the LR input image into overlapping blocks x_i , and for each of the blocks determining S1.1 a measure of its texture or a degree of its texturization, and determining S1.2 according to the measure of its texture whether the block has high texture or low texture. Then, for each of the blocks having high texture, the following steps are performed:

searching one or more similar blocks $b_{i,1}, b_{i,2}, b_{i,3}$ of same size of a current block x_0^i in one or more or each of the downsampled LR images,

determining blocks $c_{i,1}, c_{i,2}, c_{i,3}$ at corresponding positions in the LR input image, wherein the blocks in the LR image are upsampled by the reciprocal of the first factor, locally optimizing S_4 the blocks by linear approximation to obtain locally optimized patches z_a^k ,

downsampling S_5 the locally optimized patches z_a^k by the first factor to obtain a downsampled locally optimized patch $z_{a,d}^k$,

comparing the downsampled locally optimized patch $z_{a,d}^k$ with the input block x_0^i and optimizing the number and/or weights $w_{i,1}, w_{i,2}, w_{i,3}$ of the determined blocks

$c_{i,1}, c_{i,2}, c_{i,3}$, wherein a Sum-of-Square-Distance between the downsampled locally optimized patch $z_{a,d}^k$ with the input block x_0^i is minimized and wherein a locally optimized patch x_i^1 is obtained. Further, for each of the blocks having low texture, simplified upsampling S1.6 is performed. The following steps are then repeated:

5 downsampling S'_1 several times with different phases to obtain several different downsampled LR images $\Phi_0^2 - \Phi_{15}^2$, searching, determining, locally optimizing S'_4 , and downsampling S'_5 , wherein for the downsampling a different second factor is used, and wherein for each repetition the locally optimized patch x_i^1, x_i^2, x_i^3 of the previous iteration is used as input block. Finally, the locally optimized patch x_i^3 of

10 the last iteration is accumulating and averaged S_8 , and IBP is performed S_9 , wherein a HR output image S_{10} is obtained.

In one embodiment of the method, the simplified upsampling is one of linear interpolation, bilinear interpolation, cubic interpolation and bi-cubic interpolation.

15

In one embodiment of the method, the simplified upsampling for blocks with low texture comprises steps of

searching one or more similar blocks $b_{i,1}, b_{i,2}, b_{i,3}$ of same size of a current block x_0^i in one or more or each of the downsampled LR images, wherein less blocks are

20 searched than for blocks with high texture, or the number of downsampled LR images in which the searching is done is less than for blocks with high texture; determining blocks $c_{i,1}, c_{i,2}, c_{i,3}$ at corresponding positions in the LR input image, wherein the blocks in the LR image are upsampled by the reciprocal of the first factor, locally optimizing S_4 the blocks by linear approximation to obtain locally

25 optimized patches z_a^k ,

downsampling S_5 the locally optimized patches z_a^k by the first factor to obtain a downsampled locally optimized patch $z_{a,d}^k$,

comparing the downsampled locally optimized patch $z_{a,d}^k$ with the input block x_0^i and optimizing the number and/or weights $w_{i,1}, w_{i,2}, w_{i,3}$ of the determined blocks

30 $c_{i,1}, c_{i,2}, c_{i,3}$, wherein the Sum-of-Square-Distance between the downsampled locally optimized patch $z_{a,d}^k$ with the input block x_0^i is minimized and wherein a locally optimized patch x_i^1 is obtained.

In one embodiment of the method, the step of determining S1.1 for a block a measure of its texture or a degree of its texturization comprises measuring the block's energy according to a maximum luminance difference within the block, or according to a sum of luminance gradients within the block.

5

In one embodiment of the method, the step of determining S1.2 for a block whether it has high texture or low texture comprises comparing an energy measure of the block with a threshold, and a block having an energy measure above the threshold is determined to have high texture and a block having an energy measure below the threshold is determined to have low texture.

10

Fig.12 shows an apparatus according to one embodiment of the invention, where highly textured blocks are treated in a different way than lowly textured blocks. Additionally to the above-described apparatus, the apparatus in this embodiment further comprises a block energy measure determining unit EM for determining a measure of the current block's texture or a degree of its texturization, and a comparison and decision unit ECD for comparing the measure of texture with a threshold so as to determine according to the measure of its texture whether the current block has high texture or low texture. As shown in Fig.12, the one or more first search units SU_1 , one or more first position determining units PDU_1 , first optimizer OPT_1 , second downsampler DU_2 , first comparator CU_1 and first control unit CTR_1 are operative for only those current blocks that have high texture $x_{0,HT}^i$, as determined to the comparison and decision unit ECD. The blocks having low texture $x_{0,LT}^i$ are upsampled in a simplified upsampling unit SUP by using any simplified upsampling method, as described above.

15

20

25

30

In one embodiment, the simplified upsampling unit SUP comprises one or more secondary first search units SU_1' for searching one or more similar blocks of same size $b_{i,1}, b_{i,2}, b_{i,3}$ of a current block x_0^i in one or more or each of the downsampled LR images, wherein less blocks are searched than for blocks with high texture, or the number of downsampled LR images in which the searching is done is less than for blocks with high texture, one or more secondary first position determining units PDU_1' for determining blocks $c_{i,1}, c_{i,2}, c_{i,3}$ at corresponding positions in the LR

input image, wherein the blocks in the LR image are upsampled by the reciprocal of the first factor, secondary first optimizer OPT_1' for locally optimizing S_4 the blocks by linear approximation to obtain locally optimized patches z_a^k , secondary second downsampler DU_2' for downsampling the locally optimized patches z_a^k by the first factor to obtain a downsampled locally optimized patch $z_{a,d}^k$, and secondary first comparator CU_1' for comparing the downsampled locally optimized patch $z_{a,d}^k$ with the input block x_0^i and optimizing the number and/or weights $w_{i,1}, w_{i,2}, w_{i,3}$ of the determined blocks $c_{i,1}, c_{i,2}, c_{i,3}$, wherein the Sum-of-Square-Distance between the downsampled locally optimized patch $z_{a,d}^k$ with the input block x_0^i is minimized and wherein a locally optimized patch x_i^1 is obtained.

In one embodiment, the block energy measure determining unit EM for determining for a block a measure of its texture or a degree of its texturization comprises means, e.g. a processor, for measuring or calculating the block's energy according to a maximum luminance difference within the block, or according to a sum of luminance gradients within the block.

In one embodiment, the comparison and decision unit ECD for determining for a block whether it has high texture or low texture comprises a comparison unit, e.g. comparator, for comparing an energy measure of the block with a threshold, and a determining unit, e.g. switch, for determining blocks having an energy measure above the threshold to have high texture and blocks having an energy measure below the threshold to have low texture.

Advantages of the invention can also be shown in experimental results. For example, in one embodiment the base enhancement factors are set to $S = 2$ and 3. To obtain larger ratios, the method is run multiple times with appropriate base enhancements. In an embodiment, the method is directly applied on concatenated RGB values of pixels. A good compromise between LR and HR patch sizes is found when $n = 3$, i.e., LR patches x_i of size 3×3 and HR patches y_i of size 6×6 for $\{S, p\} = \{2, 1\}$ and of size 9×9 for $\{S, p\} = \{3, 2\}$. Therefore, $\tau = 3$ with $\{s_1, s_2, s_3\}_{S=2} = \{3/4, 4/5, 5/6\}$ and $\{s_1, s_2, s_3\}_{S=3} = \{3/5, 5/7, 7/9\}$. Allowed are 3×2 or 2×3 overlaps between two adjacent LR patches, leading to 6×4 or 4×6 (when $S = 2$)

and 9×6 or 6×9 (when $S = 3$) overlaps between HR patches. K is set to 8. For IBP, $M = 5$ and the NLM kernel is of size 7×7 . The kernel coefficients are calculated with blocks of sizes 7×7 using an exponential parameter $\sigma = 0,01$. Fig. 5 demonstrates a sketch of the complete SR algorithm for $\{S, p, n\} = \{2, 1, 3\}$.

5 Experimental results of the algorithm, compared with different SR methods, are illustrated in Fig.7-10 (for the reported results of [9,10], images courtesy of their project web-pages). Fig.7 demonstrates a comparison between our result (right) and a result in [10] (left) for the “koala” image. This is a good example showing that the method according to the invention can provide more natural looking
10 synthesis results, especially for fine-detailed texture areas, e.g. in the koala’s fur. Figs.8-10 compare more results on different images for an upscaling factor of four. Particularly, Fig.8 (upper: [9], lower: present invention) shows the efficiency of using the NLM kernel for reducing the jaggy and ringing artifacts, e.g., in the ridge of the chip. From Fig.9 (left: [11], right: present invention) and Fig.10 (upper:
15 bi-cubic; lower: present invention), also sharper and cleaner edges can be observed, compared to the different other methods.

In one embodiment, the method for performing super-resolution of a single image comprises a step of determining for each of the overlapping blocks x_i whether a
20 current block has high or low texture, and simplifying or skipping the subsequent steps if a current block has low texture. Thus, super-resolution according to the above-described method is performed only for the highly textured blocks, while for the lowly-textured blocks a simplified method is used. This reduces the computational burden and makes the super-resolution processing faster. The
25 determining of blocks with high or low texture is done, in one embodiment, immediately after the step of separating the LR input image into overlapping blocks x_i . In another embodiment, the step of determining blocks with high or low texture is performed initially, before said step of downsampling S_1 a LR input image. In this case, the blocks need not necessarily coincide with the overlapping
30 blocks x_i , into which the image is separated later. For example, when highly or lowly textured regions are initially determined, there is a step of determining for each of the overlapping blocks x_i whether or not it belongs to a highly or lowly

textured region. As described above, each block has either high texture or low texture, so that determining one type of blocks implicitly yields also the other type.

In conclusion, a self-contained SR method based on optimized neighbor embedding of patch pairs is described. While making use of the recurrence property (i.e. self-similarity) of similar patches across different scales of an image, the given patch is gradually expanded, relying on geometric similarities of LR and HR patch spaces at small scaling ratios. Together with an adaptive IBP, the invention provides more natural looking textures and sharper edges with less artifacts when compared to other known methods. The invention is applicable e.g. for a complete GPU-oriented and totally localized implementation. As an advantage, it is faster or achieves higher performance respectively.

While there has been shown, described, and pointed out fundamental novel features of the present invention as applied to preferred embodiments thereof, it will be understood that various omissions and substitutions and changes in the apparatus and method described, in the form and details of the devices disclosed, and in their operation, may be made by those skilled in the art without departing from the spirit of the present invention. It is expressly intended that all combinations of those elements that perform substantially the same function in substantially the same way to achieve the same results are within the scope of the invention. Substitutions of elements from one described embodiment to another are also fully intended and contemplated. For example, the terms "block" and "patch" are used synonymously. The same holds for "sampling" and "scaling".

It will be understood that the present invention has been described purely by way of example, and modifications of detail can be made without departing from the scope of the invention. Each feature disclosed in the description and (where appropriate) the claims and drawings may be provided independently or in any appropriate combination. Features may, where appropriate be implemented in hardware, software, or a combination of the two. Connections may, where applicable, be implemented as wireless connections or wired, not necessarily direct or dedicated, connections. Reference numerals appearing in the claims are

by way of illustration only and shall have no limiting effect on the scope of the claims.

5 Cited References

- [1] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graphical Models Image Process.*, vol. 53, no. 3, pp. 231–239, 1991.
- [2] M. Irani and S. Peleg, "Motion analysis for image enhancement: Resolution, occlusion,
10 and transparency," *J. Vis. Comm. Image Repres.*, vol. 4, pp. 324–335, 1993.
- [3] S. Farsiu, M.D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1327–1344, 2004.
- [4] Z. Lin and H.-Y. Shum, "Fundamental limits of reconstruction- based superresolution algorithms under local translation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1,
15 pp. 83–97, 2004.
- [5] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, "Learning low-level vision," *Int. J. Comp. Vis.*, vol. 40, no. 1, pp. 25–47, 2000.
- [6] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Comp. Graph. Appl.*, vol. 22, no. 2, pp. 56–65, 2002.
- [7] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor
20 embedding," in *IEEE Comp. Soc. Conf. Comp. Vis. Pattern Recogn.*, 2004, pp. 275–282.
- [8] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super- resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, pp. 2861–2873, 2010.
- [9] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *IEEE
25 Int. Conf. Comp. Vis.*, 2009, pp. 349– 356.
- [10] G. Freedman and R. Fattal, "Image and video upscaling from local self-examples," *ACM Trans. Graph.*, vol. 28, no. 3, pp. 1–10, 2010.
- [11] J. Salvador, A. Kochale, and M. Borsum, "Robust super- resolution for interactive video navigation," in *IEEE Int. Conf. Consumer Electronics*, 2012, pp. 333–334.
- [12] L.I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 14, pp. 259–268, 1992.
- [13] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, 2000.
- [14] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in
35 *IEEE Comp. Soc. Conf. Comp. Vis. Pattern Recognition*, 2005, pp. 60–65.

Claims

1. A method for performing super-resolution of a single low-resolution input image, comprising steps of
 - 5 - downsampling (S_1) the low-resolution input image by a first factor two or more times with different phases to obtain several different downsampled low-resolution images ($\Phi_0^1 - \Phi_8^1$);
 - separating the low-resolution input image into overlapping blocks (x_i), and for each of the blocks (x_i) searching, in one or more or each of the down-
 10 sampled low-resolution images, one or more blocks ($b_{i,1}, b_{i,2}, b_{i,3}$) having same size as a current block (x_0^i) and being similar to the current block;
 - determining blocks ($c_{i,1}, c_{i,2}, c_{i,3}$) at corresponding positions in the low-resolution input image, wherein the blocks in the low-resolution image are upsampled by the reciprocal of the first factor;
 - 15 - locally optimizing (S_4) the blocks by linear approximation to obtain locally optimized patches (z_a^k);
 - downsampling (S_5) the locally optimized patches (z_a^k) by the first factor to obtain a downsampled locally optimized patch ($z_{a,d}^k$);
 - comparing the downsampled locally optimized patch ($z_{a,d}^k$) with the input
 20 block (x_0^i) and optimizing the number and/or weights ($w_{i,1}, w_{i,2}, w_{i,3}$) of the determined blocks ($c_{i,1}, c_{i,2}, c_{i,3}$), wherein the Sum-of-Square-Distance between the downsampled locally optimized patch ($z_{a,d}^k$) with the input block (x_0^i) is minimized and wherein a locally optimized patch (x_i^1) is obtained;
 - 25 - repeating the steps of downsampling (S'_1) two or more times with different phases to obtain several different downsampled low-resolution images ($\Phi_0^2 - \Phi_{15}^2$), searching, determining, locally optimizing (S'_4), downsampling (S'_5), wherein for the downsampling a different second factor is used, wherein for each repetition the locally optimized patch (x_i^1, x_i^2, x_i^3) of the
 30 previous iteration is used as input block;
 - accumulating and averaging (S_8) the locally optimized patch (x_i^3) of the last iteration; and

- performing iterative back projection (S_9), wherein a high-resolution output image (S_{10}) is obtained.
2. Method according to claim 1, wherein the first factor and the second factor are non-integers.
 3. Method according to claim 1 or 2, wherein the accumulating and averaging (S_8) comprises, for each pixel, counting overlapping patches that contribute to the pixel, adding values of overlapping patches that contribute to the pixel and dividing the accumulated values by the number of overlapping patches.
 4. Method according to one of the claims 1-3, wherein the iterative back projection (S_9) uses a Non-local means kernel.
 5. Method according to one of the claims 1-3, further comprising, for each of the blocks (x_i) resulting from said step of separating the low-resolution input image into overlapping blocks (x_i), steps of
 - determining ($S_{1.1}$) a measure of its texture or a degree of its texturization; and
 - determining ($S_{1.2}$) according to the measure of its texture whether the block has high texture or low texture;
 wherein said steps of searching one or more similar blocks of same size, determining blocks ($c_{i,1}, c_{i,2}, c_{i,3}$) at corresponding positions in the low-resolution input image, locally optimizing (S_4) the blocks by linear approximation, down-sampling (S_5) the locally optimized patches by the first factor and comparing the downsampled locally optimized patch ($z_{a,d}^k$) with the input block are performed only for each of the blocks having high texture, and wherein a simplified upsampling ($S_{1.6}$) is performed for each of the blocks having low texture.
 6. Method according to claim 5, wherein the simplified upsampling is one of linear interpolation, bilinear interpolation, cubic interpolation and bi-cubic interpolation.

7. Method according to claim 5, wherein the simplified upsampling for blocks with low texture comprises steps of
- searching one or more similar blocks of same size ($b_{i,1}, b_{i,2}, b_{i,3}$) of a current block (x_0^i) in one or more or each of the downsampled low-resolution images, wherein less blocks are searched than for blocks with high texture, or the number of downsampled low-resolution images in which the searching is done is less than for blocks with high texture;
 - determining blocks ($c_{i,1}, c_{i,2}, c_{i,3}$) at corresponding positions in the low-resolution input image, wherein the blocks in the low-resolution image are upsampled by the reciprocal of the first factor;
 - locally optimizing (S_4) the blocks by linear approximation to obtain locally optimized patches (z_a^k);
 - downsampling (S_5) the locally optimized patches (z_a^k) by the first factor to obtain a downsampled locally optimized patch ($z_{a,d}^k$);
 - comparing the downsampled locally optimized patch ($z_{a,d}^k$) with the input block (x_0^i) and optimizing the number and/or weights ($w_{i,1}, w_{i,2}, w_{i,3}$) of the determined blocks ($c_{i,1}, c_{i,2}, c_{i,3}$), wherein the Sum-of-Square-Distance between the downsampled locally optimized patch ($z_{a,d}^k$) with the input block (x_0^i) is minimized and wherein a locally optimized patch (x_i^1) is obtained.
8. Method according to one of the claims 5-7, wherein the step of determining (S1.1) for a block a measure of its texture or a degree of its texturization comprises measuring the block's energy according to a maximum luminance difference within the block, or according to a sum of luminance gradients within the block.
9. Method according to one of the claims 5-8, wherein the step of determining (S1.2) for a block whether it has high texture or low texture comprises comparing an energy measure of the block with a threshold, and a block having an energy measure above the threshold is determined to have high texture and a block having an energy measure below the threshold is determined to have low texture.

10. An apparatus for performing super-resolution of a single low-resolution input image, comprising a separator (SEP) for separating the low-resolution input image into overlapping blocks (x_i), one or more cascaded upscaling units (CAU1-3), a high-resolution image assembling unit (IAU), and an iterative back projection unit (IU),
- 5 wherein each cascaded upscaling unit provides its output (x_1^i, x_2^i, x_3^i) as input to the next upscaling unit,
- and wherein each cascaded upscaling unit (CAU) comprises
- one or more first downsamplers (DU_1) for downsampling a low-resolution input image by a first factor several times with different phases to obtain
 - 10 several different downsampled low-resolution images ($\Phi_0^1 - \Phi_8^1$);
 - first selector (SEL_1) for selecting a current block (x_0^i);
 - one or more first search units (SU_1) for searching, for each current block (x_0^i) of the blocks (x_i), one or more similar blocks of same size ($b_{i,1}, b_{i,2}, b_{i,3}$)
 - 15 of a current block (x_0^i) in one or more or each of the downsampled low-resolution images;
 - one or more first position determining units (PDU_1) for determining blocks ($c_{i,1}, c_{i,2}, c_{i,3}$) at corresponding positions in the low-resolution input image, wherein the blocks in the low-resolution image are upsampled in one or
 - 20 more upsamplers (UPS) by the reciprocal of the first factor;
 - first optimizer (OPT_1) for locally optimizing (S_4) the blocks by linear approximation to obtain locally optimized patches (z_a^k);
 - second downsampler (DU_2) for downsampling (S_5) the locally optimized patches (z_a^k) by the first factor to obtain a downsampled locally optimized patch ($z_{a,d}^k$);
 - 25 - first comparator (CU_1) for comparing the downsampled locally optimized patch ($z_{a,d}^k$) with the input block (x_0^i); and
 - first control unit (CTR_1) for optimizing the number and/or weights ($w_{i,1}, w_{i,2}, w_{i,3}$) of the determined blocks ($c_{i,1}, c_{i,2}, c_{i,3}$), wherein the Sum-of-Square-Distance between the downsampled locally optimized patch ($z_{a,d}^k$) with the
 - 30 input block (x_0^i) is minimized and wherein a locally optimized patch (x_i^1) is obtained;

wherein the first and second downsampler (DU_1, DU_2) and upsamplers (UPS) use a different factor per each of the one or more cascaded upscaling units, and

wherein the last of the cascaded upscaling units (CAU_3) provides its output to the high-resolution image assembling unit (IAU), the high-resolution image assembling unit (IAU) accumulating and averaging (S_9) the locally optimized patch (x_i^3) of the last cascaded upscaling unit (CAU_3) and providing its output to the iterative back projection unit (IU) which provides a high-resolution image.

10

11. Apparatus according to claim 10, wherein the simplified upsampling unit performs one of linear interpolation, bilinear interpolation, cubic interpolation and bi-cubic interpolation.

15

12. The apparatus according to claim 10 or 11, further comprising

- block energy measure determining unit (EM) for determining ($S1.1$) a measure of the current block's texture or a degree of its texturization;
- comparison and decision unit (ECD) for comparing the measure of texture with a threshold so as to determine ($S1.2$) according to the measure of its texture whether the current block has high texture or low texture; and
- simplified upsampling unit (SUP) for performing, for each of the blocks having low texture ($x_{0,LT}^i$), simplified upsampling ($S1.6$);

20

wherein said one or more first search units (SU_1), one or more first position determining units (PDU_1), first optimizer (OPT_1), second downsampler (DU_2), first comparator (CU_1) and first control unit (CTR_1) are operative for only those current blocks that have high texture ($x_{0,HT}^i$) according to said comparison and decision unit (ECD).

25

13. Apparatus according to claim 12, wherein the simplified upsampling unit (SUP) comprises

30

- one or more secondary first search units (SU_1') for searching one or more similar blocks of same size ($b_{i,1}, b_{i,2}, b_{i,3}$) of a current block (x_0^i) in one or more or each of the downsampled LR images, wherein less blocks are

searched than for blocks with high texture, or the number of downsampled low-resolution images in which the searching is done is less than for blocks with high texture;

- one or more secondary first position determining units (PDU_1') for
5 determining blocks ($c_{i,1}, c_{i,2}, c_{i,3}$) at corresponding positions in the LR input image, wherein the blocks in the low-resolution image are upsampled by the reciprocal of the first factor;
 - secondary first optimizer (OPT_1') for locally optimizing (S_4) the blocks by linear approximation to obtain locally optimized patches (z_a^k);
 - 10 - secondary second downsampler (DU_2') for downsampling (S_5) the locally optimized patches (z_a^k) by the first factor to obtain a downsampled locally optimized patch ($z_{a,d}^k$);
 - secondary first comparator (CU_1') for comparing the downsampled locally optimized patch ($z_{a,d}^k$) with the input block (x_0^i) and optimizing the number
15 and/or weights ($w_{i,1}, w_{i,2}, w_{i,3}$) of the determined blocks ($c_{i,1}, c_{i,2}, c_{i,3}$), wherein the Sum-of-Square-Distance between the downsampled locally optimized patch ($z_{a,d}^k$) with the input block (x_0^i) is minimized and wherein a locally optimized patch (x_i^1) is obtained.
- 20 14. Apparatus according to one of the claims 12-13, wherein the block energy measure determining unit (EM) for determining for a block a measure of its texture or a degree of its texturization comprises means for measuring or calculating the block's energy according to a maximum luminance difference within the block, or according to a sum of luminance gradients within the
25 block.
15. Apparatus according to one of the claims 12-14, wherein the comparison and decision unit (ECD) for determining for a block whether it has high texture or low texture comprises a comparison unit for comparing an energy measure of
30 the block with a threshold, and a determining unit for determining blocks having an energy measure above the threshold to have high texture and blocks having an energy measure below the threshold to have low texture.

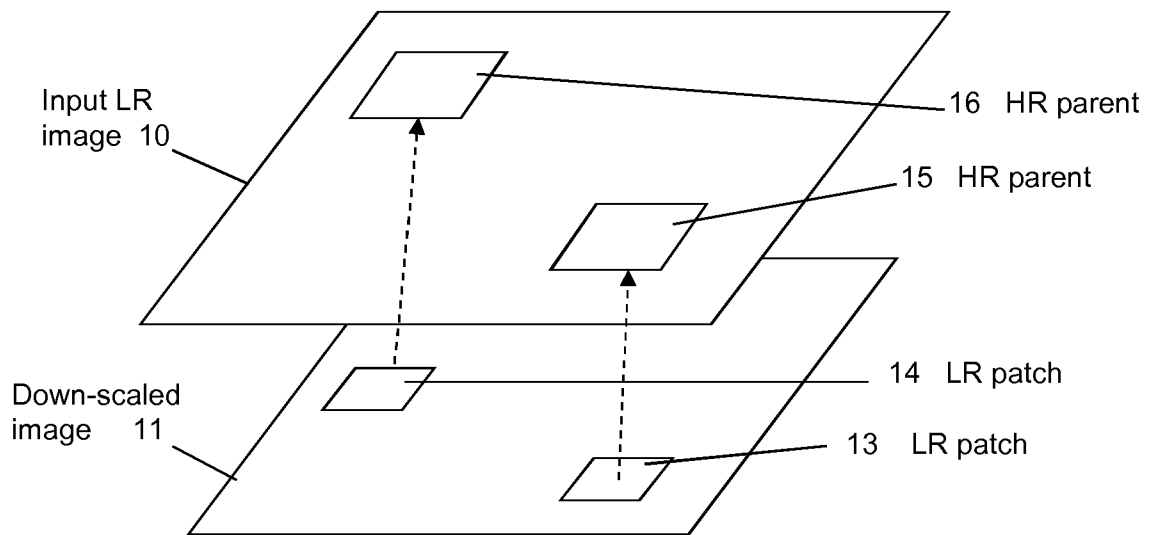
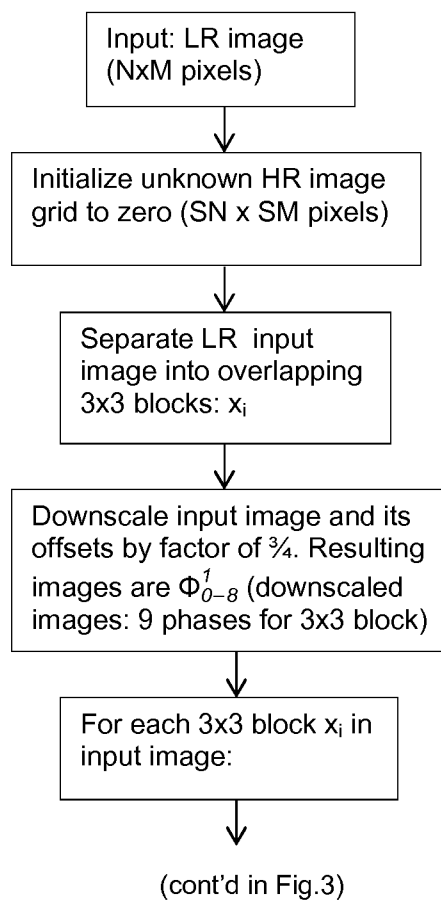


Fig.1



S1

Fig.2

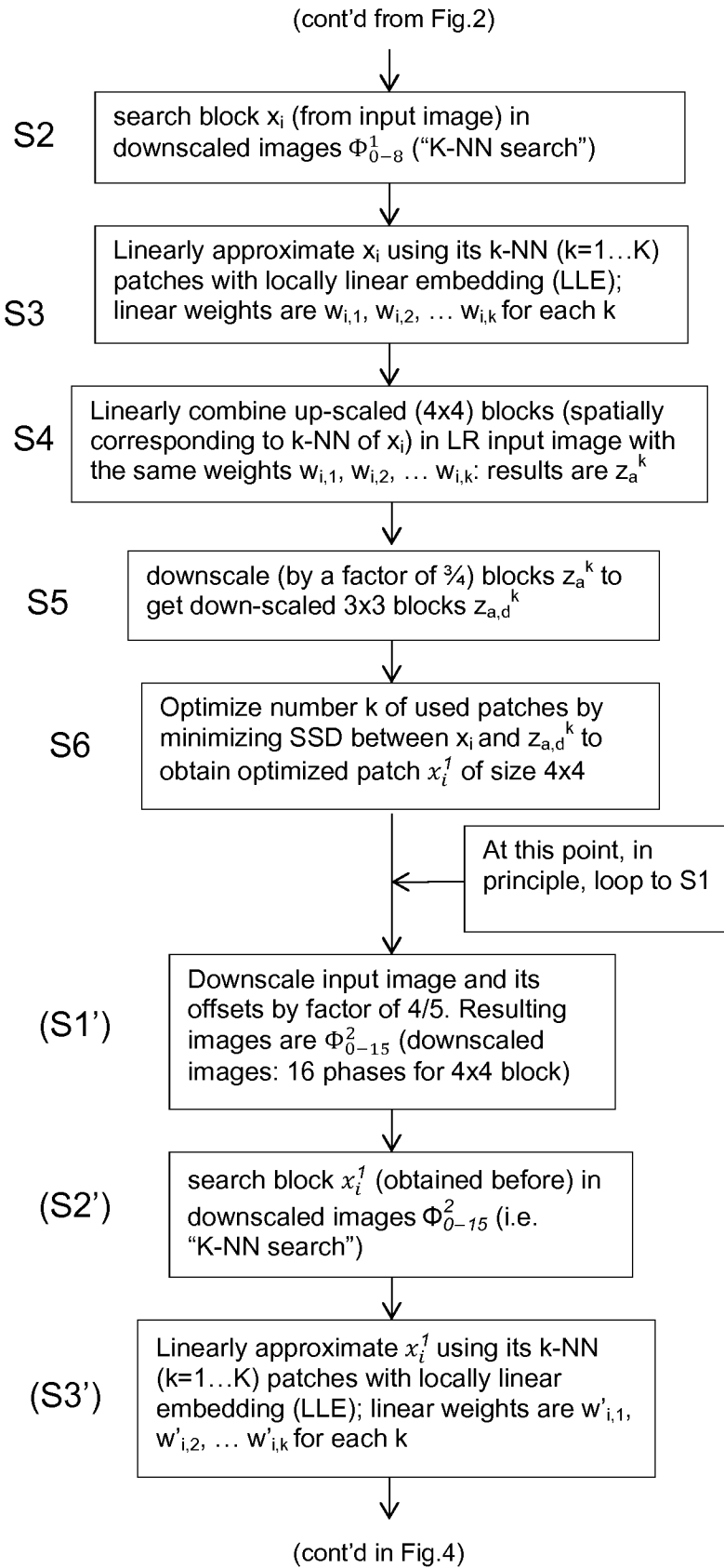


Fig.3

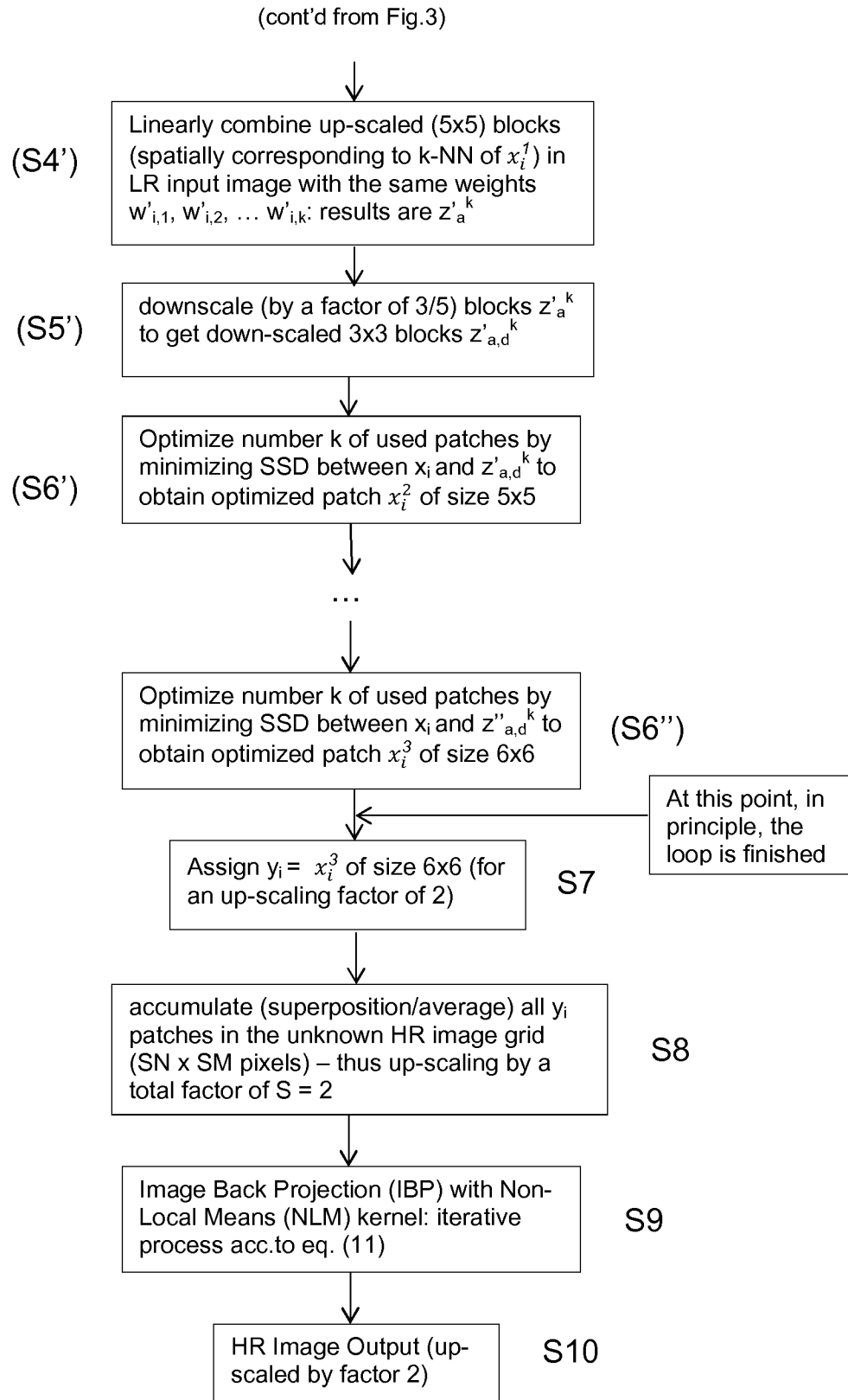


Fig.4

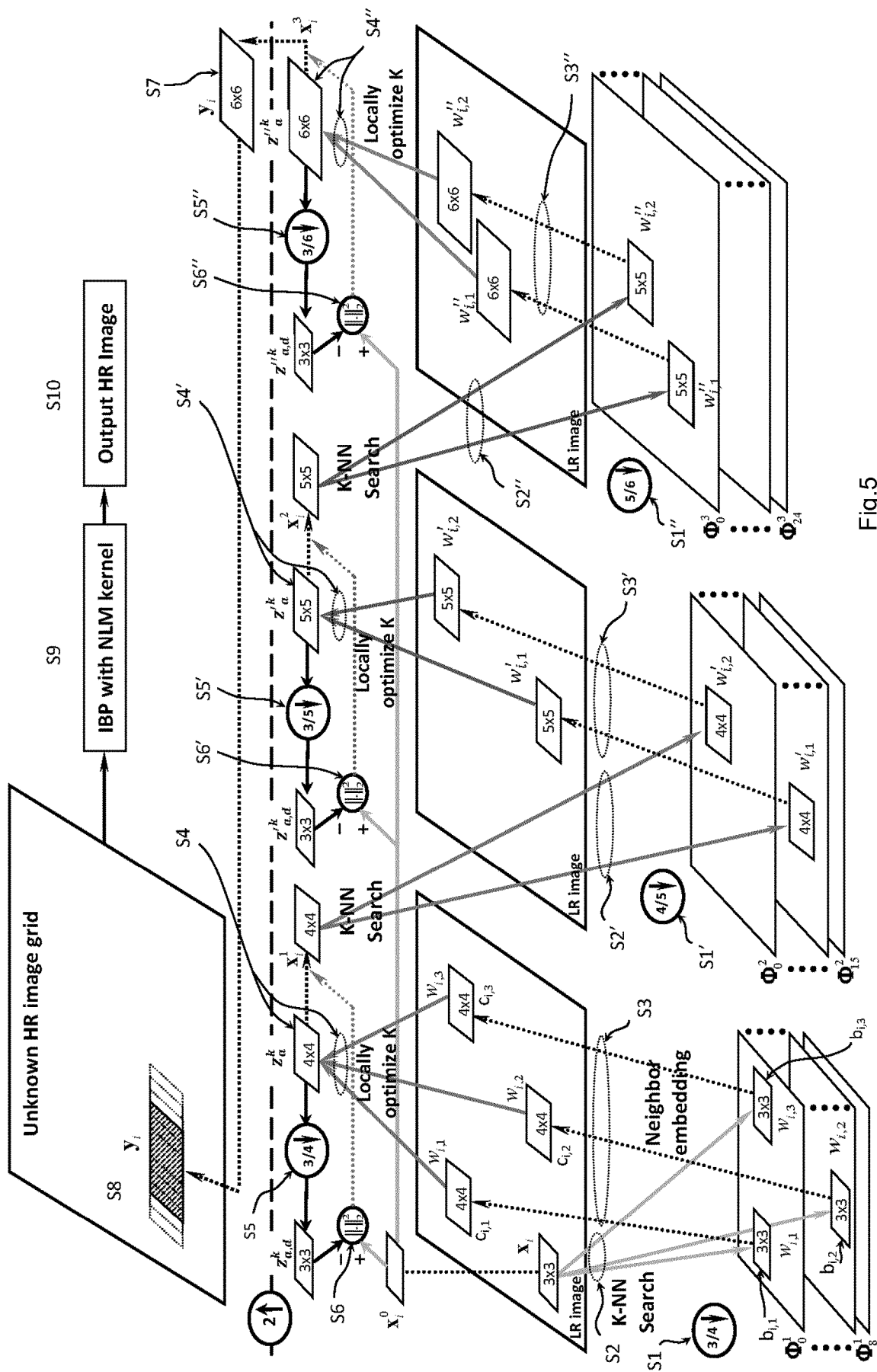


Fig.5

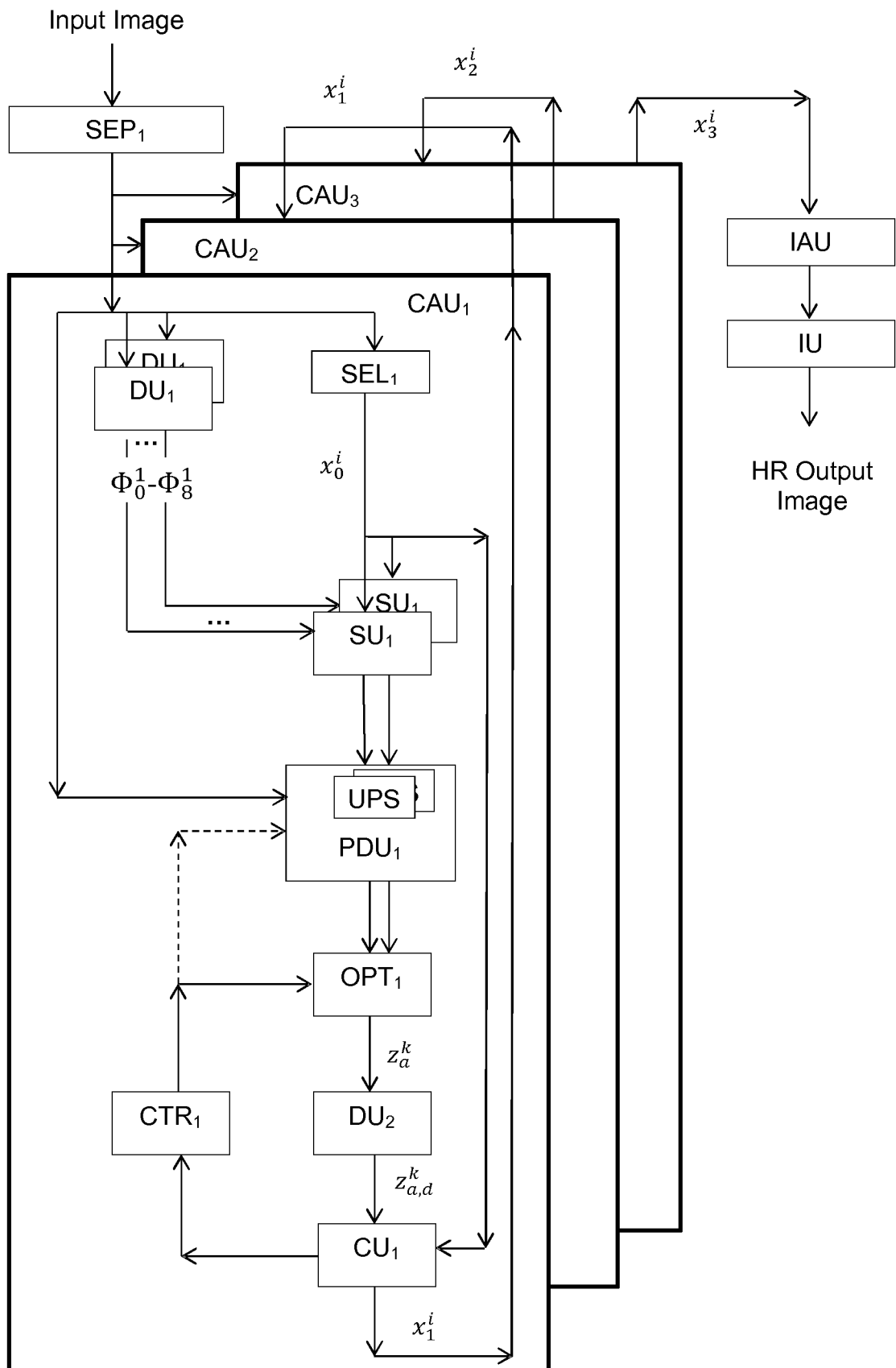


Fig.6

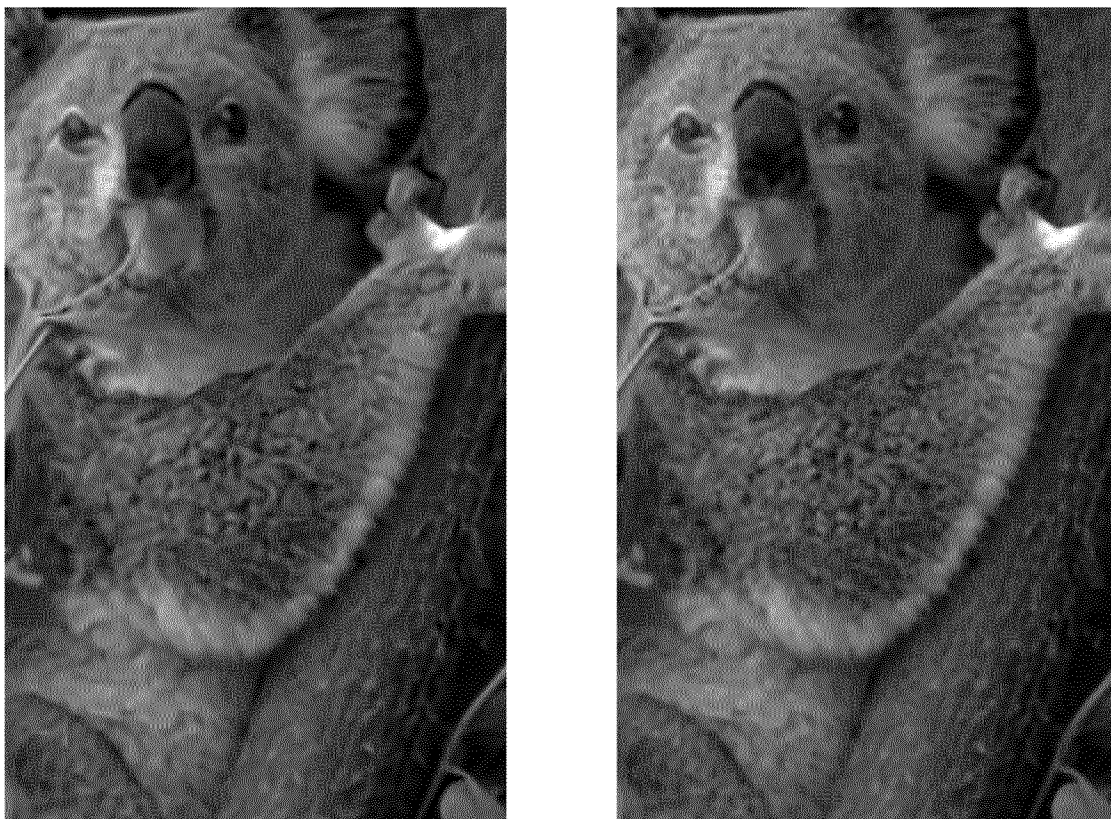


Fig.7

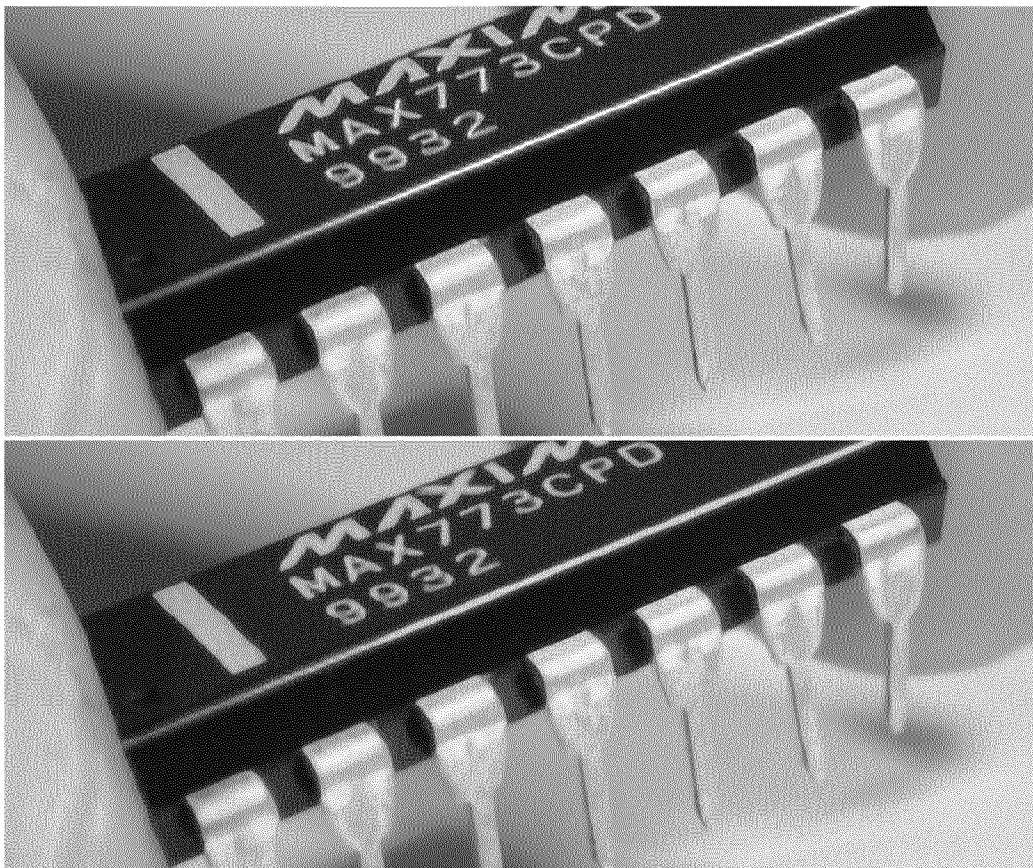


Fig.8



Fig.9

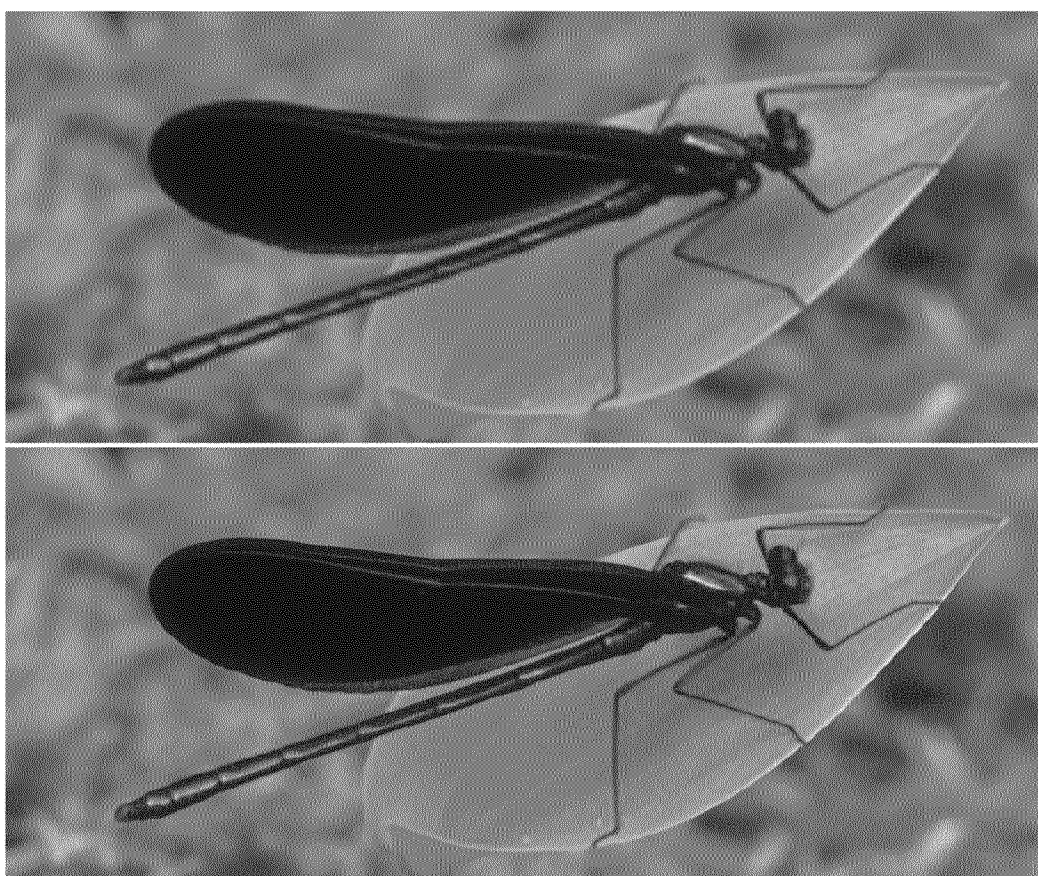


Fig.10

8/9

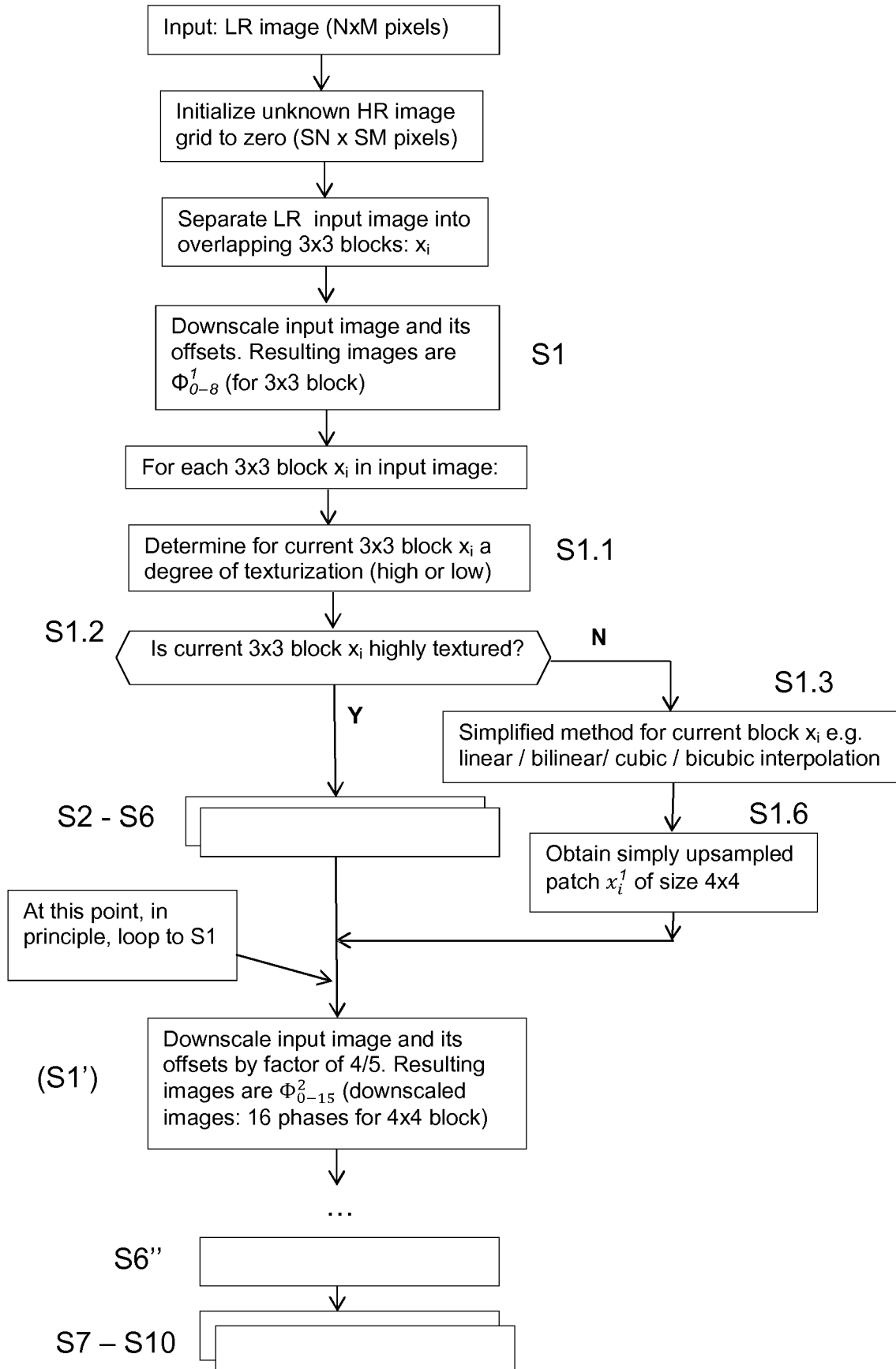


Fig.11

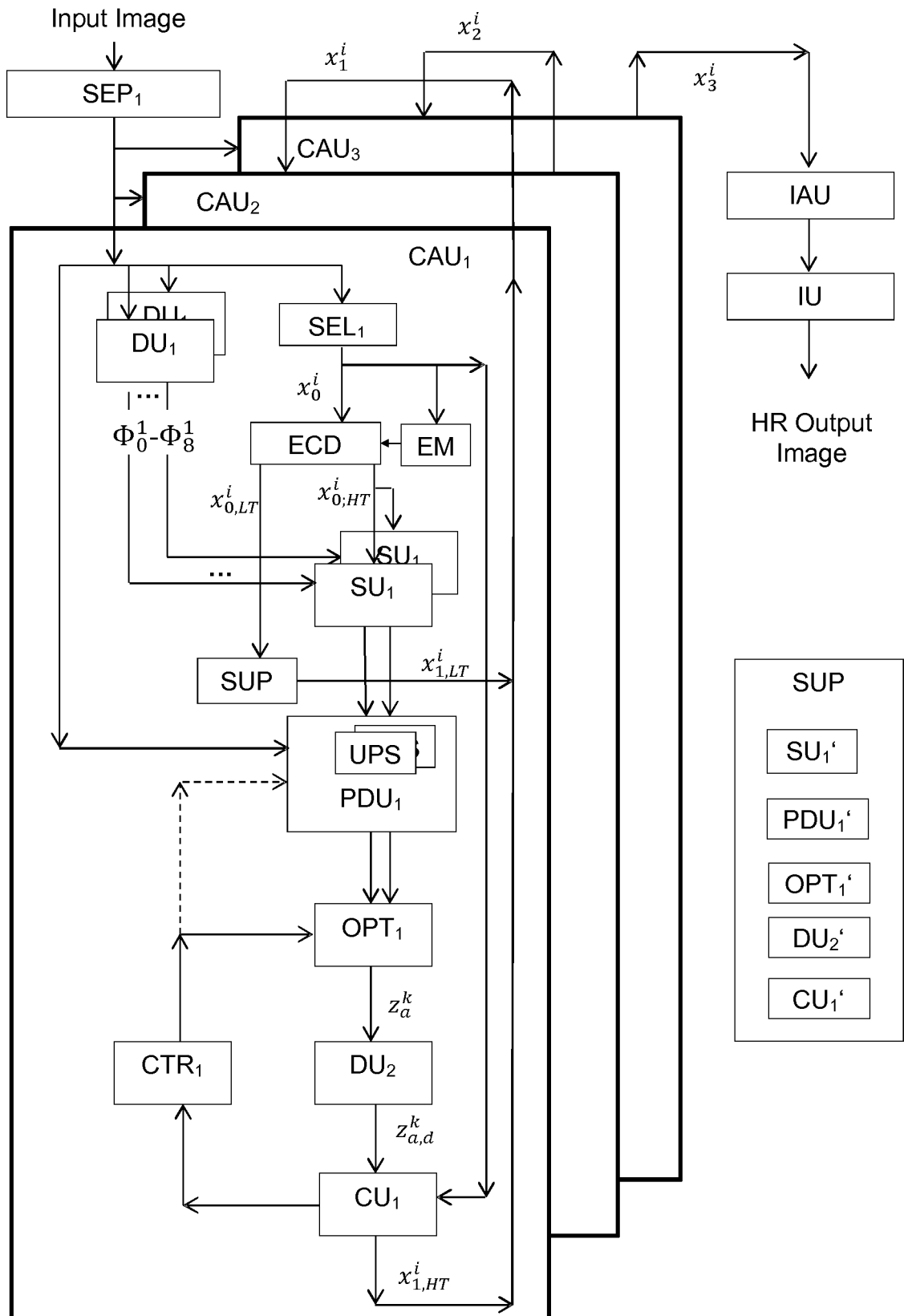


Fig.12

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2014/051117

A. CLASSIFICATION OF SUBJECT MATTER

INV. G06T3/40

ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06T

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, INSPEC, WPI Data, COMPENDEX, IBM-TDB

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	MEHMET TÜRKAN ET AL: "Self-content super-resolution for ultra-HD up-sampling", VISUAL MEDIA PRODUCTION, ACM, 2 PENN PLAZA, SUITE 701 NEW YORK NY 10121-0701 USA, 5 December 2012 (2012-12-05), pages 49-58, XP058010576, DOI: 10.1145/2414688.2414695 ISBN: 978-1-4503-1311-7 abstract page 49, paragraph 1 - page 53, paragraph 2.2 ----- -/-	1-15



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

14 May 2014

Date of mailing of the international search report

22/05/2014

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Zamuner, Umberto

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2014/051117

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	GILAD FREEDMAN ET AL: "Image and video upscaling from local self-examples", ACM TRANSACTIONS ON GRAPHICS (TOG), ACM, US, vol. 30, no. 2, 22 April 2011 (2011-04-22) , pages 1-11, XP058001113, ISSN: 0730-0301, DOI: 10.1145/1944846.1944852 cited in the application abstract page 3, paragraph 3 - page 7, paragraph 3.3 -----	1-15