(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 January 2016 (14.01.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/005242 A1

(51) International Patent Classification:
G06T 3/40 (2006.01)

(21) International Application Number:
PCT/EP2015/064974

(22) International Filing Date:
1 July 2015 (01.07.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
14306131.5 10 July 2014 (10.07.2014) EP

(71) Applicant: **THOMSON LICENSING** [FR/FR]; 1-5, rue Jeanne d'Arc, F-92130 Issy-les-Moulineaux (FR).

(72) Inventors: **GANDOLPH, Dirk**; Wallbrink 2, 30952 Ronnenberg (DE). **SALVADOR MARCOS, Jordi**; Voigtstrasse 5, 20257 Hamburg (DE). **PUTZKE-ROEMING, Wolfram**; Karl-Bodenstein-Weg 20, 31137 Hildesheim (DE). **KOCHALE, Axel**; Beckerweg 1, 31832 Springe (DE).

(74) Agent: **SCHMIDT-UHLIG, Thomas**; Deutsche Thomson OHG, European Patent Operations, 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, 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, ST, 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, 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))*

(54) Title: METHOD AND APPARATUS FOR UP-SCALING AN IMAGE

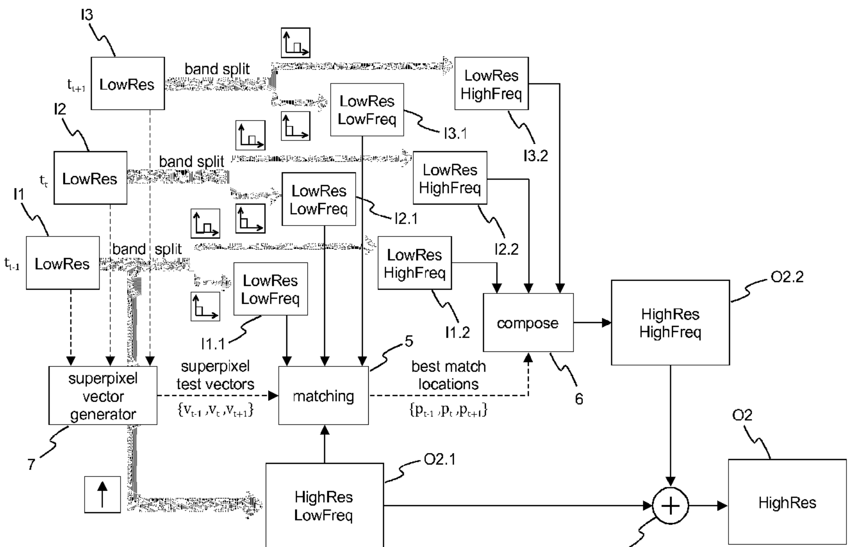

Fig. 3

(57) Abstract: A method and an apparatus (20) for up-scaling an input image (I2) are described, wherein a cross-scale self-similarity matching using superpixels is employed to obtain substitutes for missing details in an up-scaled image. The apparatus (20) comprises a superpixel vector generator (7) configured to generate (10) consistent superpixels for the input image (I2) and one or more auxiliary input images (I1, I3) and to generate (11) superpixel test vectors based on the consistent superpixels. A matching block (5) performs a cross-scale self- similarity matching (12) across the input image (I2) and the one or more auxiliary input images (I1, I3) using the superpixel test vectors. Finally, an output image generator (22) generates (13) an up-scaled output image (O2) using results of the cross-scale self-similarity matching (12).

WO 2016/005242 A1

METHOD AND APPARATUS FOR UP-SCALING AN IMAGE

FIELD

The present principles relate to a method and an apparatus for up-scaling an image. More specifically, a method and an apparatus for up-scaling an image are described, which make use of superpixels and auxiliary images for enhancing the up-scaling quality.

BACKGROUND

The technology of super-resolution is currently pushed by a plurality of applications. For example, the HDTV image format successors, such as UHDTV with its 2k and 4k variants, could benefit from super-resolution as the already existing video content has to be up-scaled to fit into the larger displays. Light field cameras taking multiple view images with relatively small resolutions each, do likewise require an intelligent up-scaling to provide picture qualities which can compete with state of the art system cameras and DSLR cameras (DSLR: Digital Single Lens Reflex). A third application is video compression, where a low resolution image or video stream can be decoded and enhanced by an additional super-resolution enhancement layer. This enhancement layer is additionally embedded within the compressed data and serves to supplement the prior via super-resolution up-scaled image or video.

The idea described herein is based on a technique exploiting image inherent self-similarities as proposed by G. Freedman et al. in: *"Image and video upscaling from local self-examples"*, ACM Transactions on Graphics, Vol. 30 (2011), pp. 12:1–12:11. While this fundamental paper was limited to still images, subsequent work incorporated multiple images to handle video up-scaling, as discussed within a paper by J. M. Salvador et al.: "Patch-based spatio-temporal super-resolution for video with non-rigid motion", Journal of Image Communication, Vol. 28 (2013), pp. 483-493.

Unfortunately, any method for up-scaling of images is accompanied by distressing quality losses.

Over the last decade superpixel algorithms have become a broadly accepted and applied method for image segmentation, providing a reduction in complexity for subsequent processing tasks. Superpixel segmentation provides the advantage of switching from a rigid structure of the pixel grid of an image to a semantic description defining objects in the image, which explains its popularity in image processing and computer vision algorithms.

Research on superpixel algorithms began with a processing intensive feature grouping method proposed by X. Ren et al. in: "*Learning a classification model for segmentation*", IEEE International Conference on Computer Vision (ICCV) 2003, pp. 10–17. Subsequently, more efficient solutions for superpixel generation were proposed, such as the simple linear iterative clustering (SLIC) method introduced by R. Achanta et al. in: "*SLIC superpixels compared to state-of-the-art superpixel methods*", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 34 (2012), pp. 2274–2282. While earlier solutions focused on still images, later developments aimed at application of superpixels to video, which require their temporal consistency. In M. Reso et al.: "*Temporally Consistent Superpixels*", International Conference on Computer Vision (ICCV), 2013, pp. 385-392, an approach achieving this demand is described, which provides traceable superpixels within video sequences.

SUMMARY

It is an object to describe an improved solution for up-scaling of an image, which allows achieving reduced quality losses.

According to one embodiment, a method for up-scaling an input image, wherein a cross-scale self-similarity matching using superpixels is employed to obtain substitutes for missing details in an up-scaled image, comprises:
- generating consistent superpixels for the input image and one or more auxiliary input images;
- generating superpixel test vectors based on the consistent superpixels;
- performing a cross-scale self-similarity matching across the input image and the one or more auxiliary input images using the superpixel test vectors; and
- generating an up-scaled output image using results of the cross-scale self-similarity matching.

Accordingly, a computer readable storage medium has stored therein instructions enabling up-scaling an input image, wherein a cross-scale self-similarity matching using superpixels is employed to obtain substitutes for missing details in an up-scaled image. The instructions, when executed by a computer, cause the computer to:
- generate consistent superpixels for the input image and one or more auxiliary input images;
- generate superpixel test vectors based on the consistent superpixels;
- perform a cross-scale self-similarity matching across the input image and the one or more auxiliary input images using the superpixel test vectors; and

- generate an up-scaled output image using results of the cross-scale self-similarity matching.

Also, in one embodiment an apparatus configured to up-scale an input image, wherein a cross-scale self-similarity matching using superpixels is employed to obtain substitutes for missing details in an up-scaled image, comprises:
- a superpixel vector generator configured to generate consistent superpixels for the input image and one or more auxiliary input images and to generate superpixel test vectors based on the consistent superpixels;
- a matching block configured to perform a cross-scale self-similarity matching across the input image and the one or more auxiliary input images using the superpixel test vectors; and
- an output image generator configured to generate an up-scaled output image using results of the cross-scale self-similarity matching.

In another embodiment, an apparatus configured to up-scale an input image, wherein a cross-scale self-similarity matching using superpixels is employed to obtain substitutes for missing details in an up-scaled image, comprises a processing device and a memory device having stored therein instructions, which, when executed by the processing device, cause the apparatus to:
- generate consistent superpixels for the input image and one or more auxiliary input images;
- generate superpixel test vectors based on the consistent superpixels;
- perform a cross-scale self-similarity matching across the input image and the one or more auxiliary input images using the superpixel test vectors; and
- generate an up-scaled output image using results of the cross-scale self-similarity matching.

The proposed super-resolution method tracks captured objects by analyzing generated temporal or multi-view consistent superpixels. The awareness about objects in the image material and of their whereabouts in time or in different views is transferred into advanced search strategies for finding relevant multi-image cross-scale self-similarities. By incorporating the plurality of significant self-similarities found for different temporal phases or different views a better suited super-resolution enhancement signal is generated, resulting in an improved picture quality. The proposed super-resolution approach provides an improved image quality, which can be measured in peak signal-to-noise ratio via the comparison against ground truth data. In addition, subjective testing confirms the visual improvements for the resulting picture quality, which is useful, as peak signal-to-noise ratio measures are not necessarily consistent with human visual perception.

The super-resolution approach works on multiple images, which might represent an image sequence in time (e.g. a video), a multi-view shot (e.g. Light Field camera image holding multiple angles), or even a temporal sequence of multi-view shots. These applications are interchangeable, which means that multi-view images and temporal images can be treated as equivalents.

In one embodiment, the solution comprises:

- up-sampling the input image to obtain a high resolution, low frequency image;
- determining match locations between the input image and the high resolution, low frequency image, and between the one or more auxiliary input images and the high resolution, low frequency image;

- composing a high resolution, high frequency composed image from the input image and the one or more auxiliary input images using the match locations; and
- combining the high resolution, low frequency image and the high resolution, high frequency composed image into a high resolution up-scaled output image.

Typically, the up-sampled image has distressing quality losses due to the missing details. However, these missing details are substituted using image blocks from the input image and the one or more auxiliary input images. While these images will only contain a limited number of suitable image blocks, these blocks are generally more relevant, i.e. fitting better.

In one embodiment, the input images are band split into low resolution, low frequency images and low resolution, high frequency images, wherein the low resolution, low frequency images are used for the cross-scale self-similarity matching and the low resolution, high frequency images are used for generating the up-scaled output image. In this way an efficient analysis of self-similarity is ensured and the necessary high-frequency details for the up-scaled output image can be reliably obtained.

In one embodiment, an image block for generating the up-scaled output image is generated by performing at least one of selecting a single image block defined by a best match of the cross-scale self-similarity matching, generating a linear combination of all or a subset of blocks defined by matches of the cross-scale self-similarity matching, and generating an average across all image blocks defined by matches of the cross-scale self-similarity matching. While the former two solutions require less processing power, the latter solution shows the best results for the peak signal-to-noise ratio.

For a better understanding the solution shall now be explained in more detail in the following description with reference to the figures. It is understood that the solution is not limited to this exemplary embodiment and that specified features can also expediently be combined and/or modified without departing from the scope of the present solution as defined in the appended claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 shows a block-diagram of a known super-resolution algorithm;

Fig. 2 shows an extended and more compact version of the block diagram of Fig. 1;

Fig. 3 depicts a super-resolution multi-image self-similarity matching using superpixels;

Fig. 4 illustrates a linear combination of image blocks, where combination weights are determined via linear regression;

Fig. 5 shows an example of an image before segmentation into superpixels;

Fig. 6 shows the image of Fig. 5 after segmentation into superpixels;

Fig. 7 shows an example of a single temporally consistent superpixel being tracked over a period of three images;

Fig. 8 shows average peak signal-to-noise ratios obtained for different up-scaling algorithms;

Fig. 9 shows average structural similarity values obtained for different up-scaling algorithms;

Fig. 10 depicts a method according to an embodiment for up-scaling an image;

Fig. 11 schematically depicts a first embodiment of an apparatus configured to perform a method for up-scaling an image; and

Fig. 12 schematically illustrates a second embodiment of an apparatus configured to perform a method for up-scaling an image.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

In the following the solution is explained with a focus on temporal image sequences, e.g. images of a video sequence. However, the described approach is likewise applicable to spatially related images, e.g. multi-view images.

The approach described in the following is based on the super-resolution algorithm by G. Freedman et al., as shown by the block-diagram in Fig. 1. Of course, the general idea is likewise applicable to other super-resolution algorithms. For simplicity the block diagram describes a solution working for single images only, while the proposed approach provides a solution for multiple images. All corresponding necessary extensions are explained later in a separate block diagram.

In Fig. 1 a low resolution input image I1 is processed by three different filters: an up-sampling filter 1 generating a low frequency, high resolution image O1.1, a low-pass filter 2 generating a low frequency, low resolution image I1.1, and a high-pass filter 3 generating a high frequency, low resolution image I1.2.

Usually the up-sampled image O1.1 has distressing quality losses due to the missing details caused by a bi-cubic or alternatively a more complex up-sampling. In the following steps a substitute for these missing details is generated by exploiting the inherent cross-scale self-similarity of natural objects. The process of generating the missing details results in a high frequency, high resolution image O1.2, which can be combined with the low frequency, high resolution image O1.1 in a processing block 4 to generate the final high-resolution output image I2.

The cross-scale self-similarities are detected by a matching process block 5. This matching process block 5 searches the appropriate matches within the low resolution image I1.1 for all pixels in the high resolution image O1.1. State of the art for the matching process is to search within fixed extensions of a rectangular search window. The matching process block 5 generates best match locations for all pixels in O1.1 pointing to I1.1. These best match locations are transferred to a composition block 6, which copies the indicated blocks from the high frequency, low resolution image I1.2 into the high frequency, high resolution image O1.2.

The block diagram in Fig. 2 shows a more compact version of the block diagram of Fig. 1, which is extended by an advanced matching technique. The additional block in Fig. 2 is a superpixel vector generator 7, which processes the input image

I1 for calculating superpixels and selects test vectors used for the matching block 5. The superpixel test vector generation substitutes the rigid rectangular search window used in Fig. 1.

The block diagram in Fig. 3 explains a further extension of the superpixel vector generation, namely a super-resolution multi-image self-similarity matching using superpixels. As its predecessor in Fig. 2 the block diagram of Fig. 3 is aware of the objects in the image material. The idea is that the objects are tracked over multiple images, which serve to generate test vectors for the matching across multiple input images in the vector generator block 7. In Fig. 3 the number of input images is three, but this number is not mandatory and can be increased or reduced by including or excluding images located in future or past direction. Similarly, a multi-view application can include or exclude further views/angles, or a temporal sequence of multi-view images can include or exclude further views/angles and/or temporally succeeding or preceding images.

The example given in Fig. 3 shows the proposed method executed for image I2 at time t_t for creating the output image O2 also at the time t_t. The input images I1 and I3 at the times t_{t-1} and t_{t+1} are additional sources to find relevant cross-scale self-similarities for the output image O2.

The matching block 5 receives the superpixel test vectors for all input images, which in this example are $\{v_{t-1}, v_t, v_{t+1}\}$, and generates best match locations for all pixels in O2.1 pointing to I1.1, I2.1, and I3.1, respectively. In the figure this is indicated by $\{p_{t-1}, p_t, p_{t+1}\}$ representing three complete sets of best match locations. Usually the dimension of a set equals the number of input images. The composition block 6 combines the indicated blocks from I1.2, I2.2, and I3.2 and copies the combination result into the high frequency, high resolution image O2.2.

In the following a more detailed description of the vector generator block 7 and the composition block 6 is given.

The multi-image superpixel vector generator block 7 generates the superpixel test vector set $\{v_{t-1}, v_t, v_{t+1}\}$ by performing the following steps:

STEP 1: Generating consistent superpixels $\{SP_{t-1}(m), SP_t(n), SP_{t+1}(r)\}$, where the indices $\{m, n, r\}$ run over all superpixels in the images. The term temporally consistent can be substituted with multi-view consistent for multi-view applications. An approach for generating temporally consistent superpixels is described in M. Reso et al.: "*Temporally Consistent Superpixels*", International Conference on Computer Vision (ICCV), 2013, pp. 385-392. Fig. 5 shows an example of an image being segmented into superpixel areas as depicted in Fig. 6, where each superpixel is represented using a different grey value. Fig. 6 is called a superpixel label map. Fig. 7 shows an example of a single temporally consistent superpixel being tracked over the period of three images, where the superpixels follow a moving object in the video scene depicted in the images at the times t_{t-1}, t_t, and t_{t+1}.

STEP 2: Generating search vectors $\{s_{t-1}(\zeta), s_t(\zeta), s_{t+1}(\zeta)\}$ separately for all superpixel images, where the index ζ runs across all image positions. One approach for generating such search vectors is described, for example, in co-pending European Patent Application EP14306130.

STEP 3: Generating object related pixel assignments for all superpixels

$$\begin{array}{ccc} SP_t & \rightarrow & SP_{t+1} \\ SP_t & \rightarrow & SP_{t+2} \\ \ldots & \rightarrow & \ldots \end{array} \quad \text{and} \quad \begin{array}{ccc} SP_t & \rightarrow & SP_{t-1} \\ SP_t & \rightarrow & SP_{t-2} \\ \ldots & \rightarrow & \ldots \end{array},$$

where the number of relations depends on the number of input images. One approach for generating such object related pixel assignments is described, for example, in co-pending European Patent Application EP14306126. In the example in Fig. 3 only the very first lines are used.

STEP 4: The final superpixel test vectors $\{v_{t-1}, v_t, v_{t+1}\}$ are determined by applying the pixel assignments found in STEP 3. For the example in Fig. 3 each separate superpixel $SP_t(n) \equiv SP_{t,n}$ in the image at the time t_t has a pixel individual assignment to $SP_{t-1}(m) \equiv SP_{t-1,m}$ and a pixel individual assignment to $SP_{t+1}(r) \equiv SP_{t+1,r}$, which can be expressed by $p_{t,n}(i) \rightarrow p_{t-1,m}(j)$ and $p_{t,n}(i) \rightarrow p_{t+1,r}(k)$, with $i \in \{1, \ldots I\}$, $j \in \{1, \ldots J\}$, and $k \in \{1, \ldots K\}$. In other words, for each pixel $p_{t,n}(i)$ located in an origin superpixel $SP_{t,n}$ in the image at the time t_t corresponding pixels $p_{t-1,m}(j)$ and $p_{t+1,r}(k)$ are required, being located within the superpixels $SP_{t-1,m}$ in the image at the time t_{t-1} and $SP_{t+1,r}$ in the image at the time t_{t+1}. I is the number of pixels contained in $SP_{t,n}$, J the number of pixels contained in $SP_{t-1,m}$, and K the number of pixels contained in $SP_{t+1,r}$. In general the numbers of pixels I, J, and K are different. Therefore, the resulting pixel mappings can be one-to-many, one-to-one, many-to-one, and a combination of them. The test vectors v_t need no assignments, as they can be taken directly, i.e. $v_t(\zeta) = s_t(\zeta)$. The test vectors v_{t-1} and v_{t+1} use the assignments according to $v_{t-1}(\zeta) = s_{t-1}\left(p_{t,n}(\zeta) \rightarrow p_{t-1,m}(\zeta)\right)$ and $v_{t+1}(\zeta) = s_{t+1}\left(p_{t,n}(\zeta) \rightarrow p_{t+1,r}(\zeta)\right)$, respectively. A larger number of input images is treated accordingly.

The block combination performed by the composition block 6 can be implemented, for example, using one of the following approaches:

a) Selection of a single block only defined by the very best match, i.e. the best among all best matches found.

b) A linear combination of all or a subset of the blocks, where the weights (linear factors) are determined via linear regression, as shown in Fig. 4.

c) Generating the average across all best matches found. This approach is preferable, as it shows the best results for the PSNR (Peak Signal-to-Noise Ratio).

Fig. 4 shows the linear regression approach for composing the high frequency, high resolution image O2.2 executed within the composition block 6. The linear regression is processed for each pixel position ζ in O2.1 individually by taking the best match locations $\{p_{t-1}, p_t, p_{t+1}\}$, fetching the best match block data $\{\vec{d}_{t-1}(p_{t-1}), \vec{d}_t(p_t), \vec{d}_{t+1}(p_{t+1})\}$ and the target block $\vec{b}$ by forming the regression equation

$$\vec{\alpha} = \begin{pmatrix} d_{t,1} & d_{t-1,1} & d_{t+1,1} & \dots \\ d_{t,2} & d_{t-1,2} & d_{t+1,2} & \dots \\ \dots & \dots & \dots & \dots \\ d_{t,q} & d_{t-1,q} & d_{t+1,q} & \dots \end{pmatrix}^{-1} \cdot \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_q \end{pmatrix} \quad \text{or} \quad \vec{\alpha} = (D)^{-1} \cdot \vec{b},$$

where q is the number of pixels in the matching block. This equation is solvable if the count of input images is less or equal to the number of pixels in the matching block. In case that the count of input images is higher it is proposed to reduce the horizontal dimension of matrix D by selecting the best matching blocks only, i.e. those blocks with the minimum distance measures.

The two diagrams in Figs. 8 and 9 show the average PSNR and SSIM (Structural SIMilarity) analyzed over a sequence of 64 images by comparing the up-scaled images against ground truth data. Shown are the comparisons between the following algorithms:

bicubic: Up-scaling via bi-cubic interpolation.

SISR: Single Image Super Resolution, the matching process searches within fixed extensions of a rectangular search window.

SRm25: Single image Super Resolution using a vector based self-similarity matching. The search vector length is 25.

SRuSPt1: Multi-image self-similarity matching using superpixels across three images $\{t_{t-1}, t_t, t_{t+1}\}$, i.e. one previous and one future image, by averaging as described above in item c).

SRuSPt5: Multi-image self-similarity matching using superpixels across eleven images $\{t_{t-5}, \ldots, t_{t-1}, t_t, t_{t+1}, \ldots, t_{t+5}\}$, i.e. five previous and five future images, by averaging as described above in item c).

SRuSPt1s: Multi-image self-similarity matching using superpixels across three images $\{t_{t-1}, t_t, t_{t+1}\}$, i.e. one previous and one future image, but selecting the best matching block as described above in item a).

SRuSPt5s: Multi-image self-similarity matching using superpixels across eleven images $\{t_{t-5}, \ldots, t_{t-1}, t_t, t_{t+1}, \ldots, t_{t+5}\}$, i.e.

five previous and five future images, but selecting the best matching block as described above in item a).

The two diagrams show that all methods using superpixel controlled self-similarity matching are superior to the matching within a fixed search area. They also reveal that an increase of input images creates an improvement for the PSNR and SSIM values. Finally, it can be seen that the SRuSPt5 algorithm analyzing eleven input images creates superior PSNR and SSIM values.

Fig. 10 schematically illustrates one embodiment of a method for up-scaling an image, wherein a cross-scale self-similarity matching using superpixels is employed to obtain substitutes for missing details in an up-scaled image. In a first step consistent superpixels are generated 10 for the input image I2 and one or more auxiliary input images I1, I3.
Based on these consistent superpixels superpixel test vectors are then generated 11. Using the superpixel test vectors a cross-scale self-similarity matching 12 is performed across the input image I2 and the one or more auxiliary input images I1, I3. Finally, an up-scaled output image O2 is generated 13 using results of the cross-scale self-similarity matching 12.

Fig. 11 depicts one embodiment of an apparatus 20 for up-scaling an input image I2. The apparatus 20 employs a cross-scale self-similarity matching using superpixels to obtain substitutes for missing details in an up-scaled image. To this end the apparatus 20 comprises an input 21 for receiving an input image I2 to be up-scaled and one or more auxiliary input images I1, I3. A superpixel vector generator 7 generates 10 consistent superpixels for the input image I2 and one or more auxiliary input images I1, I3, and further generates 11 superpixel test vectors based on the consistent superpixels. Of course, these two functions may likewise be performed by separate processing blocks. A matching block 5 performs a cross-scale self-similarity matching 12 across the input image I2 and the one or more auxiliary input images I1, I3 using the superpixel test vectors. An output image generator 22 generates 13 an up-scaled output image O2 using results of the cross-scale self-similarity matching 12. In one embodiment, the output image generator 22 comprises the composition block 6 and a processing block 4 as described further above. The resulting output image O2 is made available at an output 23 and/or stored on a local storage. The superpixel vector generator 7, the matching block 5, and the output image generator 22 are either implemented as dedicated hardware or as software running on a processor. They may also be partially or fully combined in a single unit. Also, the input 21 and the output 23 may be combined into a single bi-directional interface.

Another embodiment of an apparatus 30 configured to perform the method for up-scaling an image is schematically illustrated in Fig. 12. The apparatus 30 comprises a processing device 31 and a memory device 32 storing instructions that, when executed, cause the apparatus to perform steps according to one of the described methods.

For example, the processing device 31 can be a processor adapted to perform the steps according to one of the described methods. In an embodiment said adaptation comprises that the processor is configured, e.g. programmed, to perform steps according to one of the described methods.

CLAIMS

1. A method for up-scaling an input image (I2), wherein a cross-scale self-similarity matching using superpixels is employed to obtain substitutes for missing details in an up-scaled image, **characterized** in that the method comprises:
- generating (10) consistent superpixels for the input image (I2) and one or more auxiliary input images (I1, I3);
- generating (11) superpixel test vectors based on the consistent superpixels;
- performing a cross-scale self-similarity matching (12) across the input image (I2) and the one or more auxiliary input images (I1, I3) using the superpixel test vectors; and
- generating (13) an up-scaled output image (O2) using results of the cross-scale self-similarity matching (12).

2. The method according to claim 1, the method **comprising**:
- up-sampling the input image (I2) to obtain a high resolution, low frequency image (O2.1);
- determining (12) match locations between the input image (I2) and the high resolution, low frequency image (O2.1), and between the one or more auxiliary input images (I1, I3) and the high resolution, low frequency image (O2.1);
- composing a high resolution, high frequency composed image (O2.2) from the input image (I2) and the one or more auxiliary input images (I1, I3) using the match locations; and
- combining the high resolution, low frequency image (O2.1) and the high resolution, high frequency composed image (O2.2) into a high resolution up-scaled output image (O2).

3. The method according to claim 1 or 2, **wherein** the input image (I2) and the one or more auxiliary input images (I1,

I3) are successive images of a sequence of images or multi-view images of a scene.

4. The method according to one of the preceding claims, **wherein** the input images (I1, I2, I3) are band split into low resolution, low frequency images (I1.1, I2.1, I3.1) and low resolution, high frequency images (I1.2, I2.2, I3.2), wherein the low resolution, low frequency images (I1.1, I2.1, I3.1) are used for the cross-scale self-similarity matching (12) and the low resolution, high frequency images (I1.2, I2.2, I3.2) are used for generating (13) the up-scaled output image (O2).

5. The method according to one of the preceding claims, **wherein** an image block for generating (13) the up-scaled output image (O2) is generated by performing at least one of selecting a single image block defined by a best match of the cross-scale self-similarity matching (12), generating a linear combination of all or a subset of blocks defined by matches of the cross-scale self-similarity matching (12), and generating an average across all image blocks defined by matches of the cross-scale self-similarity matching (12).

6. A computer readable storage medium having stored therein instructions enabling up-scaling an input image (I2), wherein a cross-scale self-similarity matching using superpixels is employed to obtain substitutes for missing details in an up-scaled image, wherein the instructions, when executed by a computer, cause the computer to:
- generate (10) consistent superpixels for the input image (I2) and one or more auxiliary input images (I1, I3);
- generate (11) superpixel test vectors based on the consistent superpixels;
- perform a cross-scale self-similarity matching (12) across

the input image (I2) and the one or more auxiliary input images (I1, I3) using the superpixel test vectors; and
- generate (13) an up-scaled output image (O2) using results of the cross-scale self-similarity matching (12).

7. An apparatus (20) configured to up-scale an input image (I2), wherein a cross-scale self-similarity matching using superpixels is employed to obtain substitutes for missing details in an up-scaled image, the apparatus (20) **comprising:**
- a superpixel vector generator (7) configured to generate (10) consistent superpixels for the input image (I2) and one or more auxiliary input images (I1, I3) and to generate (11) superpixel test vectors based on the consistent superpixels;
- a matching block (5) configured to perform a cross-scale self-similarity matching (12) across the input image (I2) and the one or more auxiliary input images (I1, I3) using the superpixel test vectors; and
- an output image generator (22) configured to generate (13) an up-scaled output image (O2) using results of the cross-scale self-similarity matching (12).

8. An apparatus (30) configured to up-scale an input image (I2), wherein a cross-scale self-similarity matching using superpixels is employed to obtain substitutes for missing details in an up-scaled image, the apparatus (30) comprising a processing device (31) and a memory device (32) having stored therein instructions, which, when executed by the processing device (31), cause the apparatus (30) to:
- generate (10) consistent superpixels for the input image (I2) and one or more auxiliary input images (I1, I3);
- generate (11) superpixel test vectors based on the consistent superpixels;
- perform a cross-scale self-similarity matching (12) across

the input image (I2) and the one or more auxiliary input images (I1, I3) using the superpixel test vectors; and

- generate (13) an up-scaled output image (O2) using results of the cross-scale self-similarity matching (12).

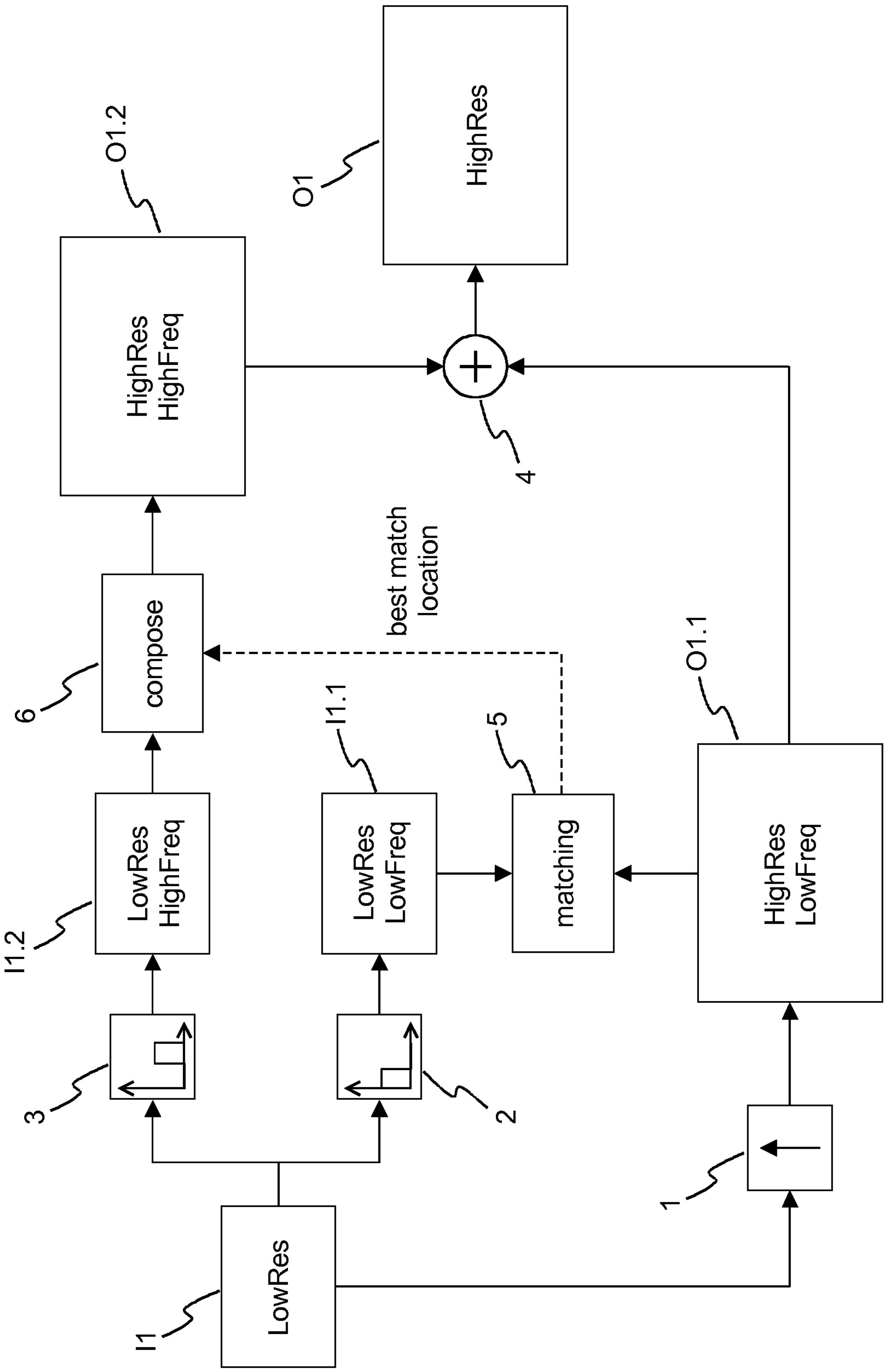
LowRes
I1
1
LowRes
HighFreq
I1.2
3
LowRes
LowFreq
I1.1
2
matching
5
HighRes
LowFreq
O1.1
compose
6
best match
location
HighRes
HighFreq
O1.2
4
HighRes
O1

Fig. 1

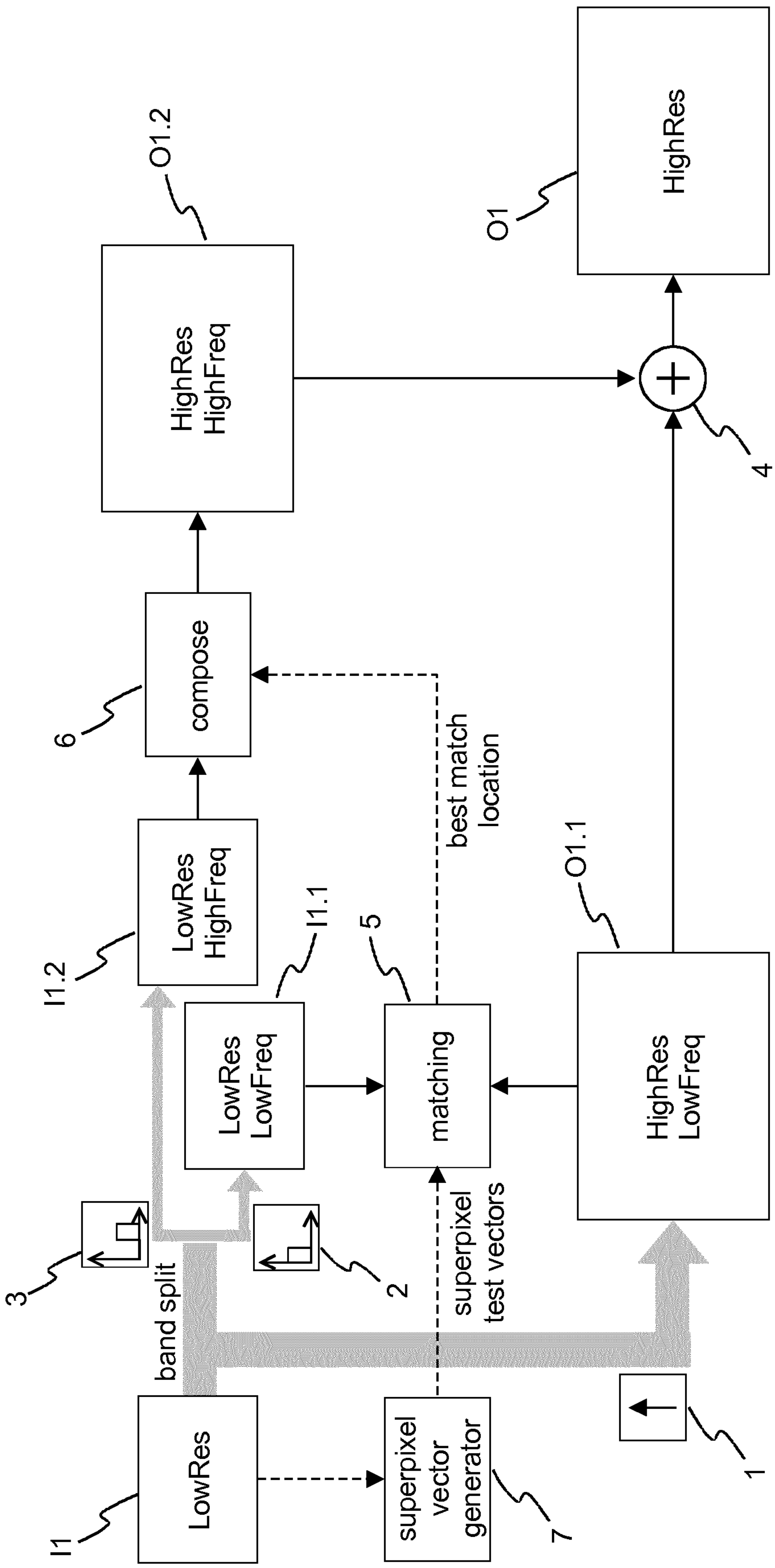
I1
LowRes
band split
3
2
I1.2
LowRes
HighFreq
I1.1
LowRes
LowFreq
5
matching
superpixel
test vectors
superpixel
vector
generator
7
1
O1.1
HighRes
LowFreq
best match
location
6
compose
O1.2
HighRes
HighFreq
4
O1
HighRes

Fig. 2

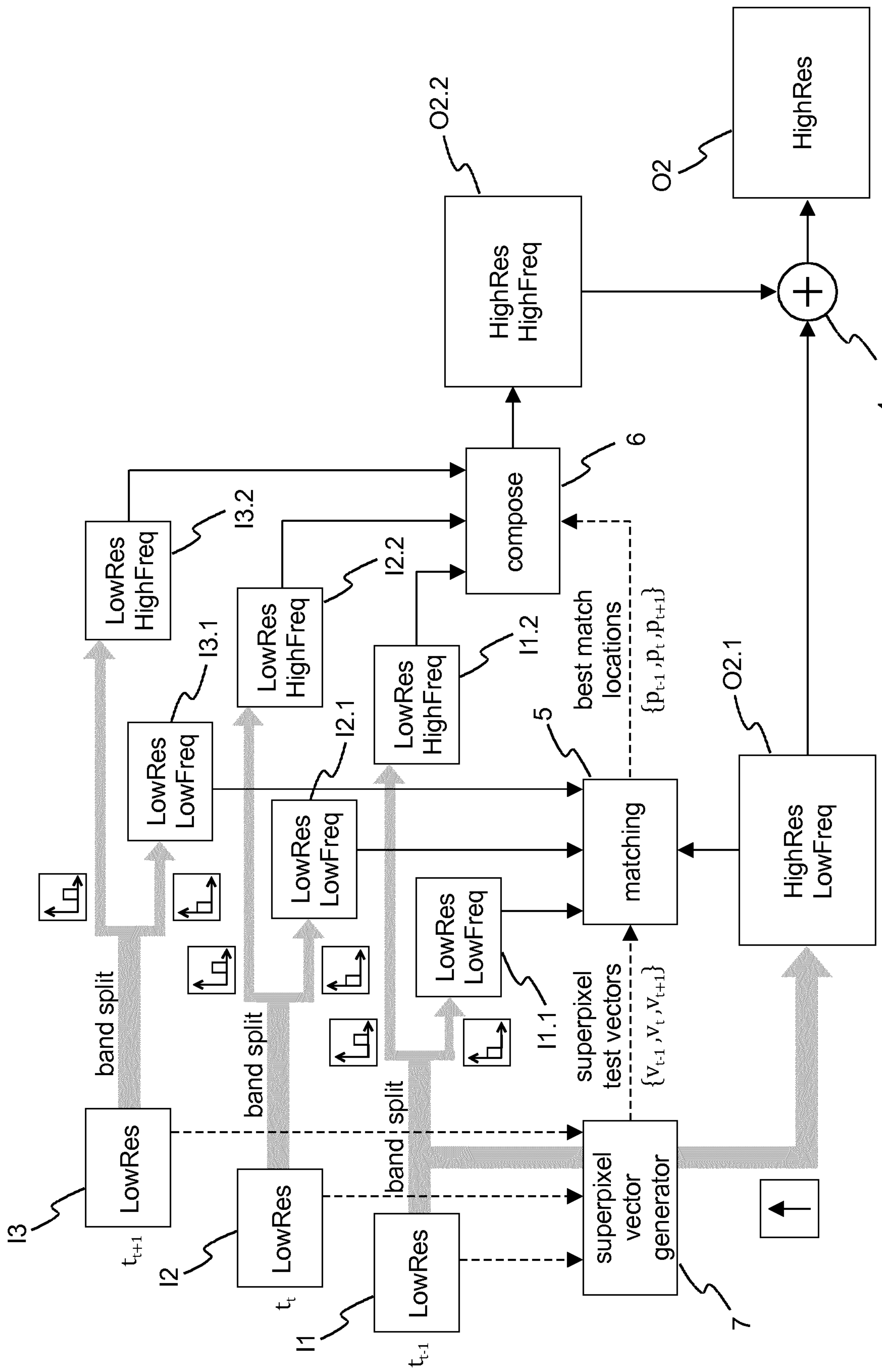
I1
I2
I3
LowRes
LowRes
LowRes
t_{t-1}
t_t
t_{t+1}
band split
band split
band split
LowRes LowFreq
LowRes LowFreq
LowRes LowFreq
LowRes HighFreq
LowRes HighFreq
LowRes HighFreq
I1.1
I1.2
I2.1
I2.2
I3.1
I3.2
superpixel vector generator
7
superpixel test vectors
$\{v_{t-1}, v_t, v_{t+1}\}$
matching
5
best match locations
$\{p_{t-1}, p_t, p_{t+1}\}$
compose
6
HighRes LowFreq
O2.1
HighRes HighFreq
O2.2
HighRes
O2
4

Fig. 3

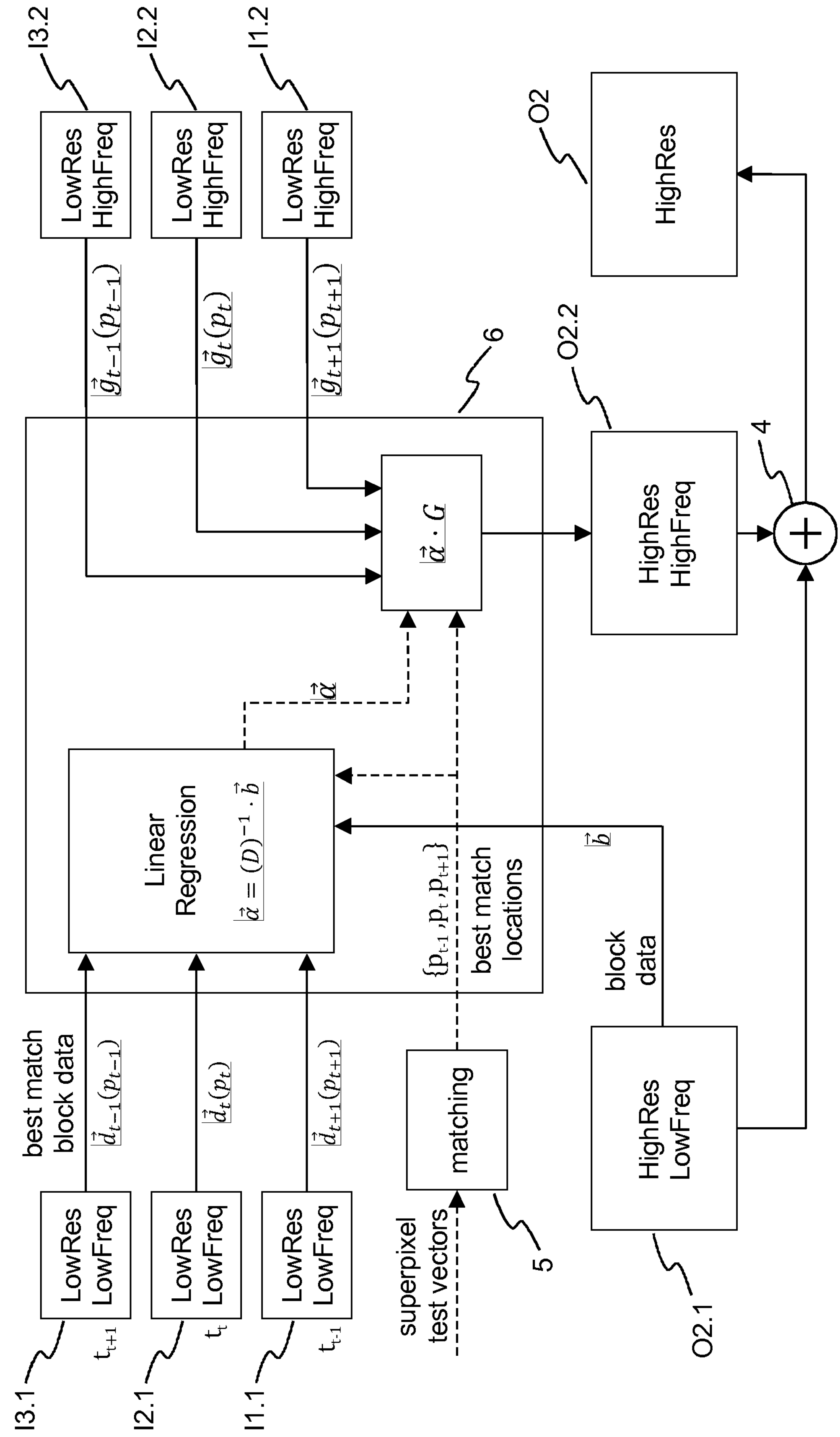
I3.2
I2.2
I1.2
LowRes HighFreq
LowRes HighFreq
LowRes HighFreq
O2
HighRes
$\vec{g}_{t-1}(p_{t-1})$
$\vec{g}_t(p_t)$
$\vec{g}_{t+1}(p_{t+1})$
6
O2.2
4
$\vec{\alpha} \cdot G$
HighRes HighFreq
$\vec{\alpha}$
Linear Regression
$\vec{\alpha} = (D)^{-1} \cdot \vec{b}$
$\vec{b}$
$\{p_{t-1}, p_t, p_{t+1}\}$ best match locations
block data
best match block data
$\vec{d}_{t-1}(p_{t-1})$
$\vec{d}_t(p_t)$
$\vec{d}_{t+1}(p_{t+1})$
matching
superpixel test vectors
5
HighRes LowFreq
O2.1
LowRes LowFreq
LowRes LowFreq
LowRes LowFreq
t_{t+1}
t_t
t_{t-1}
I3.1
I2.1
I1.1

Fig. 4

Fig. 5

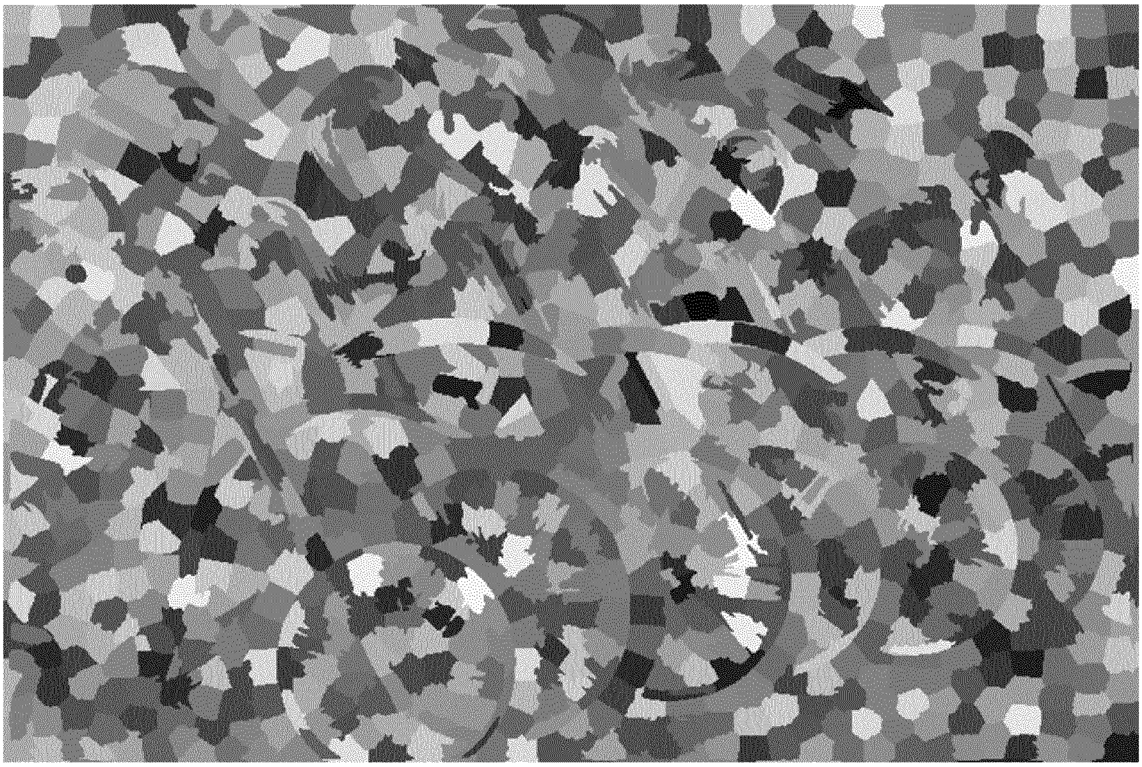

Fig. 6

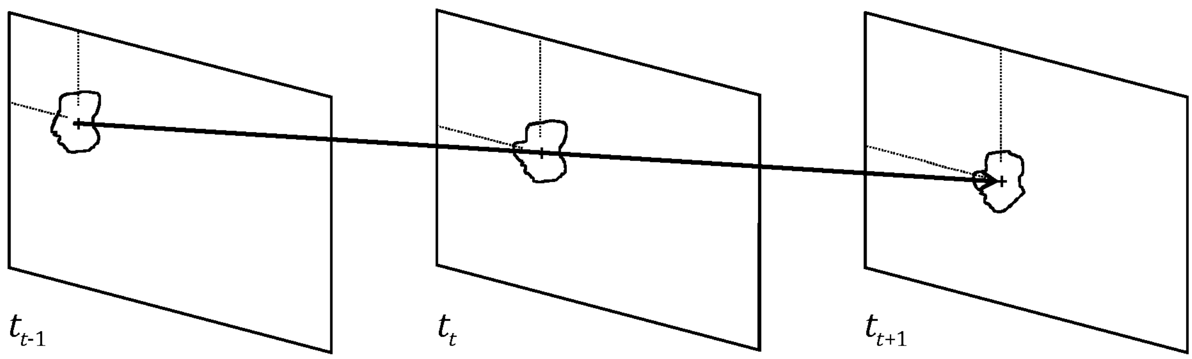
t_{t-1}
t_t
t_{t+1}

Fig. 7

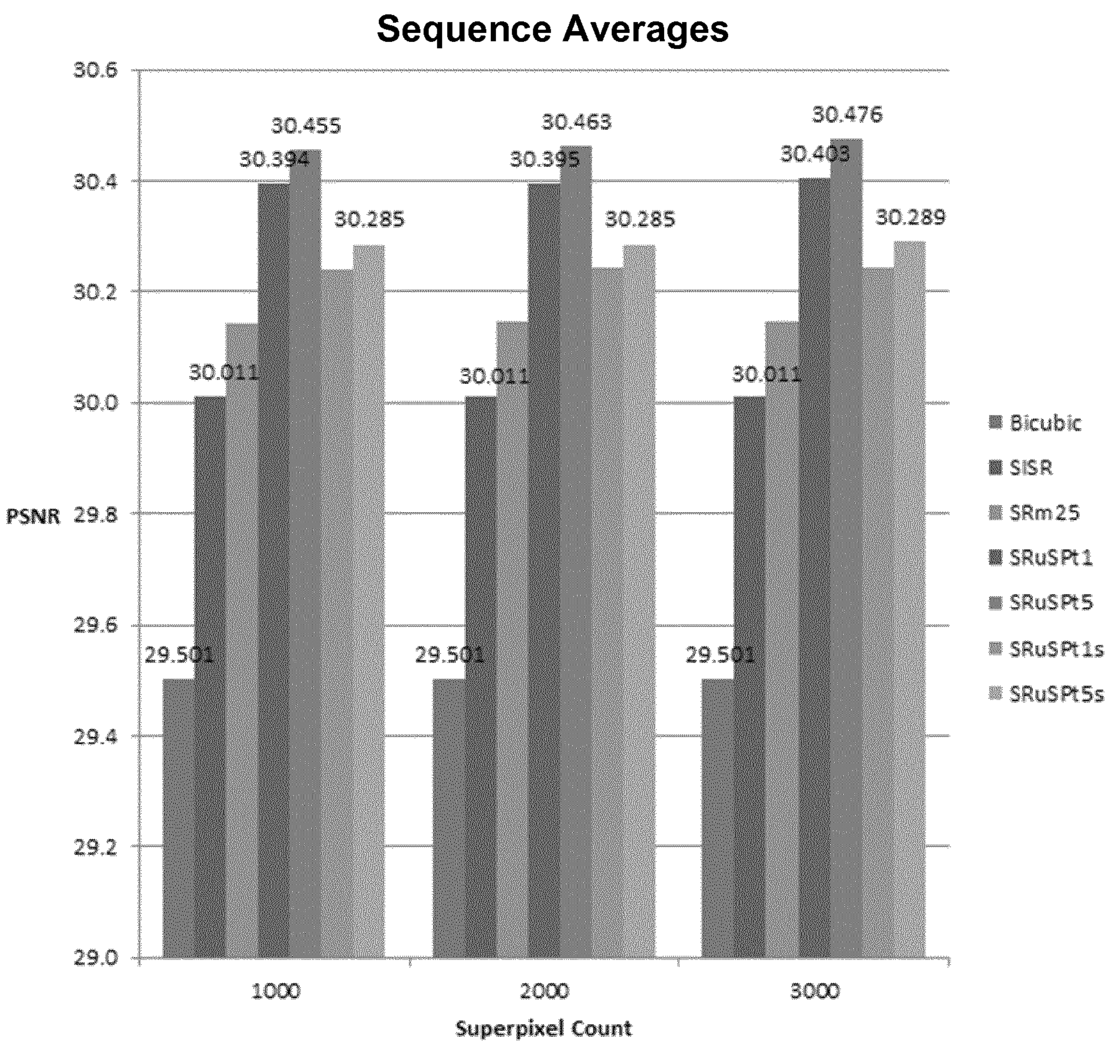
Sequence Averages
PSNR
30.6
30.4
30.2
30.0
29.8
29.6
29.4
29.2
29.0
29.501
30.011
30.394
30.455
30.285
29.501
30.011
30.395
30.463
30.285
29.501
30.011
30.403
30.476
30.289
1000
2000
3000
Superpixel Count
Bicubic
SISR
SRm25
SRuSPt1
SRuSPt5
SRuSPt1s
SRuSPt5s

Fig. 8

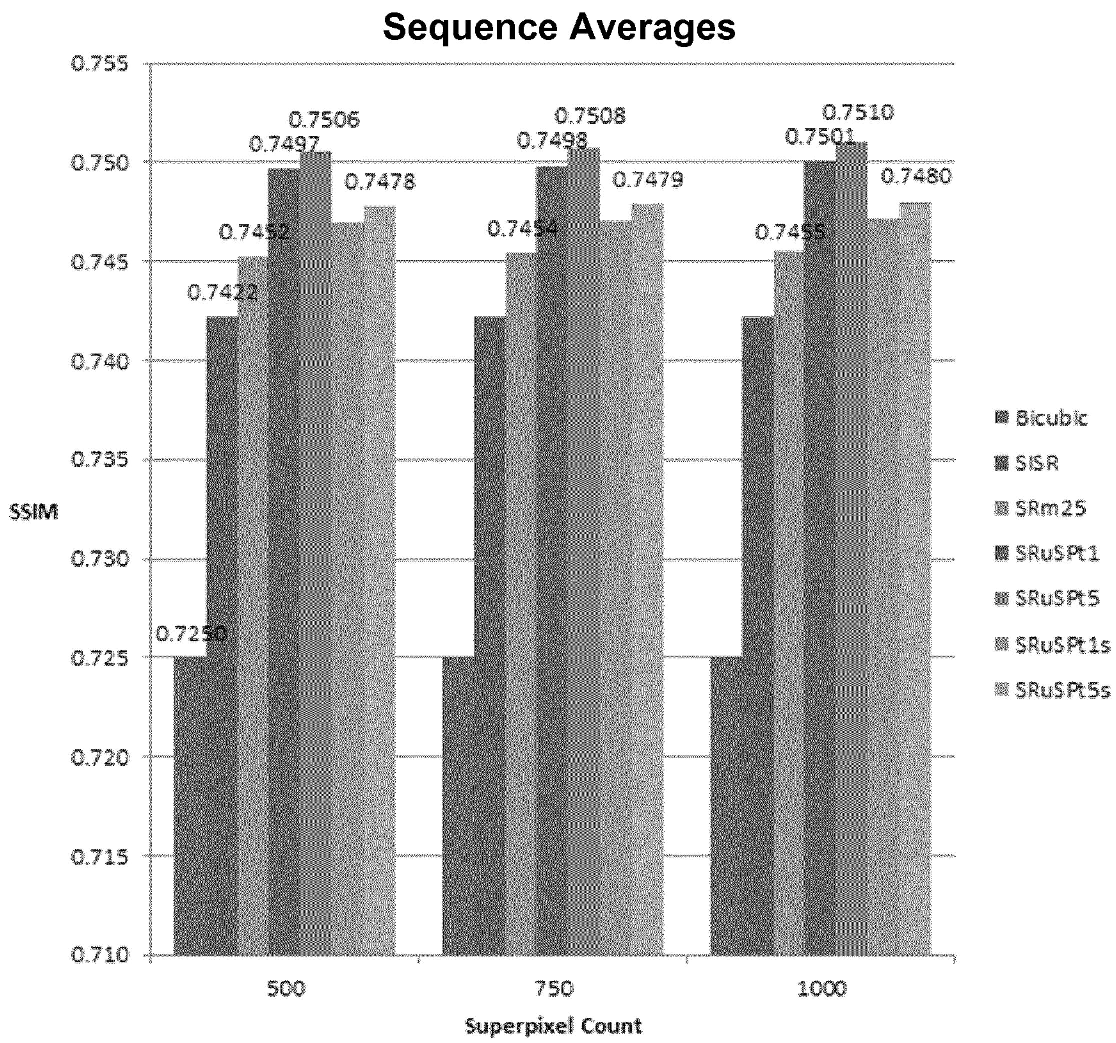
Sequence Averages
SSIM
0.755
0.750
0.745
0.740
0.735
0.730
0.725
0.720
0.715
0.710
0.7250
0.7422
0.7452
0.7497
0.7506
0.7478
0.7454
0.7498
0.7508
0.7479
0.7455
0.7501
0.7510
0.7480
500
750
1000
Superpixel Count
Bicubic
SISR
SRm25
SRuSPt1
SRuSPt5
SRuSPt1s
SRuSPt5s

Fig. 9

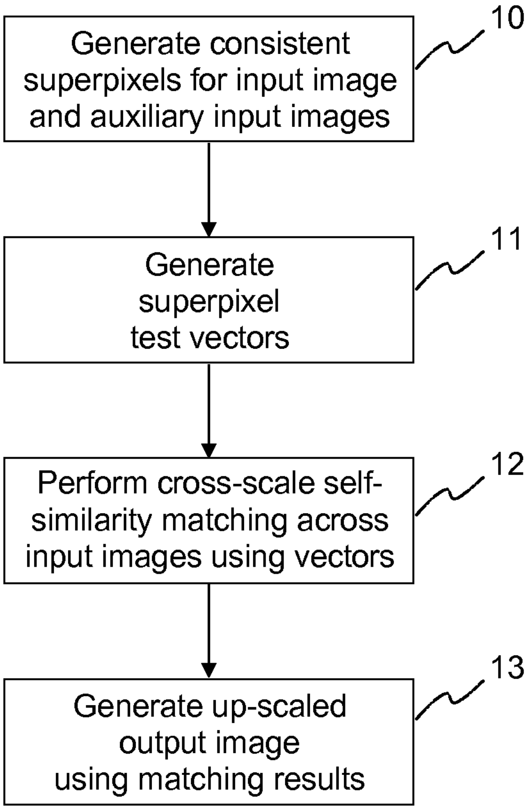
Generate consistent superpixels for input image and auxiliary input images
10
Generate superpixel test vectors
11
Perform cross-scale self-similarity matching across input images using vectors
12
Generate up-scaled output image using matching results
13

Fig. 10

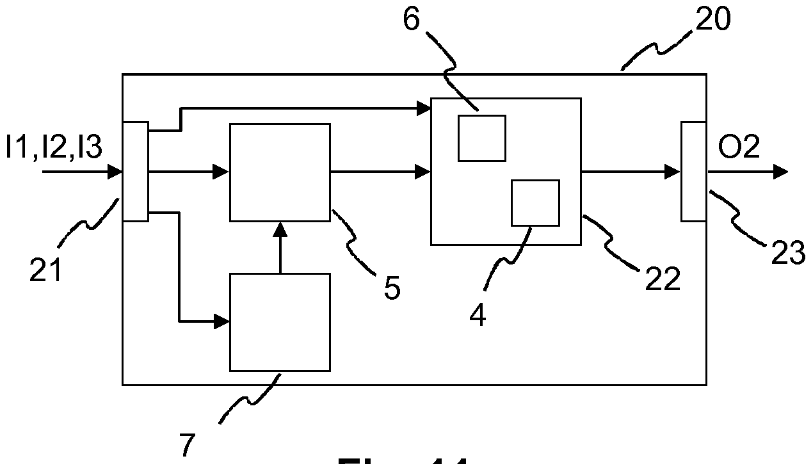

Fig. 11

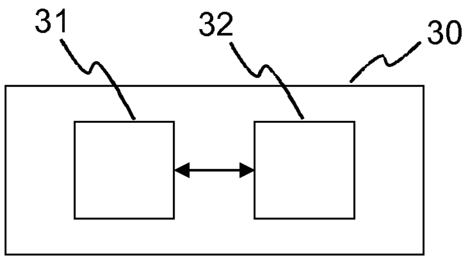

Fig. 12

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2015/064974

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, COMPENDEX, INSPEC, IBM-TDB, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	MIN-CHUN YANG ET AL: "Learning of context-aware single image super-resolution", VISUAL COMMUNICATIONS AND IMAGE PROCESSING (VCIP), 2011 IEEE, IEEE, 6 November 2011 (2011-11-06), pages 1-4, XP032081417, DOI: 10.1109/VCIP.2011.6116046 ISBN: 978-1-4577-1321-7 the whole document ----- -/--	1-8

[X] 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	Date of mailing of the international search report
2 October 2015	13/10/2015

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Rockinger, Oliver

1

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2015/064974

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	SALVADOR JORDI ET AL: "Patch-based spatio-temporal super-resolution for video with non-rigid motion", SIGNAL PROCESSING. IMAGE COMMUNICATION, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, NL, vol. 28, no. 5, 5 March 2013 (2013-03-05), pages 483-493, XP028588765, ISSN: 0923-5965, DOI: 10.1016/J.IMAGE.2013.02.002 cited in the application page 484 - page 488 -----	1-8
A	RESO MATTHIAS ET AL: "Temporally Consistent Superpixels", 2013 IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION, IEEE, 1 December 2013 (2013-12-01), pages 385-392, XP032572909, ISSN: 1550-5499, DOI: 10.1109/ICCV.2013.55 [retrieved on 2014-02-28] the whole document -----	1-8
X,P	Dirk Gandolph ET AL: "D4.4.1 Scene Compression, Simplification and Super-Resolution", , 31 July 2014 (2014-07-31), XP55167397, Retrieved from the Internet: URL:http://3d-scene.eu/pdfs/delis/SCENE-D4.4.1-20140731_final.pdf [retrieved on 2015-02-04] page 5 - page 18 -----	1-8

1