PSyCo: Manifold Span Reduction for Super Resolution

Eduardo Pérez-Pellitero[1,2], Jordi Salvador[2], Javier Ruiz-Hidalgo[3] and Bodo Rosenhahn[1]

[1]TNT Lab, Leibniz Universität Hannover
[2]Technicolor R&I Hannover
[3]Image Processing Group, Universitat Politècnica de Catalunya

Abstract

The main challenge in Super Resolution (SR) is to discover the mapping between the low- and high-resolution manifolds of image patches, a complex ill-posed problem which has recently been addressed through piecewise linear regression with promising results. In this paper we present a novel regression-based SR algorithm that benefits from an extended knowledge of the structure of both manifolds. We propose a transform that collapses the 16 variations induced from the dihedral group of transforms (i.e. rotations, vertical and horizontal reflections) and antipodality (i.e. diametrically opposed points in the unitary sphere) into a single primitive. The key idea of our transform is to study the different dihedral elements as a group of symmetries within the high-dimensional manifold. We obtain the respective set of mirror-symmetry axes by means of a frequency analysis of the dihedral elements, and we use them to collapse the redundant variability through a modified symmetry distance. The experimental validation of our algorithm shows the effectiveness of our approach, which obtains competitive quality with a dictionary of as little as 32 atoms (reducing other methods' dictionaries by at least a factor of 32) and further pushing the state-of-the-art with a 1024 atoms dictionary.

1. Introduction

The latest achievements in the research of example-based SR have been possible thanks to the exploitation of the progressively deeper understanding of the structure of the natural image patch manifold. In recent years we have witnessed a shift from costly sparse coding techniques to more efficient models. In the latter, a relatively costly offline manifold learning enables light inference with various levels of efficacy in terms of complexity and accuracy [21, 22, 26, 5, 25]. The development of these two as-
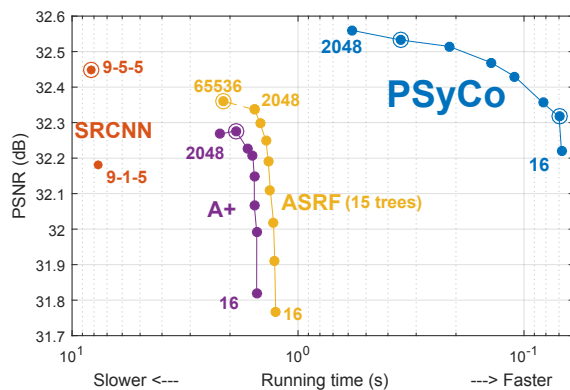

Figure 1: PSNR vs time (s) of our algorithm compared to other SR methods for dictionary sizes from 16 to 2048, in power-of-two increments. Experiment run on Set14 and $\times 2$ magnification factor. Circled points are found in Table 1.

pects is key for the pervasive adoption of SR, not only as an end application, but also as a pre-processing stage in vision applications where the resolution of the capture device is insufficient for scene understanding or feature detection.

Within the sparse coding techniques [27, 29, 14], the main idea is to enforce image pairs to have the same sparse representations over both the low resolution and the high resolution dictionaries. However, this sparsity constrain involves minimizing a L_1 norm function, both at the training and inference stage, leading to complex minimization procedures. Newer solutions to avoid sparse representations involve the use of a regression-based approach [21, 25, 26]. In this case, the main goal is to infer a mapping between low- and high-resolution manifolds. This mapping, potentially complex and non-linear, is approximated by a piecewise linearization employing several linear regressors at specific anchor points. Sublinear search structures such as trees, hashing functions or forests [17, 25, 19, 18] have been proposed in order to reduce the regressor search complexity. Other current light-inference solutions introduce deep neural net-

works for SR with encouraging results [6, 7]. Although not necessarily light, single-image self-similarity SR have shown competitive results partly thanks to improved search strategies, such as [12]. In a similar direction, there are several efforts in the literature to extend the available manifold samples. Most of the approaches aim to expand the search space by generating new data from the available one, e.g. multi-scale pyramids [10, 22], homographies [12], with the resultant increase in the number of search candidates.

In this paper we present *PSyCo* (**P**atch **Sy**mmetry **Co**llapse), a method that improves the search without increasing its number of candidates. Our contributions include the use of the dihedral group (i.e. reflections and rotations) over more complex projective transformation models and the introduction of a collapsing transform κ that collapses the 16 redundant patch variations induced by the dihedral group and the antipodality. This transform has an inherently low complexity and therefore is specially suitable for fast SR algorithms and, by extension, to other patch-based methods.

2. Related work

2.1. Problem formulation

SR techniques aim to estimate a high-resolution (HR) image X from a low-resolution (LR) image Y which has an unsatisfactory pixel resolution, assuming the following generation model:

$$Y = (X * H) \downarrow s, \tag{1}$$

where H is a low-pass filter and $\downarrow s$ denotes downsampling operator for an s magnification factor. This inverse problem is usually addressed at a patch level, which we denote with small case (e.g. x, y, c) when in their original square shape and with bold small case (e.g. $\mathbf{x}, \mathbf{y}, \mathbf{c}$) when vectorized.

2.2. Dictionary learning

The usage of a pair of sparse dictionaries ($\mathbf{D}_h$ and $\mathbf{D}_l$) was popularized by the SR method of Yang et al. [27], in which they propose a sparse prior for SR. Sparse coding represents input vectors as a weighted linear combination of a small set of basis vectors, thus extracting high level patterns of the input unlabeled data and obtaining compact and meaningful dictionaries. In their sparse SR [27], the optimization of the dictionary is performed jointly for both $\mathbf{D}_h$ and $\mathbf{D}_l$, resulting in great computational cost. During testing time, they minimize the following function:

$$\min_{\boldsymbol{\alpha}} \|\mathbf{y} - \mathbf{D}_l \boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1, \tag{2}$$

where the first term ensures a good LR reconstruction and the L_1-norm regularization term enforces sparsity in the solution. The sparse decomposition $\boldsymbol{\alpha}$ is then applied to

$\mathbf{D}_h$ to obtain the HR patch. Later work on sparse SR by Zeyde et al. [29] introduced faster algorithms for dictionary optimization (e.g. k-SVD [1]) and a different optimization scheme: the dictionaries are learned separately, obtaining first $\mathbf{D}_l$ independently from $\mathbf{D}_h$, and afterwards the latter is generated with the sparse encoding of $\mathbf{D}_l$. The results improve both in time and quality those of the original work of Yang et al. [27].

Despite alleviating some of the most time-consuming processing of its predecessor, the sparse decomposition in [29] is still the bottleneck during inference time. As a natural solution for that, Timofte et al. proposed the Anchored Neighborhood Regression (ANR) [21] where there is no sparse decomposition during inference time, but instead a selection within a discrete set of points (i.e. anchor points) for which a linear ridge regressor has been trained off-line. This method coincided with other similar regression-based SR algorithms such as [25, 26], and also triggered several follow-ups, e.g. [22, 19, 17].

2.3. Coarse to fine regression

Regression-based SR tackles the upscaling problem by estimating the mapping function between the LR and the HR manifolds, most commonly through an off-line learning stage. Under a locally linear assumption [4], this function can be split into several piecewise linear regression functions that are trained with a subset of the full training set, e.g. a cluster or neighborhood. For the sake of clarity, we introduce the regression framework using the opening work of ANR [21], as this is fundamentally the framework we use in our proposed method.

Let us denote by $\mathbf{D}_l$ the set of k anchor points and $\mathbf{D}_h$ the correspondent HR counterparts, both of them obtained with k-SVD as in the scheme of Zeyde et al. [29]. For each atom $\mathbf{d}_i$ in $\mathbf{D}_l$, a certain neighborhood or cluster $\mathbf{N}_l^i$ is found through a nearest neighbor (NN) search from a training pool of examples, which in ANR is the same $\mathbf{D}_l$. The associated ridge regressor R_i is then trained with the closed form expression $R_i = \mathbf{N}_h^i (\mathbf{N}_l^{i\mathsf{T}} \mathbf{N}_l^i + \lambda \mathbf{I})^{-1} \mathbf{N}_l^{i\mathsf{T}}$.

During inference, the closest anchor point $\mathbf{d}^*$ to the input features $\mathbf{y}_F$ is obtained through NN search and the associated regressor R^* is applied in the following form:

$$\mathbf{x} = \mathbf{c} + R^* \mathbf{y_F}, \tag{3}$$

where $\mathbf{c}$ is a first coarse approximation of $\mathbf{x}$ (obtained by e.g. bicubic, Iterative Back Projection (IBP) [13, 27]). In the later A^+ [22] and the Dense Local Regressor [17], $\mathbf{N}_l$ is obtained from a large pool of training samples (in the order of millions), which greatly improves quality results.

2.4. Search space and manifold collapse

Finding meaningful examples for SR is crucial both for internal learning (where the search space is limited by the
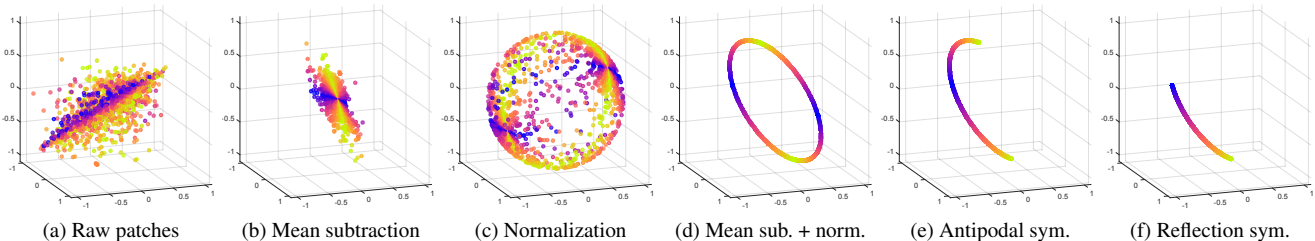
| (a) Raw patches | (b) Mean subtraction | (c) Normalization | (d) Mean sub. + norm. | (e) Antipodal sym. | (f) Reflection sym. |

Figure 2: Reduction of the manifold's span and complexity by the procedures introduced in Section 3. The manifold is composed of three dimensional (i.e. 3×1) patches in the range of $[-1, 1]$ extracted from images.

image) and external learning. In this direction, Timofte et al. [22] proposed to generate new training data from different multi-scale images, Zhu et al. [30] proposed to deform patches based on optical flow and, more recently, Huang et al. [12] incorporate 3D scene geometry for cross-scale self-similarity using a modified PatchMatch [2].

Another approach to improve the NN search consists in reducing variability of the manifold through the knowledge of its redundancy. In the early work of Freeman et al. [9], the concept of improving the NN search through the *collapse* of the manifold's variability was already addressed. In their learning process, to predict the highest frequency band they only consider the mid-frequency band and discard the rest of low-frequencies (LF), thus *collapsing the training data for all possible LF values into one value*. Similarly in concept, when subtracting the mean to a patch, all possible means are mapped to a single 0-mean patch. The benefit of removing the undesired variability of the manifold versus generating more data is obvious as the first one obtains the same advantages while not increasing the number of search candidates. In this paper we further deepen the knowledge of natural image patch manifold, analyzing the redundancy present within the manifold due to the dihedral group of transforms (i.e. rotation, vertical and horizontal reflections), which are invariant across scales and easily invertible (i.e. a lossless $f^{-1}(x)$ exists).

3. Reducing the manifold span

In this section we first overview two basic patch pre-processing steps (mean subtraction and normalization) and their effects within the manifold, followed by a geometric transformation model that can reduce the manifold span (extended in Section 4) and its analysis in the Discrete Cosine Transform (DCT) space. An overview of the presented transformation is shown in Fig. 2.

3.1. Mean subtraction and normalization

Mean subtraction is an inexpensive process widely adopted in SR applications, as it is specially beneficial since the mean presents no variations across scales. Bevilaqua et
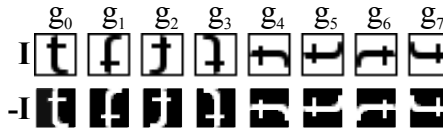

Figure 3: D_4 dihedral transforms applied to a 20x20 patch and its corresponding antipodal versions denoted with $-\mathbf{I}$.

al. [3] concluded in their feature analysis that the centered luminance patches are the best suited for their non-negative neighbor embedding SR. Within the manifold structure, mean subtraction collapses all the possible patches to lie on the hyperplane $\mathbf{1}^\intercal \mathbf{x} = 0$, as shown in Fig. 2b.

Patch normalization is also a simple yet effective process very present in low-level vision, often interpreted as an illumination normalization. Patch normalization removes the undesired variability derived from scalar multiplication: All positive scalar variations are represented by a single unitary vector (i.e. a certain patch structure). In terms of manifold transformation, normalization enforces the patches to lie in the unitary hypersphere, as we show in Fig. 2c. The combination of both mean subtraction and normalization limits the span of the manifold to the intersection of the mean hyperplane and the unitary hypersphere, a ring in the 3-dimensional example of Fig. 2d.

3.2. Antipodality

Antipodal points (i.e. points that are diametrically opposed in the unitary sphere: $\mathbf{x}_A = -\mathbf{x}$) cannot be properly collapsed by patch normalization as norms are strictly positive, so any two normalized antipodal points are located at the furthest away Euclidean distance (the diameter of the hypersphere) while actually the structure of the patch is exactly the same (see Fig. 3). In our previous work we already introduced antipodal invariance for SR [16, 15]. It is possible to collapse antipodal variability together with dihedral transformations as described in Section 4 and illustrated in Fig. 2e-2f.

3.3. Transformation models

Within the space of patches, numerous 2D geometric transformations have been proposed in order to model physical displacements in the 3D world, improve invariance to those transforms (e.g. rotation for object detection) or expand the search space both in testing and training. A general model for such transformations is the projective transformation model, also referred as *homography* or *collinearity*.

The projective transformation properly describes the possible transformations of a pinhole camera when moving to an arbitrary viewpoint. Homographies are widely used in several applications involving multiple cameras or camera motion [8, 11], and they have been also used recently in SR [12] in order to increase the number of relevant patches in the NN search.

Homographies show two main drawbacks when applied to SR. Firstly, as small patches present a very scarcely sampled grid, transforming its geometry requires interpolating values, which leads to a high-frequency loss. Secondly, the homography transform has 8 degrees of freedom, therefore being computationally expensive to explore and estimate (e.g. Huang et al. [12] use an affine transform enriched with some perspective deformation limited to a discrete set of detected planes).

In this paper, we propose the usage of the dihedral group D_4 (for polygons of 4 sides, e.g. patches) [23], which is a subset of affine transformations that only includes rotations and reflections. This finite group $G = \{g_j\}_{j=0}^{7}$ contains 8 structure-preserving transforms which just re-distribute the elements within a patch and therefore do not require any interpolation. We can obtain the set of 8 dihedral transforms G via a combination of the following matrices in the 2D space:

$$g_x = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \; g_y = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \; g_{\mathsf{T}} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad (4)$$

where g_x and g_y denote the reflections along the x and y axis respectively, and g_{T} denotes the transpose operation. All the transforms forming the dihedral group are linear and scale invariant, and a straightforward inverse function exists. In Fig. 3 we show the behavior of the dihedral group of transforms and how they affect a given patch.

3.4. Dihedral group in the DCT space

In this section we analyze the effect of the dihedral group G in the domain of the DCT, as there are some useful properties that lay the groundwork for our proposed method. The DCT b of a patch x of size $M \times N$ reads:

$$b(k,l) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} x(m,n) \cos \frac{\pi(m+\frac{1}{2})k}{M} \cos \frac{\pi(n+\frac{1}{2})l}{N}.$$

$$(5)$$

As the DCT is linear, applying the transpose operator (i.e. g_{T} in the 2D space) results in a transpose in the transformed space, i.e. $b^{\mathsf{T}} = f_{DCT}(x^{\mathsf{T}})$. As for the reflection operators (i.e. g_x and g_y in the 2D space), they result in a change of sign in some of its components:

$$\begin{aligned} b_{g_x}(k,l) &= b(k,l) \cdot (-1)^l \\ b_{g_y}(k,l) &= b(k,l) \cdot (-1)^k \end{aligned} . \qquad (6)$$

The behavior of the proposed dihedral transforms in the DCT space is therefore reduced to transpositions and sign changes in a defined set of coefficients. Fig. 5 left shows which components of the DCT are expected to change whenever there is a reflection or transposing operator. This simple and predictable behavior in the DCT space facilitates the observation of mirror symmetries.

4. Manifold symmetries

The transform group G presented in Section 3.3 defines 8 points in the $M \times N$-dimensional manifold of natural patches for a given patch primitive x (see Fig. 5 right). This is a dihedral symmetric shape within the manifold surface, since a symmetric structure is defined if there exists a non-trivial group of action that defines an isomorphism. Our goal is to exploit the symmetries defined by G together with antipodality in order to efficiently collapse redundant variability of our manifold span.

Our proposed Symmetry-Collapsing Transform (SCT) builds on the work of Zabrodsky et al. [28], where they proposed a continuous Symmetry Distance (SD) which measures how symmetric a given structure is. This metric δ is defined in the shape space Ω, where each shape is represented by a sequence of r points $\{P_i\}_{i=0}^{r-1}$. The metric reads:

$$\delta(P,Q) = \frac{1}{r} \sum_{i=0}^{r-1} \|P_i - Q_i\|^2, \qquad (7)$$

which is an averaged point to point Euclidean distance. In order to achieve invariance to symmetry, a Symmetry Transform (ST) of a shape P is defined as the symmetric shape closest to P in terms of Equation (7), and thus SD is defined as $SD = \delta(P, ST(P))$. The metric is therefore the point to point Euclidean distance of a given shape to its closest symmetric shape.

Zabrodsky et al. [28] present different ST depending on the type of symmetry to be accounted (e.g. rotational, mirror-symmetry). For the specific case of the mirror-ST, with a known mirror symmetry axis, the procedure for every pair of points $\{P_0, P_1\}$ is:

Fold by reflecting the point across the mirror symmetry axis obtaining $\left\{\mathring{P}_0, \mathring{P}_1\right\}$ (i.e. $P_0 \equiv \mathring{P}_0$).

Average both points to obtain a new average point A_0.

Unfold the average point A_0 in order to obtain A_1.

We show an overview of the original mirror-ST in Fig. 4 (steps 1 and 2a). In the original algorithm, the ST aims to obtain a regular polygon which can be thereafter compared to the input shape in order to estimate its point to point distance. Our goal is to obtain a transform that reduces variability while respecting the SD.

To achieve this reduction, we present a modified ST, which we denote as SCT, that moves all the possible symmetric points to a reference side of the mirror axis, thus reducing redundant variability. For that purpose, assuming a single mirror axis, all the points are fold into the reference side where P_0 lies, and the element of the applied symmetry group (i.e. g_j) is saved. This is similar to a mean subtraction , where all possible different means of a given patch are collapsed to a single 0-mean patch and the mean is saved in order to differentiate among them. We show an overview of our proposed SCT in Fig. 4 (steps 1 and 2b), where we highlight that the resulting distances are conserved with respect to the original algorithm. Although folding the points back to their original position is not necessary for the distance calculation in our SCT, we can do it at any point as the inverse SCT.

The initial ST and SD extend to any finite point-symmetry group G in any dimension, where the folding and unfolding are performed by applying the group elements [28]. However, when extending to more than 3D, finding the symmetry axes that minimize SD is non-trivial.

In order to (a) keep the transform under a reasonable complexity, (b) easily and analytically find the mirror axes of G and (c) benefit from behavior of G in the DCT domain, we propose a representation based on the first vertical and horizontal harmonics $b(1,0)$ and $b(0,1)$. Each of these coefficients is affected only by one reflection and the transpose is plainly mapped to a coefficient switch. Semantically, $b(1,0)$ and $b(0,1)$ are the coefficients statistically containing more energy that represent the response to vertical and horizontal variations, resembling the original vertical and horizontal 2D space of Zabrodsky et al. [28]. The three resulting mirror planes are straightforwardly obtained as $b(1,0) = 0$, $b(0,1) = 0$ and $|b(1,0)| - |b(0,1)| = 0$, as shown in Fig. 5 right. At this stage, there is still ambiguity within this projected space as an antipodal point can be confused by a patch affected by vertical and horizontal reflections (as both vertical and horizontal coefficients have a sign change). In order to disambiguate, we include another dimension and a fourth mirror plane in $b(3,3) = 0$ which is not affected by transpose, nor vertical or horizontal reflection (as it is a DCT base with inner dihedral symmetry). This fourth axis, which we fold in the first place, represents the negative unitary matrix $-\mathbf{I}$ (i.e. sign change) to be applied both patch-wise and within the DCT domain before collapsing the rest of symmetries.

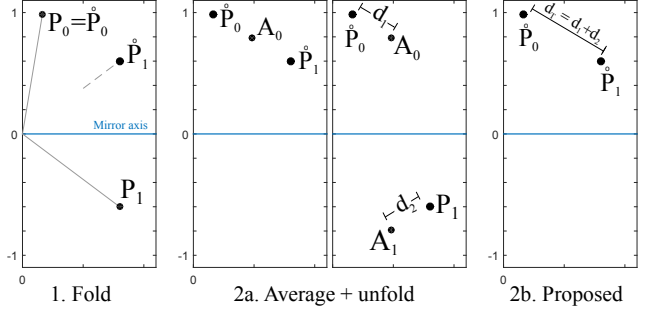The final proposed transform $\mathring{c} = \kappa(c, \varphi(c))$ produces

Figure 4: Mirror-Symmetry Transform of a single pair of points as proposed by Zabrodsky et al. [28] (1 and 2a) and our proposed SCT (1 and 2b).
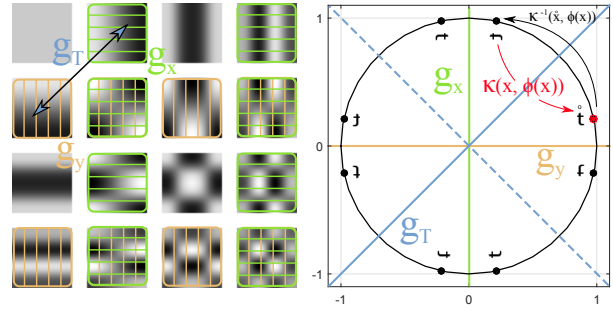

Figure 5: **Left**: Coefficients of a DCT that are affected by g_x, g_y (resulting in a sign change, Eq. 6) and g_T (resulting in a transpose of coefficients). **Right**: Overview of our $\kappa(x, \varphi(x)$ with real patches, highlighting the symmetry axes associated to each operator.

collapsed patches (denoted by the ring accent) using the four defined axes, where $g_j = \varphi(c)$ retrieves the element within the group G together with the disambiguation of the sign (i.e. $-\mathbf{I}$ when $b(3,3) < 0$). The inverse $c = \kappa^{-1}(\mathring{c}, \varphi(c))$ applies the same elements of the symmetry group that were used in the collapse in a reverse order, restoring the patch to its original appearance.

5. Application to SR

In this section we propose a novel SR algorithm that makes use of our proposed $\mathring{c} = \kappa(c, \varphi(c))$, which we name *PSyCo* (**P**atch **Sy**mmetry **Co**llapse). As briefly mentioned in the related work, we build on the ANR [21] framework of anchored regression. We denote 0-mean patches with the line accent (e.g. $\bar{\mathbf{c}}$). The main idea is to train our regression ensemble (both k anchor points in $\mathbf{D_1}$ and the associated regressors $\{R_i\}$) with the ground truth and coarse collapsed patches $\{\mathring{\mathbf{x}}, \mathring{\mathbf{c}}\}$ so that during training time the system is optimized for the reduced span of the manifold which is to be used. We obtain our coarsely approximated images C with IBP as presented in [27, 15]. The k-SVD input is a matrix of 0-mean patches without symmetric redundancy which have
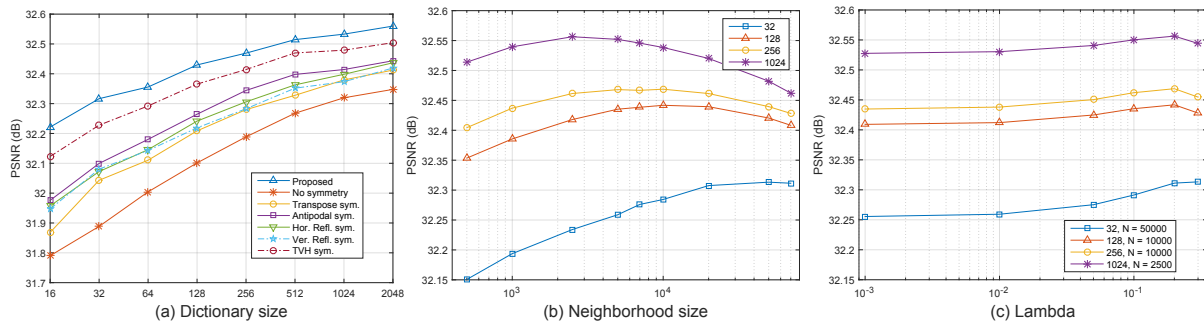
Figure 6: Different configurations of *PSyCo*. (a) Shows the PSNR performance for different mirror symmetries, (b) shows the impact of the neighborhood size and (c) the impact of the regularization weighting term λ.

been stacked as columns, denoted by $\overset{\circ}{\overline{\mathbf{C}}}$. After that, a NN search with the angular similarity $\left|\overset{\circ}{\overline{\mathbf{d}}}_i^\intercal \overset{\circ}{\overline{\mathbf{c}}}\right|$ is performed for each atom $\mathbf{d}_i$ in $\mathbf{D}_l$ to construct each neighborhood $\mathbf{C}_i$ as a fixed-size subset of the whole training data $\mathbf{C}$. Once the anchor points and neighborhoods have been defined, each regressor R_i is trained with the following closed-form expression:

$$R_i = (1 + \lambda)(\overset{\circ}{\mathbf{X}}_i - \overset{\circ}{\mathbf{C}}_i)\overset{\circ}{\overline{\mathbf{C}}}_i{}^\intercal(\overset{\circ}{\overline{\mathbf{C}}}_i \overset{\circ}{\overline{\mathbf{C}}}_i{}^\intercal + \lambda\mathbf{I})^{-1}. \quad (8)$$

During inference time, the NN search and regression is performed with $\left\{\overset{\circ}{\overline{\mathbf{d}}}, \overset{\circ}{\overline{\mathbf{c}}}\right\}$ and after regression the symmetric transformation needs to be reverted so that the patches recover their original orientation. The regression stage reads:

$$\tilde{\mathbf{x}} = \mathbf{c} + \kappa^{-1}(R^* \overset{\circ}{\overline{\mathbf{c}}}, \varphi(\overline{\mathbf{c}})), \quad (9)$$

and the final image $\tilde{X}$ is obtained by an overlapping reconstruction strategy, as it is common in SR [21, 26, 19].

5.1. Configuration

In this section we validate the contributions of our proposed transform, assessing the impact of collapsing each of the axes separately, and also the combination of those exclusively corresponding to the dihedral group G and the impact of the complete system, which also tackles antipodal symmetries. Figure 6a shows *PSyCo* with several mirror-axes configurations and dictionary sizes. First, we would like to asses the benefits of our symmetric transform when compared to untransformed patches. The quality is around 0.4 dB higher for small dictionary sizes (e.g. 16, 32) and around 0.2 dB for 1024 atoms. We find remarkable the fact that our symmetry transform performs always slightly better than a $\times 16$ times larger dictionary without any symmetry accounted. This supports the idea that with our manifold collapse we can effectively cover the 16 different appearances of a given primitive patch without increasing the search space, plus an additional quality gain as the training of the regressors is better (i.e. due to more meaningful patches in the neighborhoods).

When it comes to assess the incidence of each type of transform separately, we find that all have similar impact, being the antipodal symmetry slightly better-performing than the reflection or the transpose. We also note that each symmetry axis is roughly comparing equally to a $\times 2 - 4$ times larger untransformed dictionary. The dihedral symmetries together surpass that of the antipodal, and we observe that its quality performance surpasses by a great margin that of the $\times 8$ larger dictionary without any symmetry.

In Fig. 6b and 6c we show the behavior of the only two parameters to be selected in our SR algorithm. The neighborhood size has higher impact and its optimal value increases for smaller dictionary sizes (as each neighborhood covers more span within the manifold). The regularization weighting term λ has a lesser impact and its optimal value increases for big neighborhoods and small dictionaries. We also note that $\lambda = 0.2$ is a good compromise across all possible configurations, and thus we recommend its usage for a first approach when optimizing the neighborhood size.

6. Results

We perform several experiments comparing with current SR state of the art. The experimental set-up is:

Datasets: We perform our SR benchmarking with `Set5`, `Set14`, `kodak` and `Urban100` ([12] used different resolutions for each s, we set the resolution corresponding to $s = 4$ for all s). **Compared methods**: We compare against the current SRCNN deep learning method presented by Dong et al. [7] with their recommended 9-5-5 network (note the superior performance when compared to the 9-1-5 network of their earlier publication [6]), the A^+ anchored regression algorithm of Timofte et al. [22], the recently published SR forest with alternative training ASRF of Schulter et al. [19] and the Transformed Self-Exemplars Single-Image SR of Huang et al. [12]. **Conditions**: We present two different configurations, with 32 and 1024 atoms. For the first one, we set a neighborhood size of 42000 and $\lambda = 0.25$; for the second one the neighborhood size is set to 2750 and $\lambda = 0.18$. We train A^+, ASRF and *PSyCo* with

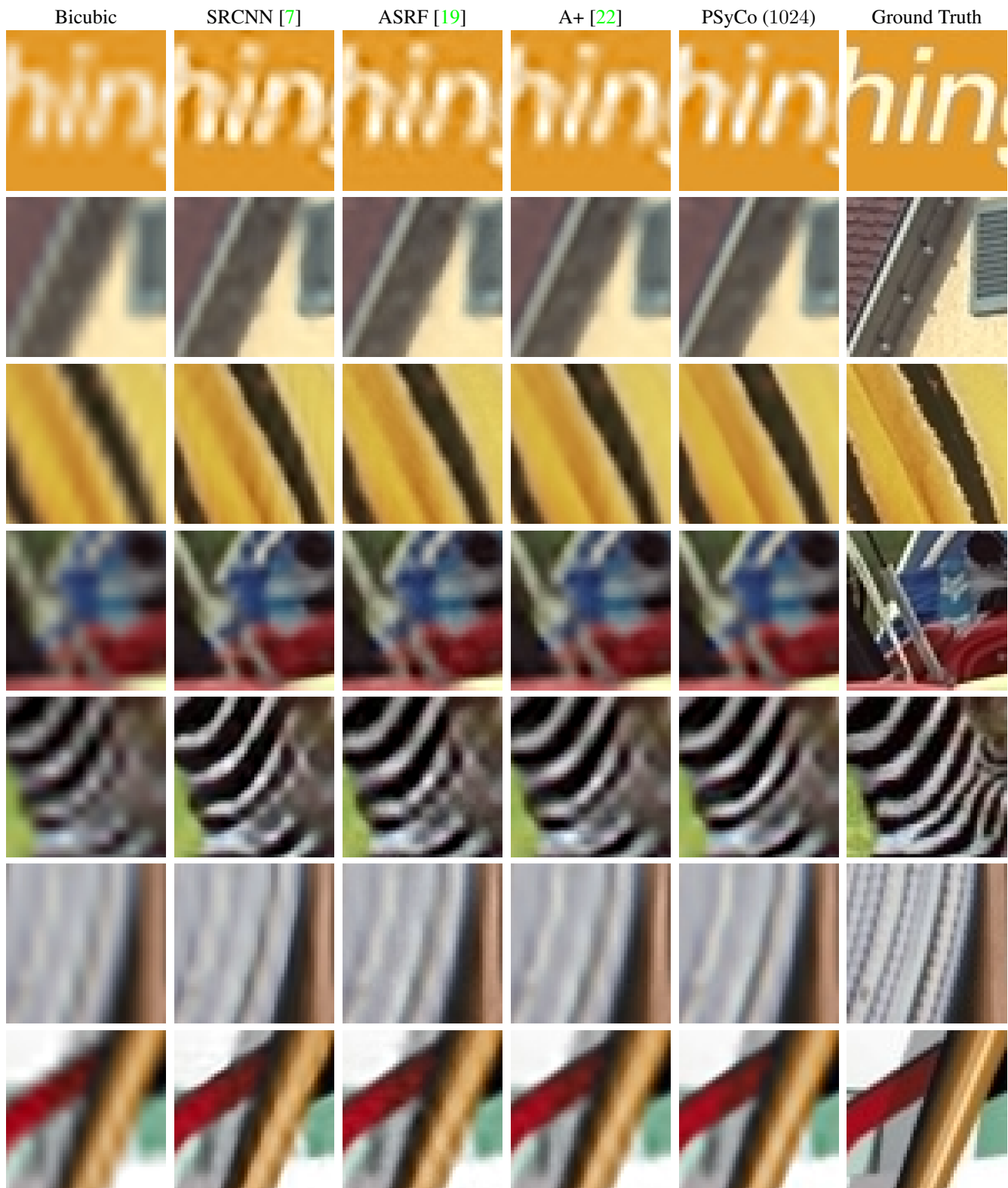| Bicubic | SRCNN [7] | ASRF [19] | A+ [22] | PSyCo (1024) | Ground Truth |

Figure 7: Close-ups of the results for visual qualitative assessment of a $\times 4$ magnification factor from the datasets in the benchmark. Best-viewed zoomed in.

	s	Bicubic			SRCNN [7]			TSelfEx [12]			ASRF [19]			A+ [22]			PSyCo (32)			PSyCo (1024)		
		PSNR	IFC	Time	PSNR	IFC	Time	PSNR	IFC	Time	PSNR	IFC	Time	PSNR	IFC	Time	PSNR	IFC	Time	PSNR	IFC	Time
Set5	2	33.66	6.083	0.002	36.66	8.036	4.722	36.50	7.811	42.521	36.69	8.556	1.278	36.55	8.477	0.684	36.57	8.504	**0.038**	**36.88**	**8.642**	0.185
	3	30.39	3.580	0.002	32.75	4.658	5.226	36.62	4.748	31.008	32.57	4.926	1.026	32.59	4.923	0.401	32.63	4.961	**0.049**	**32.93**	**5.083**	0.456
	4	28.42	2.329	0.002	30.48	2.991	9.962	30.33	3.166	26.728	30.20	3.191	1.071	30.28	3.248	0.226	30.32	3.275	**0.055**	**30.62**	**3.379**	0.210
Set14	2	30.23	6.105	0.002	32.45	7.784	8.204	32.23	7.591	98.645	32.36	8.175	2.134	32.28	8.140	1.421	32.32	8.173	**0.068**	**32.55**	**8.280**	0.346
	3	27.54	3.473	0.002	29.29	4.338	8.098	29.16	4.371	70.176	29.12	4.531	1.674	29.13	4.338	0.828	29.13	4.557	**0.097**	**29.36**	**4.660**	0.350
	4	26.00	2.237	0.002	27.50	2.751	8.305	27.40	2.893	63.873	27.31	2.919	1.386	27.32	2.751	0.614	27.30	2.981	**0.106**	**27.57**	**3.055**	0.334
Kodak	2	30.85	5.711	0.003	32.81	7.149	14.367	32.65	6.782	195.70	32.76	7.387	3.360	32.71	7.380	2.489	32.65	7.419	**0.151**	**32.89**	**7.481**	0.559
	3	28.43	3.214	0.003	29.65	3.895	15.026	29.52	3.806	135.243	29.63	4.025	2.555	29.57	4.053	1.4614	29.57	4.071	**0.172**	**29.74**	**4.136**	0.558
	4	27.23	2.026	0.003	28.17	2.423	14.069	28.14	2.462	115.652	28.17	2.530	2.204	28.10	2.594	1.081	28.07	2.603	**0.201**	**28.28**	**2.663**	0.571
Urban100	2	26.87	6.245	0.007	29.51	7.989	29.358	29.52	7.937	398.307	29.35	8.450	5.878	29.20	8.413	4.857	29.27	8.445	**0.297**	**29.64**	**8.589**	1.050
	3	24.46	3.620	0.005	26.24	4.584	29.439	**26.45**	4.843	286.027	26.00	4.801	4.630	26.03	4.867	3.012	26.04	4.867	**0.348**	26.36	**5.031**	1.104
	4	23.14	2.361	0.004	24.52	2.963	30.233	**24.80**	3.314	238.735	24.28	3.110	3.733	24.32	3.208	2.165	24.35	3.217	**0.365**	24.62	**3.351**	1.095

Table 1: Performance of ×2, ×3 and ×4 magnification in terms of averaged PSNR (dB), IFC and execution time (s) on datasets Set5, Set14, Kodak and Urban100. Best results in bold and runner-up in blue.

the same 91 images provided by Yang et al. in their sparse coding SR [27]. As for SRCNN, we use the network provided by their authors which has been trained with the ImageNet dataset (in the order of hundred thousand images). **Implementation**: For the compared methods we used the code publicly available from the author's website. Our code is a MATLAB + MEX implementation. **Experiments**: We upscale images by the magnification factors ×2, ×3 and ×4 with the authors' recommended configurations and measure PSNR, time and Information Fidelity Criterion (IFC) [20], which has the highest correlation with perceptual scores for SR evaluation [24]. Additionally, for those methods which depend on a dictionary, we test a ×2 upscaling factor on Set14 for several dictionary sizes and measure PSNR and times (see Fig. 1) to compare performances for equal dictionary sizes. **Evaluation**: In Table 1 we show the averaged PSNR, IFC and times of the benchmark. *PSyCo* with 1024 atoms obtains the best PSNR values, around 0.3dB higher across all s and datasets when compared to the most related algorithm A^+. We also outperform the most competitive methods (SRCNN and ASRF) in PSNR by up to 0.3dB. In terms of time, both our configurations are the fastest of the benchmark, specially our proposed (32), which is an order of magnitude faster than any other method. We also note that our methods are trained in less than two hours, which contrasts with the SRCNN method trained with ImageNet. The measured IFC values are consistently the highest among the benchmark, and we highlight the fact that for most s, *PSyCo* with 32 atoms obtains the runner-up IFC, confirming the good performance of our time- and memory-effective configuration. In Fig. 7 we show some quantitative results. We highlight the generally sharper edges and the less proliferation of ringing and aliasing artifacts which results in better preserved structures (e.g. first row in Fig. 7).

7. Conclusions

In this paper we present a new method for regression-based SR that builds around a novel manifold collapsing transform κ. This transform eliminates the undesired variability of the manifold due to the dihedral group of symmetries (i.e. rotation, vertical and horizontal reflections) and the antipodal symmetry (i.e. points that are diametrically opposed in the unitary sphere). Our contributions are: (1) We recommend the use of the dihedral transformation group over more complex projective transformation models. The dihedral group is specially suitable for SR as it is scale invariant and easily invertible. Furthermore, we perform a frequency analysis of the dihedral group in the DCT domain, where the group members are mapped as a combination of transpose and sign changes. (2) We modify the ST of Zabrodsky et al. [28] by skipping point averaging and unfolding, so that the resulting transform collapses the variability of the data while still preserving the original SD. (3) We select a set of projections for which we define symmetry axes corresponding to those of the dihedral and antipodal symmetries. The complexity of our proposed κ is inherently low, as it requires as little as 3 inner products and a matrix re-ordering (i.e. g_j). We exhaustively test our transform applied to SR, and also compare it with other recent state of the art. We consistently obtain ×16 − ×32 smaller dictionaries when aiming at a certain PSNR (see Fig 1). For a fixed dictionary size, we greatly improve in terms of quality both objectively and qualitatively. Our method with 1024 atoms greatly surpasses the state of the art in terms of PSNR and IFC, and with a 32 atoms dictionary we achieve competitive quality while being an order of magnitude faster.

References

[1] M. Aharon, M. Elad, and A. Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. on Signal Processing*, 54(11), 2006. 2

[2] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman. Patchmatch: A randomized correspondence algorithm for structural image editing. *ACM Trans. Graph.*, 28(3), 2009. 3

[3] M. Bevilacqua, A. Roumy, C. Guillemot, and M.-L. Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In *BMVC*, 2012. 3

[4] H. Chang, D.-Y. Yeung, and Y. Xiong. Super-resolution through neighbor embedding. In *CVPR*, 2004. 2

[5] Z. Cui, H. Chang, S. Shan, B. Zhong, and X. Chen. Deep network cascade for image super-resolution. In *ECCV*, 2014. 1

[6] C. Dong, C. Loy, K. He, and X. Tang. Learning a deep convolutional network for image super-resolution. In *ECCV*, 2014. 2, 6

[7] C. Dong, C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. *TPAMI*, PP(99), 2015. 2, 6, 7, 8

[8] R. Eshel and Y. Moses. Homography based multiple camera detection and tracking of people in a dense crowd. In *CVPR*, 2008. 4

[9] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael. Learning low-level vision. *IJCV*, 40(1), 2000. 3

[10] D. Glasner, S. Bagon, and M. Irani. Super-resolution from a single image. In *ICCV*, 2009. 2

[11] M. Grundmann, V. Kwatra, and I. Essa. Auto-directed video stabilization with robust l1 optimal camera paths. In *CVPR*, 2011. 4

[12] J.-B. Huang, A. Singh, and N. Ahuja. Single image super-resolution from transformed self-exemplars. In *CVPR*, 2015. 2, 3, 4, 6, 8

[13] M. Irani and S. Peleg. Improving resolution by image registration. *CVGIP: Graphical Models and Image Processing*, 53(3), 1991. 2

[14] X. Lu, H. Yuan, P. Yan, Y. Yuan, and X. Li. Geometry constrained sparse coding for single image super-resolution. In *CVPR*, 2012. 1

[15] E. Pérez-Pellitero, J. Salvador, J. Ruiz-Hidalgo, and B. Rosenhahn. Antipodally invariant metrics for fast regression-based super-resolution. *TIP*, PP(99), 2016. 3, 5

[16] E. Pérez-Pellitero, J. Salvador, J. Ruiz-Hidalgo, and B. Rosenhahn. Half hypersphere confinement for piecewise linear regression. In *WACV*, 2016. 3

[17] E. Pérez-Pellitero, J. Salvador, I. Torres, J. Ruiz-Hidalgo, and B. Rosenhahn. Fast super-resolution via dense local training and inverse regressor search. In *ACCV*, 2014. 1, 2

[18] J. Salvador and E. Pérez-Pellitero. Naive Bayes Super-Resolution Forest. In *ICCV*, 2015. 1

[19] S. Schulter, C. Leistner, and H. Bischof. Fast and accurate image upscaling with super-resolution forests. In *CVPR*, 2015. 1, 2, 6, 7, 8

[20] H. Sheikh, A. Bovik, and G. de Veciana. An information fidelity criterion for image quality assessment using natural scene statistics. *TIP*, 14(12), 2005. 8

[21] R. Timofte, V. D. Smet, and L. V. Gool. Anchored neighborhood regression for fast example-based super-resolution. In *ICCV*, 2013. 1, 2, 5, 6

[22] R. Timofte, V. D. Smet, and L. V. Gool. A+: Adjusted anchored neighborhood regression for fast super-resolution. In *ACCV*, 2014. 1, 2, 3, 6, 7, 8

[23] A. Yang, S. Rao, K. Huang, W. Hong, and Y. Ma. Geometric segmentation of perspective images based on symmetry groups. In *ICCV*, 2003. 4

[24] C.-Y. Yang, C. Ma, and M.-H. Yang. Single-image super-resolution: A benchmark. In *ECCV*, 2014. 8

[25] C.-Y. Yang and M.-H. Yang. Fast direct super-resolution by simple functions. In *ICCV*, 2013. 1, 2

[26] J. Yang, Z. Lin, and S. Cohen. Fast image super-resolution based on in-place example regression. In *CVPR*, 2013. 1, 2, 6

[27] J. Yang, J. Wright, H. T.S., and Y. Ma. Image super-resolution via sparse representation. *TIP*, 19(11):2861–2873, 2010. 1, 2, 5, 8

[28] H. Zabrodsky, S. Peleg, and D. Avnir. Symmetry as a continuous feature. *TPAMI*, 17(12), Dec 1995. 4, 5, 8

[29] R. Zeyde, M. Elad, and M. Protter. On single image scale-up using sparse-representations. In *Proc. International Conf. on Curves and Surfaces*, 2012. 1, 2

[30] Y. Zhu, Y. Zhang, and A. Yuille. Single image super-resolution using deformable patches. In *CVPR*, 2014. 3