Geometrical Grouplets

Stéphane Mallat
CMAP, Ecole Polytechnique, 91128 Palaiseau Cedex
To appear in Applied and Computational Harmonic Analysis, 2008
18 février 2008

Abstract

Grouplet orthogonal bases and tight frames are constructed with association fields that group points to take advantage of geometrical image regularities in space or time. These association fields have a multiscale geometry that can incorporate multiple junctions. A fast grouplet transform is computed with orthogonal multiscale hierarchical groupings. A grouplet transform applied to wavelet image coefficients defines an orthogonal basis or a tight frame of grouping bandlets. Applications to noise removal and image zooming are shown.

1 Geometric Image Processing

Geometry is at the heart of perception and is required to improve state of the art image processing algorithms. A drawing with few lines gives a powerful perception of shapes and textures. Motion fields provide geometrical information necessary for video processing. Understanding how to represent complex geometrical structures and derive sparse representations is a key issue for image processing. Recent neurophysiological experiments have lead to interesting new models.

The “Gestalt” psychophysic school gives a unifying theory of geometric perception through grouping processes. Neurophysiological studies have shown the existence of “neural integration” processes that may be responsible for these visual geometric groupings. This neural integration seems to first connect simple cells that implement a transformation similar to a wavelet transform in the V1 visual cortex. According to physiological models [16], a simple cell in V1 provides a wavelet coefficient at a given retina location, with a fixed scale and orientation selectivity that depends upon its position in a hypercolumnar structure [1]. These
cells also have non-linear behavior that have been interpreted as gain control feedback loops [11]. Experiments have shown the existence of “horizontal connections” that bind simple cells depending upon the image geometry [25]. The relation between horizontal connections and Gestalt grouping was done by Fields, Hayes and Hesse [8] with “pop out” experiments on images of Gabor patches. They model this horizontal connectivity through an association field between simple cells having a similar orientation selectivity. These association fields seem to be part of a hierarchical computational structure, with neurons in V2 integrating some geometric information provided by V1 simple cells, through feed-forward and feed-back connections. Some of these V2 neurons have been shown [16] to be sensitive to higher level geometrical structures such as the illusory contours of Kanizsa triangles [9]. Several ad-hoc models have been proposed [13] to explain the formation of association fields and the neuronal integration that is performed in V2, but physiological data are not precise enough to validate any precise model.

Wavelet transforms can adapt the processing resolution to the local image regularity but they cannot take advantage of geometrical directional regularity. From a mathematical point of view, building a hierarchical geometrical representation over wavelet coefficients is a natural idea because wavelet coefficients inherit the geometrical regularity of the image [15]. Several new bases have been elaborated to take advantage of the anisotropic regularity of geometrical image structures. Curvelets [3], contourlets [6], bandlets [14], wedgelets [7] are examples among others. Asymptotic results have been established for “cartoon-like” image models, defined as piecewise regular with edges along piecewise regular curves. For such images, these new bases provide more efficient asymptotic approximation, compression and denoising results than wavelet bases. Numerical improvements have also been demonstrated on real images. However, these improvements are not as spectacular as one may expect from the asymptotic theorems. This may be due to the relatively small portion of regular geometrical curves in real images, whereas textures often have a more complex geometry not efficiently represented by these bases.

For video compression and geometric image processing, more flexible directional multiscale decompositions have been developed with lifting schemes. A lifting is computed over embedded grids by associating each point to a neighbor point in a direction of maximum regularity. For video, the association field is derived from a computed motion between two images, and the lifting is calculated along a “motion threads” defined by the associations of pixels from one image to the next [23, 17]. In space, edge adaptive lifting schemes associate coefficients along the local orientation of edges, often calculated with a gradient operators [10, 12] or with block based optimisations [4]. Directional liftings have also been implemented over wavelet coefficients [5]. Adaptive directional lifting schemes satisfy perfect reconstruction properties, but they are not orthogonal and their stability is often
not controlled.

This paper introduces grouplet orthogonal bases and tight frames that build a stable geometrical image representation with an orthogonal weighted Haar lifting. In the spirit of the “Gestalt” psychophysical school, the geometry is constructed with grouping processes which define multiscale association fields. Grouplet orthogonal bases are constructed in Section 2 with these multiscale association fields. Section 2.2 describes more flexible grouplet tight frames that do not require embedded grid structures, with causal association fields. Noise removal and super-resolution image zooming algorithms are obtained by thresholding grouplet coefficients in Section 2.3. Section 3 introduces a hierarchical geometrical image representation by applying multiscale grouplet transforms over wavelet coefficients, which yields grouping bundled orthogonal bases and tight frames. Applications are also described to remove noise from images and restore fine geometrical textures.

2 Multiscale Groupings with Association Fields

2.1 Orthogonal Grouplet Bases

An orthogonal multiscale grouping is implemented with a weighted Haar lifting applied successively to points that are grouped by an association field. The Haar transform is reviewed and then modified to define a grouplet transform.

**Haar transform** An average signal \( a[n] \) is initially set to be equal to the input signal \( f[n] : a[n] = f[n] \) for \( 0 \leq n < N \). Then, for a scale \( 2^i \) that increases from \( 2^i \) to \( 2^{i+1} \), and for all \( 0 \leq n < 2^{i+1} \), the Haar transform groups consecutive average coefficients \( a[2n] \) and \( a[2n+1] \), and it computes a next scale average

\[
a[n] = \frac{a[2n] + a[2n+1]}{2},
\]

(1)

which together with a normalized difference:

\[
d_j[n] = \left( a[2n+1] - a[2n] \right) \sqrt{\frac{2^{j-1}}{2} \cdot 2^{j-1} \cdot 2^{j-1}}, \tag{2}
\]

The resulting coefficient \( a[n] \) at a scale \( 2^j \) is the average of the signal values \( f[p] \) for \( p \in [2^j n, 2^j n + 2^{j-1} - 1] \) whereas the difference \( d_j[n] \) is proportional to the difference between the averages of \( f[p] \) on \( [2^j n, 2^j n + 2^{j-1} - 1] \) and on \( [2^{j-1} n, 2^{j-1} n + 2^{j-1} - 1] \). At the largest scale \( 2^j \), the average coefficients are renormalized \( a_j[n] = 2^{j/2} a[n] \) to be an inner product between \( f \) and a vector of unit norm. The set of coefficients \( \{d_j[n], a_j[n]\}_{1 \leq j \leq J, n} \) are inner products of \( f \) with vectors of a Haar orthonormal basis.
**Association Fields on Embedded Grids**  An orthogonal grouplet transform modifies the Haar transform with an association field that groups together points that are not necessarily neighbors. Let us consider a
\( \tilde{G}_{j+1} \). This defines column embedded subgrids. Fig. 1 illustrates an association field defined with column embedded subgrids. If the image has geometrically regular structures that rather propagate vertically, then \( G_j \) is decomposed into even and odd rows corresponding respectively to \( G_{j+1} \) and \( G_{j+1} \). This defines rows embedded subgrids. When the image has no preferential direction, one may choose more isotropic subgrids, such as quincunx subgrids.

Although this paper mostly concentrates on geometrical structures of two-dimensional images, association fields and their corresponding grouplet bases may be defined in any dimension. For videos, \( G_j \) is a three dimensional grid which is a time sequence of two-dimensional image sampling grids. At each scale \( 2^j \), time embedded subgrids can decompose \( G_j \) into even time images \( G_{j+1} \) and odd time images \( G_{j+1} \). An association field \( A_j(\tilde{m}) \) typically stores a time displacement from \( \tilde{m} \in G_j \) to the previous image point \( m \in G_j \). Other space-time subsampling grids may be chosen to potentially take advantage simultaneously of spatial and time geometric regularities. However, in more than two dimensions, no embedded subgrid scheme characterized by a subsampling matrix with reduction factor 2 can have an isotropic behavior similar to a two-dimensional quincunx grid [24].

**Orthogonal Grouplets** A grouplet transform is a weighted multiscale Haar transform. The average signal is also initialized to be the input signal \( a[n] = f[n] \) for all \( n \in G_j \). A new array \( s[n] \) is created to store the support size of the averaging kernel which computes \( a[n] \) from \( f \). At the initialization, since \( a[n] = f[n] \) it results that \( s[n] = 1 \).

For scales \( 2^j \) that increase from 1 to \( 2^j \), the successive groupings are performed in a predefined order over all points in \( G_j \). Let \( N_j \) be the number of samples of \( G_j \). Let \( \alpha_j \) be an invertible mapping between \( 1 \leq n \leq N_j \) and points \( \tilde{m} = \alpha_j(n) \in G_j \). For \( n \) going from 1 to \( N_j \), each \( \tilde{m} = \alpha_j(n) \) is associated to a point \( m = \tilde{m} + A \).
These values are stored “in place” at \( m \in G_j \): \( a[m] = \hat{a} \) and \( s[m] = \hat{s} \).

The flexibility of the grouping process can take advantage of various geometric regularity including periodic patterns and junctions. At junction points, several points may be grouped with a single one, with successive orthogonalizations, as illustrated by Fig. 1.

At the largest scale \( 2^j = 2^j \), average coefficients are normalized by setting

\[
\forall m \in G_j \quad a_j[m] = a[m] \sqrt{s[m]}.
\]

The grouplet transform associates to a signal \( f[n] \) of size \( N \) the family of \( N \) grouplet coefficients \( \{d_j[m], a[m]\}_{1 \leq j \leq J, \hat{m} \in \mathbb{G}, m \in \mathbb{G}} \). The grouplet representation includes not only these coefficients but also the \( N(1 - 2^{-j}) \) multiscale association field coefficients \( \{A_j[\hat{m}]\}_{1 \leq j \leq J, \hat{m} \in \mathbb{G}} \). For each \( \hat{m} \in \mathbb{G} \), (3), (4) and (5) is computed with \( O(1) \) operations. Like in a Haar transform, the \( N \) grouplet coefficients are thus computed with \( O(N) \) operations. The over
Let us first prove that the supports of $\tilde{h}_{k,m}[n]$ and $\tilde{h}_{k,\tilde{m}}[n]$ do not intersect. Observe that the support of $\tilde{h}_{k,m}$ is the union of all points aggregated by the association fields $A_l$ for $l \leq j$. This grouping defines a tree over the support of $\tilde{h}_{k,m}$ whose root is $m$. Indeed, an association field $A_l$ groups a point in $G$ to a point in $G$ which thus belongs to a grid $G_k$ for some $k > l$ or to $G_l$. Since the grouping goes from a finer to a larger scale grid, there cannot be any cycle and the root is the point $m$ at the largest scale.

Similarly the support of $\tilde{h}_{k,m}$ has a tree structure rooted at $\tilde{m}$. We also know that $\tilde{m}$ and $m$ do not belong to the support respectively of $\tilde{h}_{k,m}$ and $\tilde{h}_{k,\tilde{m}}$. Indeed, $m$ cannot belong to the support of $\tilde{h}_{k,\tilde{m}}$ since $m$ is in a coarser grid than $\tilde{m}$, and $\tilde{m}$ cannot belong to the support of $\tilde{h}_{k,m}$ because $\tilde{m}$ has not yet been grouped to a coarser scale point by the aggregation process.

Since the supports of $\tilde{h}_{k,m}$ and $\tilde{h}_{k,\tilde{m}}$ are trees and none of their roots belong to the other tree, necessarily these trees are disjoint which means that these supports do not intersect.

Let us now prove by induction on $k$ that $\tilde{h}_{k,m}[n] = 1/s[m]$ over a support of size $s[m]$. For $k = 0$ this property is valid since $\tilde{h}_{0,m}[n] = \delta[n - m]$ and $s[m] = 1$. Suppose that the induction hypothesis is valid for $k$. It implies that $\tilde{h}_{k,m} = 1/s[m]$ and $\tilde{h}_{k,\tilde{m}} = 1/s[\tilde{m}]$ over their support of size respectively $s[m]$ and $s[\tilde{m}]$. Since $\tilde{h}_{k,m}$ and $\tilde{h}_{k,\tilde{m}}$ have disjoint supports, it results from (9) that $\tilde{h}_{k+1,m} = 1/(s[m] + s[\tilde{m}])$ over its support of size $s[m] + s[\tilde{m}]$. The updating formula (5) with $s[m] = s$ thus verifies that the induction hypothesis is valid for $k + 1$. Incorporating the normalization (6) at the coarsest scale finishes the proof of (i).

To prove (ii), as a result of (3) we can write
\[
    g_{j,\tilde{m}} = (\tilde{h}_{k,\tilde{m}} - \tilde{h}_{k,m}) \frac{s[m]s[\tilde{m}]}{\sqrt{s[m] + s[\tilde{m}]}}.
\]  
(10)

Since $\tilde{h}_{k,\tilde{m}} = 1/s[\tilde{m}]$ and $\tilde{h}_{k,m} = 1/s[m]$ over their support of size respectively $s[\tilde{m}]$ and $s[m]$, and their support are disjoint, it results that $g_{j,\tilde{m}}$ takes only two different values over its support. Moreover we verify that $\sum n g_{j,\tilde{m}}[n] = 0$ and hence that $g_{j,\tilde{m}}$ has 1 vanishing moment. □

Theorem 2.1 proves that the support of grouplet vectors have a tree structure with a support size equal to $s[m]$. Junctions correspond to points where grouplet supports have several “sons” defined by multiple groupings with a single junction point. The following theorem proves that grouplets define an orthonormal basis.

**Theorem 2.2.** The grouplet family
\[
    \left\{ g_{j,\tilde{m}}[n] , h_{j,m}[n] \right\}_{1 \leq j \leq J, \tilde{m} \in \mathcal{G}_j, m \in \mathcal{G}_j}
\]

is an orthonormal basis of $L^2(G_i)$. 

7
Proof.

Lemma 2.1. For any \( a, \tilde{a}, s > 0 \) and \( \tilde{s} > 0 \), we define

\[
\tilde{s} = s + \tilde{s}, \quad d = (a - \tilde{a}) \sqrt{\frac{s \tilde{s}}{s}} \quad \text{and} \quad \hat{a} = \frac{s a + \tilde{s} \tilde{a}}{\tilde{s}}.
\] (11)

The operator that transforms \((a \sqrt{s}, \tilde{a} \sqrt{\tilde{s}})\) into \((d, \hat{a} \sqrt{\tilde{s}})\) is a unitary operator.

This lemma is proved by observing that the two rows of the \(2 \times 2\) matrix of this operator are \((s, \tilde{s})/s + \tilde{s}\) and \((s, -s)/s + \tilde{s}\), which are two orthonormal vectors. It is therefore a unitary operator.

Lemma 2.1 proves that the operator which transforms \((a[m], \sqrt{s[m]}, a[\tilde{m}], \sqrt{s[\tilde{m}]}]\) into \((\hat{a}, \sqrt{\tilde{s}}, d_j[\tilde{m}])\) as defined by (3), (4) and (5) is a unitary operator. Let \(a_j[m] = a[m] \sqrt{s[m]}\) be the normalized average array at the end of the loop over the index \(n\) that goes from 1 to \(N_j\), for a given \(j\). The operator which transforms \(\{a_{j-1}[m]\}_{m \in G_{j-1}}\) into \(\{a_j[m], d_j[\tilde{m}]\}_{m \in G_j, \tilde{m} \in \tilde{G}_j}\) is therefore a cascade of \(N_j\) unitary operators and is thus a unitary operator. At the initialization \(a_0[m] = a[m] \sqrt{s[m]} = f[m]\) for all \(m \in G_0\). It results that iterating over the variable \(j\) defines a cascade of unitary operators that transforms \(f[m]\) in to the grouplet coefficients

\[
\left\{d_j[\tilde{m}], a_j[m]\right\}_{1 \leq j \leq J, \tilde{m} \in \tilde{G}_j, m \in G_j}.
\]

As a consequence, the grouplet vectors defined by (7) and (8) define an orthonormal basis.

Inverse Orthogonal Grouplet Transform Since grouplets define an orthonormal basis, the input signal can be reconstructed from its grouplet coefficients with:

\[
f[n] = \sum_{j=1}^{J} \sum_{m \in G_j} d_j[\tilde{m}] g_{j, \tilde{m}}[n] + \sum_{m \in G_j} a_j[m] h_{J,m}[n].
\] (12)

Similarly to the fast inverse Haar transform, the fast inverse grouplet transform inverts each grouping operator in the reverse order of the forward transform, for a scale \(2^J\) that decreases from \(2^J\) to 1.

The averaging size array is recalculated at the coarsest scale \(2^J\) from the association fields. We initialize \(s[m] = 1\) for all \(m \in G_0\) and for \(j\) going from 1 to \(J\) and for all \(\tilde{m} \in \tilde{G}_j\):

\[
\tilde{s} = s[m] + s[\tilde{m}] \quad \text{and} \quad s[m] = \tilde{s} \quad \text{for} \quad m = \tilde{m} + A_j[\tilde{m}].
\]
At the coarsest scale, $2^J$, the average coefficient normalization (6) is then inverted

$$\forall m \in G_J \ , \ a[m] = \frac{a_J[m]}{\sqrt{s[m]}} .$$

(13)

Each grouping transform is then inverted at each scale in their reverse order. For $j$ going from $J$ to 1 and for $n$ going from $N_j$ to 1 let $\tilde{m} = \alpha_j(n)$ and $m = \tilde{m} + A_j[\tilde{m}]$. To invert the grouping transform, the averaging size is updated by inverting (5)

$$\tilde{s} = s[m] - s[\tilde{m}]$$

(14)

and the finer scale average coefficients are computed from the larger scale average coefficient and the difference coefficient by inverting the unitary transform defined by (3) and (4):

$$a[\tilde{m}] = a[m] + d_j[\tilde{m}] \sqrt{s[\tilde{m}]/s[m]}$$

(15)

$$\tilde{a} = a[m] - d_j[\tilde{m}] \sqrt{s[\tilde{m}]/s[m]} .$$

(16)

These reconstructed values are stored “in place” $a[m] = \tilde{a}$ and $s[m] = \tilde{s}$. At the end of this double loop over all groupings at all scales, this inversion reconstructs the original signal $a[m] = f[m]$ for all $m \in G_s$.

For each $\tilde{m} \in G_j$ the reconstruction operations (14), (15) and (16) requires $O(1)$ operations. The $N$ signal coefficients $f[m]$ are thus recovered with $O(N)$ operations.

**Multiscale Association Fields and Best Basis**  There are as many grouplet orthonormal bases as possible groupings of signal samples. Computing multiscale association fields is thus equivalent to choose a “best” grouplet basis to decompose a particular image. The representation does not only include the grouplet coefficients $d_j[m]$ but also the association fields $A_j[m]$. For image compression applications, it is necessary to code both the grouplet coefficients and the association fields, with as few bits as possible. Few bits are needed to encode the difference coefficients $d_j[\tilde{m}]$ if few of these coefficients are non-zero after quantization. This means that each $\tilde{m} \in G_j$ should be associated to a point $m = \tilde{m} + A_j[\tilde{m}]$ such that $|a[m] - a[\tilde{m}]|$ is small. The ability to code efficiently an association field depends upon its regularity. If it is highly regular then it can be approximated with few coefficients that can be encoded for example in a wavelet or polynomial spline basis. We thus need to compute regular association fields that produce sparse grouplet coefficient sequences.

Grouplet orthonormal bases have a flexible geometry but the resulting dictionary is more complex than the “tree-structured” dictionaries used for wedglets [7]
or bandlets [14]. As a consequence, optimizing the multiscale association fields to minimize the distortion-rate of a grouplet signal compression is a more complex problem, that is beyond the scope of this paper. A finer mathematical analysis of grouplet dictionary properties for video processing can be found in [19]. We describe below a simple block matching algorithm that is used in numerical experiments.

Block matching algorithms are often used to compute displacements, for motion estimation in video. For each $\hat{m} \in \hat{G}_j$, the search for a best match is performed in a neighborhood of $\hat{m}$ of points in $G_j$, that we write $N_j(\hat{m})$. A block matching computes for all $p$ the $1^1$ or $1^2$ distance between a block of points $p$ of $j$ around $\hat{m}$ and the corresponding block around $m$, and it finds the
Fig. 2 – Possible matches for \( \tilde{m} \in \mathcal{G}_j \) in a neighborhood \( \mathcal{N}_a(\tilde{m}) \subset \mathcal{G}_j \) that is an interval in the previous column in (a) and a wider set of points covering more directions in (b).

Block matching computes regular multiscale association fields with directions along which the image is smoothly varying.

Fig. 3 – Grouping of an association field at scales \( 2^1 \), \( 2^2 \) and \( 2^3 \) computed by block matching for a seismic image shown in transparency.

Fig. 4(b) shows orthogonal grouplet coefficients over \( J = 6 \) scales, computed from the multiscale association fields displayed in Fig. 3. Black, gray and white coefficients correspond to negative, zero and positive coefficients with the same dynamic range as in the original seismic image in Fig. 4(a). There are mostly gray coefficients which means that most “difference” coefficients of the multiscale grouping orthogonal transform are nearly zero. Fig. 4(c) shows example of grouplet vectors \( g_{j,m}[n] \) for several scales \( 2^j \). These grouplet vectors have a support which is elongated along the association field, with possible junctions and a tree structure.

Fig. 4(d) gives another example of orthogonal grouplet coefficients computed on Lena’s hat, with \( J = 6 \) scales. The grouplet representation is clearly much more
sparse than the original image.

Fig. 4 – (a,d) : Original images. (b,e) : Orthogonal grouplet coefficients over 6 scales, displayed with the same dynamic range as the original image. (c) Examples of grouplet vectors \( g_{j,\tilde{m}} \).

Matching Pursuit If the multiscale association field \( A_j[\tilde{m}] \) is computed together with the grouplet decomposition the resulting algorithm is similar to an orthogonal matching pursuit. In the algorithm previously described, for each scale \( 2^j \), the iteration begins by a computation of the association field \( A_j[\tilde{m}] \) for all \( \tilde{m} \in \mathcal{G}_j \) followed by a computation of all difference coefficients \( d_j[\tilde{m}] \) for all \( \tilde{m} \in \mathcal{G}_j \). A second possibility is to compute for each \( \tilde{m} = \alpha_j(n) \) the value of \( A_j[\tilde{m}] \), then calculate \( d_j[\tilde{m}] \) and update \( a[m] \) for \( m = \tilde{m} + A_j[\tilde{m}] \), before computing the next association field value for \( \tilde{m} = \alpha_j(n + 1) \). This reordering of calculations does not change the overall computational complexity. This calculation is similar to a matching pursuit decomposition [20] where each vector is sequentially computed by a greedy optimization which finds a best match. In this case, the pursuit is orthogonal with no further orthogonalization required. It is computed in a fine to coarse order, with fine scale vectors first and coarse scale vectors afterward. This
pursuit is performed in the dictionary of all possible grouplets corresponding to all possible association fields.

2.2 Grouplet Tight Frames

Redundant grouplet tight frames can be constructed without an embedded subgrid structure. More flexible multiscale association fields are defined with a causality property which is sufficient to construct grouplets having an energy conservation as in the orthogonal case. These association fields have more flexibility as in the models [8] of Fields, Hayes and Hesse [8] for horizontal connections between simple cells in the visual cortex region V1. The redundancy of these grouplet tight frames also improves noise removal with thresholding estimators.

Causal Multiscale Association Fields To avoid grouping a point \( \hat{m} \) with \( m \) and then group \( m \) with \( \hat{m} \), a causality is imposed on the association field, with a partial ordering of all points in the signal sampling grid \( G_0 \). An association field \( A_j \) is said to be causal if it groups each \( \hat{m} \in G_0 \) with a point \( m = \hat{m} + A_j[\hat{m}] \in G_0 \) that is located before \( \hat{m} \) in the sense of this partial ordering. If there is no point before \( \hat{m} \) then we set \( A_j[\hat{m}] = \text{NULL} \).

A causal multiscale association field must also associate points whose distance increase with the scale \( 2^j \), to reach long range groupings for large scales. For a given pseudo distance \( d(x, y) \) defined over points in \( \mathbb{R}^2 \), we impose that for any scale \( 2^j \) if \( \hat{m} + A_j[\hat{m}] = m \) then \( 2^{j-1} \leq d(m, \hat{m}) < 2^j \). A pseudo distance satisfies the symmetry and triangular inequality of a distance but one can have \( d(x, y) = 0 \) with \( x \neq y \). We say that a causal association field is strictly multiscale if for any scale \( 2^j \) and any \( m = A_j[\hat{m}] \) we have \( d(m, \hat{m}) = 2^{j-1} \).

The partial ordering and the pseudo distance can be adapted to the type of geometrical regularity that must be captured by the grouping. For motion in a video frame sequence, at each scale \( 2^j \), a time causal strictly multiscale association field can associate an image point \( \hat{m} \) to a point \( m \) located in a previous image at distance of \( 2^{j-1} \) frames before. This association can be interpreted as a displacement over \( 2^{j-1} \) time steps. In this case, the partial ordering of image points is defined by the corresponding time frame location and the pseudo distance by the number of time steps between frames.

For spatial
adapted to perform a grouplet transform of wavelet coeffi
\[ p \text{ and the corresponding block around } \hat{m}, \text{ and it finds the } p = m \text{ which minimizes this distance:} \]
\[ m = \arg \min_{p \in N(\hat{m})} \sum_{n \in B(p)} \left| a[n] - a[n + \hat{m} - p] \right|^k \text{ with } k = 1 \text{ or } k = 2. \tag{20} \]

We set \( A_j[\hat{m}] = m - \hat{m} \). This process is illustrated by Fig. 5. The association field \( A_j \) is thus derived from \( A_{j-1} \) plus an adjustment whose amplitude depends upon the size \( K \) of the neighborhoods \( N(\hat{m}) \), which is typically small. This adjustment is necessary to avoid accumulations of errors when following long range geometric structures. To compute the finest scale association field \( A_1 \), we initialize \( A_0[\hat{m}] = 0 \) for all \( \hat{m} \in G_0 \). For a signal of size \( N \), this block matching compute the \( JN \) values of the \( J \) association fields \( A_j \) with \( O(JNKP) \) operations. If the multiscale association field adjustment is suppressed then all association fields are derived from the finest scale association field \( A_1 \) and the computational complexity becomes \( O(NKP + JN) \).

Fig. 6 shows a strictly multiscale causal association field over two scales calculated on a zoom of Lena’s hat. A partial ordering of angle \( \theta = 0 \) is used, which means that coefficients are ordered columns by columns. A pseudo distance of angle \( \theta = 0 \), measuring the horizontal distance, is used by the grouping. At a scale \( 2^j \), the grouping is performed between points whose columns have a distance of \( 2^{j-1} \).

**Tight Frame Grouplet Transform** A grouplet frame transforms a signal \( f[n] \) into multiscale difference coefficients and coarse scale average coefficients defined on the original signal grid \( G_0 \). Similarly to orthogonal grouplet coefficients, these coefficients are computed with successive grouping transformations according to multiscale association fields. All points of \( G_0 \) are visited in an order coherent with the grouplet partial ordering.

Suppose that \( G_0 \) includes \( N \) samples. One can define an ordering function \( \alpha(n) \) that is compatible with the grouplet partial ordering. It is an invertible mapping that associates to any integer \( 1 \leq n \leq N \) a point \( \alpha(n) \in G_0 \) such that if \( m = \alpha(n) \) is before \( p = \alpha(u) \) with respect to the grouplet partial ordering, then \( n < u \).

In a video time sequence, for a time partial ordering an ordering function sorts all image pixels in the time order, and within each image the pixels can be sorted in any spatial order. In a rectangular image with \( M_1 \) columns and \( M_2 \) rows, for a partial ordering of angle 0 (horizontal), an ordering function can be defined by \( (m_1, m_2) = \alpha(m_2 + m_1M_2) \), and for a partial ordering of angle \( \pi/2 \) (vertical) the ordering function can be \( (m_1, m_2) = \alpha(m_1 + m_2M_1) \).

For \( j = 0 \) and all \( m \in G_0 \), the grouplet frame transform initializes \( a[m] = f[m] \) with an averaging size \( s[m] = 1 \). For \( j \) going from 1 to \( J \), and for all \( n \) going from
These values are stored “in place” : \( a[m] = \tilde{a} \) and \( s[m] = \tilde{s} \). If \( A_j[\tilde{m}] = \text{NULL} \) then we perform a transform as if the associated point was 0 and set \( d_j[\tilde{m}] = a[\tilde{m}] \sqrt{s[\tilde{m}]} \).

At the coarsest scale \( 2^J \), as in the orthogonal grouplet transform, the average coefficients are normalized

\[
\forall m \in G_0, \quad a_J[m] = a[m] \sqrt{s[m]}.
\]  

(24)

A redundant grouplet transform is thus an operator \( G \) that associates to a signal \( f[n] \) of size \( N \) a family of \( J+1 \) grouplet coefficient arrays \( \{ d_j[m], a_j[m] \}_{1 \leq j \leq J, m \in \mathbb{N}} \) together with \( J \) multi-scale association field arrays \( \{ A_j[m] \}_{1 \leq j \leq J, m \in \mathbb{N}} \). For a signal of size \( N \), (21), (22) and (23) compute the \((J + 1)N\) grouplet coefficients with \( O(NJ) \) operations.

Fig. 6 shows the grouplet tight frame coefficients of a textured image computed from a multiscale association field at scales \( 2^1 \) and \( 2^2 \). Most multiscale difference coefficients have a much smaller dynamic range than the original image coefficients.

Each \( d_j[m] \) and \( a_j[m] \), for \( 1 \leq j \leq J \) and \( m \in G_0 \), is calculated with a linear operator and can thus be written as an inner product with a grouplet vector :

\[
d_j[m] = \langle f, g_{j,m} \rangle \quad \text{and} \quad a_j[m] = \langle f, h_{j,m} \rangle.
\]

The family of grouplets \( \{ g_{j,m} \}_{1 \leq j \leq J, m \in \mathbb{N}} \cup \{ h_{j,m} \}_{m \in \mathbb{N}} \) is entirely specified by the causal association fields \( \{ A_j \}_{1 \leq j \leq J} \). The following theorem characterizes the support of these grouplets and Theorem 2.4 proves that this family defines a tight frame.

**Theorem 2.3.** For any \( m \in G_0 \), \( g_{j,m}[n] \) has 1 vanishing moment. If the causal association fields are strictly multiscale then

- (i) \( h_{j,m}[n] = 1/\sqrt{s[m]} \) over its support whose size is \( s[m] \). Support points are before \( m \) at a distance strictly smaller than \( 2^J \).
- (ii) \( g_{j,m}[n] \) takes a constant positive value on support points before \( n \) and a constant negative value on support points after \( n \). The distance between \( m \) and all support points is smaller than \( 2^{J-1} \).

**Proof.** If for any \( n \in G_0 \), \( f[n] = c \) then one can derive from (21) and (23) that \( a[m] = c \) for all \( m \in G_0 \). It results from (22) that for all \( m \) and \( j \), \( d_j[m] = 0 \) and hence that \( g_{j,m} \) has 1 vanishing moment.

Let us now suppose that the causal association field is strictly multiscale, which means that if \( A_j[\tilde{m}] = m \) then \( d(m, \tilde{m}) = 2^{j-1} \). After applying \( k \) grouping operators, each \( a[m] \) is the result of a linear operator applied to \( f[n] \) and hence can be written \( a[m] = (\tilde{h}_{k,m}, f) \). Let \( m = \tilde{m} + A_j[\tilde{m}] \in G_{j+1} \) be the point grouped with \( \tilde{m} \in G_{j+1} \) by the \( k+1 \) unitary operator. As a result of (23) we have

\[
\tilde{h}_{k+1,m}[n] = \frac{s[m] \tilde{h}_{k,m}[n] + s[\tilde{m}] \tilde{h}_{k,\tilde{m}}[n]}{s[m] + s[\tilde{m}]}.
\]  

(25)
Let us first prove that the supports of $h_{k,m}[n]$ and $\tilde{h}_{k,m}[n]$ do not intersect. Each $h_{k,m}$ has a tree structure rooted at $m$ because of the causality of the association field. Indeed each point of its support is related by the causal association field that groups each point to a unique parent that is before. The root $m$ is thus before all other support points. Since the association field is strictly multiscale, The maximum distance between a support point and $m$ is bounded by adding distance between consecutive points that are grouped. At a scale $2^j \leq j$ the distance is $2^{l-1}$ because the association fie
Proof. Let \( a_j[m] = a[m] \sqrt{s[m]} \) be the normalized average array at the end of the loop over the index \( n \) that goes from \( 1 \) to \( N \), for a given \( j \). We first prove the following Lemma.

**Lemma 2.2.** For any \( j > 1 \)

\[
2 \sum_{m \in \mathbb{G}_i} |a_{j-1}[m]|^2 = \sum_{m \in \mathbb{G}_i} |a_j[m]|^2 + \sum_{m \in \mathbb{G}_i} |d_j[m]|^2. \tag{33}
\]

To prove this lemma, observe that the grouplet operator can be decomposed as a product of isometric operators. At a fixed scale \( 2^j \), the grouplet transform begins with a duplication operator that transforms \( a_{j-1}[m] \) into a pair of signals

\[
(a[m] \sqrt{s[m]} = a_{j-1}[m], d[m] = a_{j-1}[m]) \quad \text{for all } m, \hat{m} \in \mathbb{G}_i.
\]

Then the grouplet iteration over \( n \) from \( 1 \) to \( N \) transforms \( \{a[m] \sqrt{s[m]}, d[m]\}_{m, \hat{m} \in \mathbb{G}_i} \) with a cascade of unitary operators. For each \( n \), if \( A[\hat{m}] = \text{NULL} \) then the operator is the identity. If \( A[\hat{m}] \neq \text{NULL} \) then Lemma 2.1 proves that the operator which transforms \( \{a[m] \sqrt{s[m]}, d[m]\} = a[\hat{m}] \sqrt{s[\hat{m}]} \) into \( (\hat{a}, \hat{d}[\hat{m}] = d_j[\hat{m}] \) as defined by (21), (22) and (23) is indeed a unitary operator. It results that \( \{a[m] \sqrt{s[m]}, d[m]\}_{m, \hat{m} \in \mathbb{G}_i} \) is transformed from its initial values \( \{a_{j-1}[m], a_{j-1}[\hat{m}\} \}_{m, \hat{m} \in \mathbb{G}_i} \) into the final values \( \{a_j[m], d_j[\hat{m}]\}_{m \in \mathbb{G}_i, \hat{m} \in \mathbb{G}_i} \) by a cascade of unitary operators, which is therefore a unitary operator. It implies that norm of these pairs of signals is conserved which proves (33).

Since \( a_0[n] = f[n], (f, g_{1,m}) = d_j[m] \) and \( (f, h_{1,m}) = a_j[m] \), the theorem energy conservation (32) is proved with a substitution on (33) for \( j \) going from \( 1 \) to \( J \).

The grouplet transform is obtained with a loop over operators at increasing scales \( 2^j \). A left inverse is obtained with a product of the left inverse of each elementary operator in the reverse order. The pseudo inverse is the product of the pseudo-inverses of each elementary operator. We saw that at each scale \( 2^j \) the grouplet transform begins with a duplication operator followed by a cascade of unitary operators. The inverse of each unitary operator is also unitary and is therefore the pseudo inverse. The duplication operator transforms an array \( x[m] \) into a pair \( (x[m], x[m]) \) for all \( m \in \mathbb{G}_i \). Its pseudo-inverse is the averaging that transforms a pair of arrays \( (x[m], y[m]) \) into \( (x[m] + y[m])/2 \) for all \( m \in \mathbb{G}_i \). For each \( j \), the inverse grouplet algorithm is equivalent to this cascade of inverse unitary operators followed by this averaging. Instead of computing the averaging at the end of the loop over \( n \) as in the proof, the reconstruction algorithm computes immediately the averaging (31) for each pair of coefficients. The result is the same because the coefficient \( a[\hat{m}] \) will not be modified anymore by the continuing cascade of inverse operators, which is guaranteed by the causality of the association fields. The averaging is computed immediately because it requires less memory storage. \( \square \)
General frame theory [18] proves that the pseudo-inverse implements a reconstruction within the dual frame. Theorem 2.4 proves that grouplets define a tight frame and hence that the dual frame is equal to the original frame. The pseudo-inverse algorithm thus computes

$$f[n] = \sum_{j=1}^{J} \frac{1}{2^j} \sum_{m \in \mathcal{G}_j} d_j[m] g_{j,m}[n] + \frac{1}{2^j} \sum_{m \in \mathcal{G}_j} a_j[m] h_{j,m}[n].$$  

(34)

2.3 Noise Removal and Image Zooming

A signal contaminated $f[n]$ by a Gaussian white noise $W[n]$ can be estimated by thresholding the coefficients of $X[n] = f[n] + W[n]$ in an orthogonal basis or in a frame. With grouplets, this amounts to a conditional averaging of signal coefficients which have been grouped by the multiscale association fields. Image zooming by geometric interpolation and regularization is a second application that is described. Missing pixels are estimated with adaptive interpolation in directions in which the image is regular.

Orthogonal Grouplets  The efficiency of denoising with a thresholding estimator in a basis depends upon the sparsity of the signal within the basis. Grouplets define orthogonal bases and the optimization of the association fields amounts to finding a “best” basis. Ideally one would like to find the “best” basis in which the estimation by thresholding produces a minimum risk. Estimating such a best basis is beyond the scope of this paper but is studied for video denoising in [19].

The grouplet denoising approach described in the following gives numerical illustrations of grouplet properties, with no optimality result. The association field is computed on the noisy image $X[n]$ with the block matching optimization (17). For a Gaussian white noise $W[n]$ of variance $\sigma^2$, the blocks are square windows of width proportional to $\sigma$ so that the averaging regularization compensates the noise variance. The orthogonal grouplet coefficients of $X[n]$ below a threshold $T$ are set to zero and a signal estimate $\hat{F}[n]$ is reconstructed:

$$\hat{F}[n] = \sum_{j=1}^{J} \sum_{m \in \mathcal{G}_j} \rho_T((X, g_{j,m})) g_{j,m}[n] + \sum_{m \in \mathcal{G}_j} \rho_T((X, h_{j,m})) h_{j,m}[n],$$  

(35)

with $\rho_T(x) = 0$ if $|x| < T$ and $\rho_T(x) = x$ otherwise. In our numerical experiments $T = \mu \sigma$ with $\mu = 3$.

Fig. 7(b) shows a noisy seismic image contaminated by an additive Gaussian white noise. The PSNR is 26db. Fig. 7(c) shows the resulting grouplet coefficients calculated from the multiscale association fields obtained with block matchings.
Fig. 7(d) shows the coefficients above threshold and the reconstructed seismic image in Fig. 7(e) has a PSNR of 27.3db. Setting a coefficient to zero is equivalent to average the image over the corresponding grouplet support. This grouplet thresholding thus performs an adaptive multiscale averaging of the signal along the directions of the multiscale association fields. The reconstructed signal has less noise but the discontinuities of the grouplets create blocking artifacts that appear in Fig. 7(e).

![Images](a) Original image, (b) Noisy seismic image (PSNR = 26db), (c) Orthogonal noisy grouplet coefficients, (d) Thresholded noisy grouplet coefficients, (e) Image reconstructed from thresholded orthogonal grouplet coefficients (PSNR = 27.3db). (f) Image reconstructed from thresholded tight frame grouplet coefficients (PSNR = 29.5db).

**Grouplets Tight Frames** Denoising by thresholding on a grouplet tight frame improves the estimation. The denoising estimator thresholds the frame grouplet coefficients and reconstructs a signal with a left inverse. To minimize the expected risk with a Gaussian white noise, it is necessary to use the pseudo-inverse, in which
case the estimator is defined by

$$
\hat{F}[n] = \sum_{j=1}^{J} \frac{1}{2^j} \sum_{m \in \mathcal{V}_i} \rho_T((X, g_{j,m})) g_{j,m}[n] + \frac{1}{2^j} \sum_{m \in \mathcal{V}_i} \rho_T((X, h_{j,m})) h_{j,m}[n]. \quad (36)
$$

With a pseudo-inverse, a tight frame thresholding is similar to averaging thresholding estimators in several orthonormal bases, which explains why it reduces the risk compared to a thresholding estimator in a single basis.

A multiscale association field is calculated from the noisy seismic image in Fig. 7(b) with a block matching algorithm. The resulting grouplet tight frame coefficients are also thresholded with $T = 3 \sigma$. Fig. 7(f) shows the resulting estimation obtained by applying an inverse grouplet transform with a PSNR of 29.5db. As expected, the risk is much smaller than with a grouplet orthogonal basis and the blocking artifacts produced by the orthogonal basis are reduced in the tight frame estimation. Fig. 8 gives another example of image denoising on a grey-level image. Grouplet tight frame denoising yields again a PSNR that is 2db above an ortho-
nal grouplet denoising. Although the tight frame redundancy reduces the artifacts, the reconstructed image still has irregularities due to the grouplet non-regularity. This motivates applying a grouplet transform not on the image but on a wavelet transform of the image, to inherit the regularity of wavelets as explained by Section 3.

When computing the association field with a block matching, a grouplet tight frame denoising has similarities with a non local mean denoising algorithm [2]. A non local mean denoising averages a noisy signal sample with other non local “neighbors” defined as image points having a block neighborhood which is sufficiently close to the block neighborhood of the original noisy signal sample. Thresholding a grouplet transform performs a similar averaging since points are also grouped, potentially with a block matching, however the grouping is performed with multiscale procedure and the averaging decision depends upon a different thresholding criteria that involves a normalized inner product.

**Image Zooming** Image zooming requires to compute intermediate pixels to enlarge the image size. Images are most often sampled at a sampling rate that is below the Nyquist rate. Since the sampling density
$X[m] = \text{NULL}$.

Let $N$ be the total number of samples in $G_0$ and $2^J$ be the maximum scale. For
\( \hat{s} = 0 \) then \( s[m] = 0 \). Otherwise if \( s[\tilde{m}] \neq 0 \) and \( \hat{s} \neq 0 \) then

\[
\tilde{a}_1 = a[m] - d_j[\tilde{m}] \frac{\sqrt{s[\tilde{m}]} \sqrt{\hat{s} s[m]}}{\sqrt{\hat{s}} s[m]}
\]

\( \tilde{a}_2 = a[m] + d_j[\tilde{m}] \frac{\sqrt{s[\tilde{m}]} \sqrt{\hat{s} s[m]}}{\sqrt{s[\tilde{m}]} s[m]} \). \hspace{1cm} (39)

The first average is stored “in place”, \( a[m] = \tilde{a}_1 \) and \( s[m] = \hat{s} \). The second reconstructed average is calculated as in (31)

\[
a[m] = \tilde{a}_2 \quad \text{with} \quad \tilde{a}_2 = \frac{a[\tilde{m}] + \tilde{a}_2}{2} .
\]

![Fig. 9](image.png)

**Fig. 9** – (a) : Subsampled noisy seismic image. (b) : Regularized zoom, with twice more columns, computed by thresholding tight frame gouplet coefficients.

Fig. 9(b) shows an example of zooming calculated from an aliased seismic noisy image in Fig. 9(a). The zoom increases the number of columns by a factor 2. Both images are shown at the same size. Despite the presence of noise, the regularization and direction interpolation performed by the gouplet thresholding restores in Fig. 9(b) geometric structures that are barely visible in Fig. 9(a).

### 3 Grouping Bandlets

Instead of applying a gouplet transform directly on the original image, one may rather apply this transform on a wavelet transform of the image, which already provides a sparse representation. Indeed wavelet coefficients are obtained through a convolution with a dilated wavelet and hence inherit the geometric regularity of the original image. Transforming these wavelet coefficients along directions where
they are geometrically regular can thus further improve the sparsity of wavelet representation. This is the central idea of bandlet constructions. Several approaches have already been proposed to construct bandlet orthonormal bases by applying an orthogonal transform on wavelet coefficients [14, 22]. Applying a grouplet transform over wavelet coefficients is a different approach that yields grouping bandlet coefficients. This construction has a hierarchical structure that could be compared to association field models of horizontal connections between V1 simple cells [8].

Section 3.1 begins by defining orthogonal grouping bandlet bases from orthogonal wavelet bases. Section 3.2 introduces tight frame grouping bandlets constructed over steering pyramid wavelets and separable tight frame wavelets. It compares noise removal by thresholding in a bandlet tight frame and in a wavelet tight frame.

### 3.1 Orthogonal Grouping Bandlet Bases

Orthogonal grouping bandlet bases are constructed by applying an orthogonal grouplet transform on orthogonal wavelet coefficients. Separable orthogonal wavelet bases of $L^2([0, 1]^2)$ are obtained by dilating and translating three separable wavelets $\{\psi^k\}_{1 \leq k \leq}$
For the second direction \( k = 2 \), large wavelet coefficients correspond to sharp transitions horizontally that belong to structures spread across several rows. The orthogonal grouplet transform is thus defined with row embedded subgrids that correspond to a subsampling of the rows of the wavelet coefficient images \( W_2^k f[n] \). The resulting grouplet vectors are elongated vertically across rows.

For the third direction \( k = 3 \), the wavelets \( \psi^3_j \) perform a high pass filtering along vertical and horizontal directions. Large coefficients thus belong to geometric structures that are spread across several rows and columns. A grouplet transform can thus take advantage of this geometric regularity by defining column embedded subgrids or row embedded subgrids. Fig. 10(c) shows grouping bandlet coefficients obtained with grouplet transforms using different multiscale association fields for each wavelet image \( W_3^k f[n] \). Bandlet coefficients are di
where \( \{ g_{k,j,m}^l[n], h_{l,j,m}^l[n] \}_j,m \) is the grouplet basis corresponding to the multiscale association fields computed over orthogonal wavelet coefficients for the direction \( k \) and scale \( 2^l \).

![Diagram](image1)

**Fig. 11** – (a) : Examples of orthogonal grouplets. (b) : Orthogonal grouping bandlets corresponding to the grouplets in (a).

Grouplets have no regularity but grouping bandlets in (43) have the same regularity as the wavelets \( \psi^k(x) \). The support of \( b_{l,j,m}^k(x) \) has a width proportional to \( 2^l \) around the grouplet \( g_{j,m}^k \) and a length typically proportional to \( 2^l \). Fig. 11 gives examples of grouplets and the corresponding bandlets.

### 3.2 Tight Frames of Grouping Bandlets and Denoising

To remove noise in images, thresholding translation invariant wavelet frame coefficients yields better estimators than thresholding orthogonal wavelet coefficients. Grouplet tight frames also yields better estimators by using a more flexible model of association fields than grouplet orthogonal bases. Applying grouplet tight frames to wavelet tight frame defines a tight frame of bandlet functions.

A separable wavelet frame is constructed by oversampling an orthogonal or a biorthogonal wavelet transform. From three separable orthogonal or biorthogonal wavelets \( \psi^k(x) \) for \( k = 1, 2, 3 \), the oversampled is defined by:

\[
Wf^k_l[n] = f \star \tilde{\psi}_l^k(n) \quad \text{with} \quad \tilde{\psi}_l^k(x) = 2^{-l} \psi^k(-2^{-l}x) \quad \text{for} \quad k = 1, 2, 3.
\]

Fig. 12 gives an example at scales \( 2^l = 2^1, 2^2 \).

A wavelet thresholding removes a Gaussian white noise of variance \( \sigma^2 \) from a noisy image \( X \) by setting to zero all dyadic wavelet coefficient \( Wf^l_kX[n] \) whose amplitude is below a threshold \( T = \mu \sigma \), and by applying the inverse wavelet transform. In images of 512 by 512 pixels, \( \mu \) is typically equal to 3. When the image is regular or if the edges or textures are too weak, the thresholding sets all coefficients to zero which averages the noisy coefficients. The two zooms in Fig.
14 and 15 show that it removes fine textures such as the fine strip hat texture or hairs, while randomly leaving isolated fine structures corresponding to wavelet coefficients above threshold. This wavelet denoising is improved by a grouplet transform over wavelet coefficients, that takes advantage of...
let, in order to compensate for the partial "noise fitting" that may be produced by the association fields computed on noisy coefficients with a block matching. Fig. 14 and 15 compare the denoising results of a translation invariant wavelet thresholding and of the grouping bandlet tight frame thresholding. Grouping bandlets restore fine textures on Lena's hat or on its hairs because the elongation of these bandlets along these geometrical structures better compresses the signal and yields coefficients above threshold which restore these geometrical structures. On Lena, a grouping bandlet thresholding estimation improves a wavelet thresholding estimation by more than $1db$. However, the multiscale association field must be computed and regularized carefully to avoid creating geometric bandlet artifacts in regular image regions. Optimizing the multiscale association fields to minimize the grouping bandlet estimation risk is an open problem that is beyond the scope of this paper.

**Fig. 14** – (a) : Original image. (b) : Noisy image. (c) : Translation invariant wavelet thresholding. (d) : Grouping bandlet tight frame thresholding.

**Acknowledgments** I would like to thank very much Valérie Charoing for our collaboration and the many discussions we had on this topic, and Gabriel Peyré for his generous help on numerical experiments and figures.
Fig. 15—(a) : Original image. (b) : Noisy image. (c) : Translation invariant wavelet thresholding. (d) : Grouping bandlet tight frame thresholding.

Références


34

