

Psychovisual Image Coding via an Exact Discrete Radon Transform

Jean-Pierre Guédon, Dominique Barba, Nicole Burger

Systèmes Electroniques et Informatiques, IRESTE
Atlantpole, Route de Gachet, CP 3003, Nantes, France, F44087

ABSTRACT

The goal of this paper is to describe a new fully-reversible image transform specifically designed for an efficient (pseudo-critical) coding while preserving a psychovisual Fourier domain description. There is now strong evidence for the presence of directional and angular sensitivity in the cells of the human visual cortex and the representation proposed here has for main objective to respect this human like filter bank. The decomposition is performed using a discrete Radon transform for the angular patches and by splitting each projection with a 1D spline wavelet for the radial part. Consequently, the whole algorithm is performed in the spatial domain. Finally, we show that the transform is both well-suited for psychovisual quantization and channel adapted coding.

Keywords : Discrete Radon transform, Psychovisual decomposition, Wavelet decomposition, Image coding, source-channel adapted coding.

1. INTRODUCTION

1.1 Human visual system decomposition

Taking into account the human visual system is of prime importance for implementing an image coding scheme which only keeps the main content of the possibly visible information. As described in the literature, the foveal response of the human visual system (HVS) can be modeled as a filter bank^{1,2}. A first set of psychovisual experiments² demonstrates the radial/angular channel description in the spatial Fourier domain which is presented in Figure 1. Each patch is localized by two numbers standing for the corona and angular location. The 2D Fourier transform (FT) of the initial image can be dispatched according to these regions and subsequent filter banks are computable as done in². The original image is then splitted into sub-images corresponding to each patch by signal processing tools. Then the quantization is performed not on the sub images values but on a local contrast scale^{2,3}. This local contrast explains the value for a pixel in the original image as the ratio of the pixel luminance in a sub-image, corresponding to the (*i*th radial band, *j*th angular sector) patch, divided by the pixel luminance relative to the sub-image reconstructed from its lower-frequency ($0 \leq k < i$) coronas (i.e. the local mean luminance value)

$$C_{i,j}(m,n) = \frac{L_{i,j}(m,n)}{\sum_{k=0}^{i-1} \sum_l L_{k,l}(m,n)} \quad (1)$$

The major result with this contrast quantization is that a psychovisual quantizer built up on these premises gives linear laws for each patch. This local contrast definition is also consistent with the way the HVS behaves and gives a true psychovisual quantization of the image. However, the corresponding coding scheme needs a Fourier transform of the luminance image before performing the patch coding. The quantization step obtained from (1) is implemented in the

spatial domain thus involving an inverse Fourier transform. The decoding scheme is also based on Fourier direct and inverse transforms.

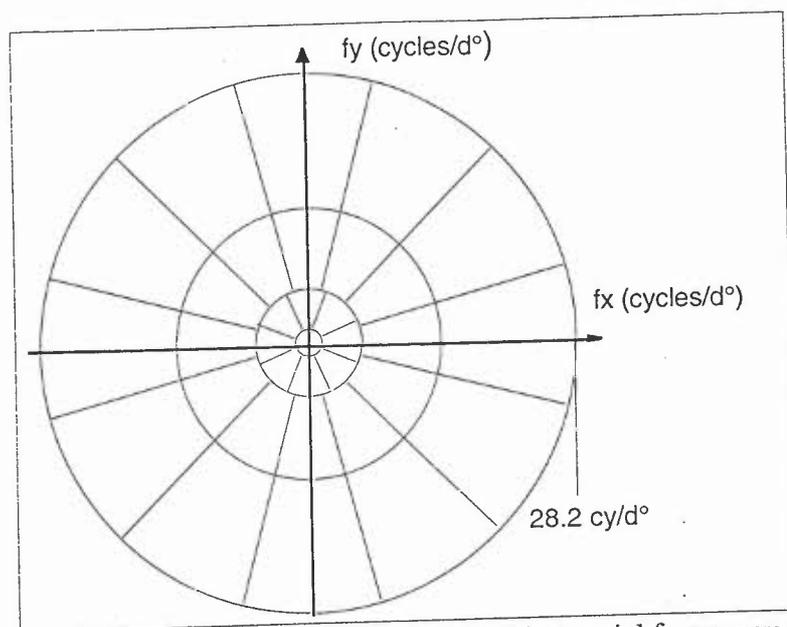


Figure 1 : The psychovisual decomposition in the spatial frequency domain

1.2 Fourier / Radon transform

Let $f(x,y)$ be a continuous image. Its Radon transform ⁴ noted Rf is composed of all the projections of the image along an axis t for every angle $\theta \in [0,\pi[$, i.e.

$$Rf(x,y) = p_{\theta}(t) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y) \delta(t - x\sin\theta + y\cos\theta) dx dy . \quad (2)$$

The inverse transform exists but this is generally an ill-posed problem. This ill-posed nature can be overcome in the discrete case. The link between Fourier and Radon transform is given by the Central Slice Theorem. If we note $P_{\theta}(v)$ the 1D Fourier transform of the projection $p_{\theta}(t)$ and

$F(\lambda,\mu)$ the 2D Fourier transform of $f(x,y)$, then we have,

$$P_{\theta}(v) = F(\lambda=v \sin\theta, \mu=-v \cos\theta). \quad (3)$$

This theorem explains why the Radon space is well adapted to the HVS angular decomposition since each central slice in the Fourier domain is directly related to a projection of the image. In practice, a $N \times N$ digital image gives a discrete Fourier transform onto a Cartesian grid. By choosing projections whose 1D Fourier transform samples exactly matches the grid, an exact image transform will be defined. Benefits from the use of a discrete Radon transform lie in defining a spatial implementation and permitting an exact representation of the original signal. A direct consequence is that the scheme is implemented with fast algorithms.

2. THE DISCRETE TRANSFORM

2.1 The direct transform

The definition of the discrete transform starts from the discrete quantized image and gives discrete integers. Following Katz ⁵, we shall only use specific angles of the form $\tan \theta = \frac{p}{q}$

where $p, q \in \mathbb{Z}$ and prime each other ($\text{GCD}(p, q) = 1$). To avoid interpolation onto the projections the sampling rate is also angle-dependent. The corresponding discrete Radon transform is defined as a simple pixel summation by

$$R_{\theta}f(k, l) = \text{proj}_{p, q}(m) = \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} f(k, l) \Delta(m - kp + ql), \quad (4)$$

with

$$\Delta(m - kp + ql) = 1 \text{ if } m = kp - ql \text{ and } 0 \text{ elsewhere.} \quad (5)$$

For a pixel size of $\Delta \times \Delta$, $\text{proj}_{p, q}(m)$ is sampled at $m\Delta \cos\theta - \sin\theta l$ when $\theta \in [0, \frac{\pi}{2}[$, and at $m\Delta \cos\theta + \sin\theta l$ if $\theta \in [\frac{\pi}{2}, \pi[$. This relationship is related to (Eq. 4) by $\sin\theta = \frac{p}{\sqrt{p^2+q^2}}$ and $\cos\theta = \frac{q}{\sqrt{p^2+q^2}}$. An example is given in Figure 2.

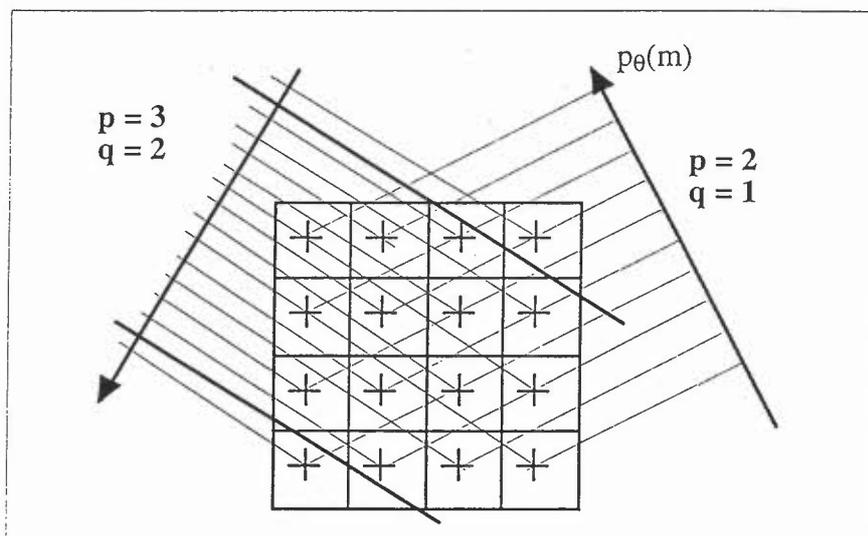


Figure 2 : The Discrete Radon transform computed for two angles

The main advantage of this definition is that a projection cell value is simply obtained by integer addition of pixel values. The main differences compared to standard discrete Radon transforms lie first in the angular-dependent sampling rate and secondly in the specific angle values which forbid every implementation for tomographic reconstruction. Notice that the transform is a discrete to discrete operator as demonstrated by the algorithm described in Appendix A. It is also of importance to remark that some projection cells will correspond to no pixel in the image because of the p/q ratio and thus are loaded with a null value (see Figure 2, for $p=3, q=2$). To fully describe the transform, the number of projections (for a given image size N) and their directions have to be derived.

Katz has shown that, according to the image resolution, a bound on the number M of projections necessary for a reversible transform can be found. More precisely, Katz's theorem says that for a set of (p_i, q_i) , the selected integer pairs such that $\tan\theta_i = \frac{p_i}{q_i}$, the null space of the discrete Radon transform is empty if the image size is bounded by $N < 1 + \text{Max}(\sum_{i=1}^M |p_i|, \sum_{i=1}^M |q_i|)$. Even if this inequation does not give the lowest bound for the number of angles M with respect

to N , it allows for a rough idea on their relationship as given by Table 1 where an arbitrary choice (not optimal) on the (p_i, q_i) has been made.

M	4	7	11	18	29
N	3	8	16	32	63

Table 1: Bounds on the number of angles M versus the image size achieved

Notice that no choice on the projection angles is made by this bound. For instance, two projections are computed for a 2×2 image in Figure 3: selecting $\theta=0$ and $\pi/2$ (4-a) does not lead to a unique solution whereas it is the case in (3-b) for $\theta=\pi/4$ and $3\pi/4$. However, the number of samples is no longer critical in this case. It is a pseudo-critical sampling as in (3-c) by choosing angles $\theta=0$ and $\pi/4$ (5 samples in the transform domain for 4 initial values). The first choice (3.a) does not produce a unique transform as demonstrated by its Fourier transform: the sum of all cells in a projection (which equals the sum of pixels in the image) is a redundant information contained in the sample value at the origin of the Radon domain: we end up with only 3 decorrelated samples. On the contrary, configurations (3.b) and (3.c) are acceptable because there are enough decorrelated samples.

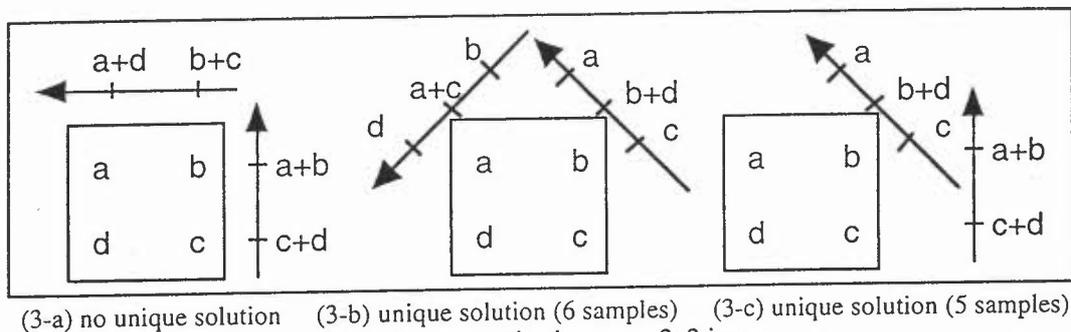


Figure 3: two projections on a 2×2 image

The Radon transform is the most appropriate tool for describing the angular features of the psychovisual patches. Concerning the radial part, it is well-known that a multiresolution analysis uses higher and higher frequencies with increasing resolution. Defining basis functions onto a disk domain would give us the final patches in the Fourier domain. However, due to the different sampling path in each projection and the fact that different projections have a different number of samples, the covered area is a square which is also consistent with the Central Slice Theorem.

Notice that m_θ , the number of bins onto the projection, depends on the (p, q) values and equals

$$m_\theta = 1 + (N-1) \cdot (|p| + |q|) \quad (6)$$

So the smallest (p, q) values produce the highest number of summed pixels in a bin with the lowest m_q value. On the other side, for high (p, q) values (i.e. $|p| + |q| \cong N$) m_θ is roughly the same as the pixels number: a bin value corresponds to only very few pixels. This will be the key point for the inverse transform. Another crucial point is the choice of the set of (p, q) values. As shown in Figure 3 for $N=2$, $\{(0,1), (1,1)\}$ and $\{(1,1), (-1,1)\}$ are two sets providing the unique solution.

The choice of the projection set leads to the algorithm complexity to efficiently solve for the inverse transform. At this point, an elegant solution uses Farey's series. A Farey series of order N is the set F_N of irreducible ratios $\frac{p}{q}$ in $[0, 1]$ ($q \leq N$) arranged in increasing order. For instance, $F_6 = (\frac{0}{1}, \frac{1}{6}, \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{5}{6}, \frac{1}{1})$. There are interesting properties on

F_N like the fact that two following elements $\frac{p}{q}$ and $\frac{p'}{q'}$ verify $p'q - pq' = 1$. Another more simplistic choice would be based on powers of 2 like $(\frac{0}{1} \frac{1}{4} \frac{1}{2} \frac{3}{4} \frac{1}{1})$. This is not very interesting on the psychovisual side : the angles have to be nicely disposed into $[0, \pi[$ to reduce the non-regularity of the grid obtained in the Fourier domain : this can be precisely obtained with the Farey's angles.

The direct transform, as described in Appendix 1, has a complexity order on $O(MN^2)$ where M is the number of projections and N^2 the pixel number.

2.2 The monoresolution inverse transform

A generic algorithm

The computation of the inverse transform lies in inverting a linear system $Ax=b$, where the matrix A is $N^2 \times N^2$. We avoid solving for the inverse matrix (which is nonetheless very stable in this case) by ordering the relative degree of pixels' summation onto the projections. The image corners are first reconstructed since directly corresponding to bin values. We then take bins containing higher and higher number of pixels and solve at each time triangular systems which are easy to inverse (only 1 and 0 values). As an example, we showed that for a 32×32 image ($N=32$) a set of angles θ such that $q \in [0, 4]$ and $p \in [-4, 4]$ is larger enough and the solution is obtained without even taking the angles $(\theta=0, \frac{\pi}{2}, \frac{\pi}{4}, \frac{3\pi}{4})$. However, even with this small set of angles, the total number of bins is almost three times the number of pixels. Since a pixel reconstruction only use already reconstructed pixels and a single element of a projection, the non-relevant bins can be omitted (not transmitted). In this case, the location of each transmitted bin has to be known by the receptor to avoid the transmission of a data structure.

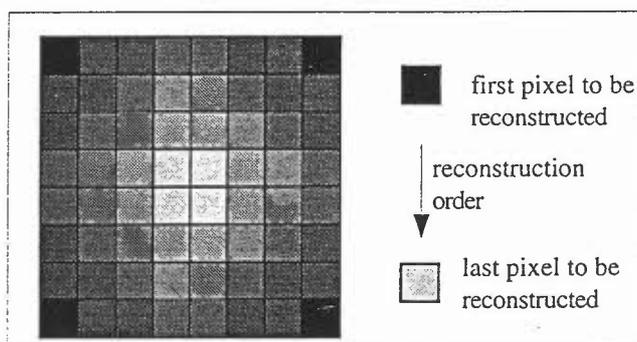


Figure 4 : The inverse transform progression

An implementation based on sub-images

This simple idea was directly taken from the standard block cosine transform. The initial image is splitted into square sub-images and the transform is applied onto each data block. Algorithms for the inverse transform corresponding to $N=4, 8, 16$ and 32 were obtained. A sub-image of dimension 32×32 ($N=32$) was taken using angles picked up in the Farey's series F_4 . Each pixel (i, j) of the sub-image is computed using a single cell element and a number of other pixels of the order of $\inf\{\frac{i}{p}, \frac{n-j}{q}\} + \inf\{\frac{i}{q}, \frac{n-j}{p}\}$. The number of compelled values ranges from 1 (image corner) to N (image center), so the algorithm order is in between N^2 and N^3 . Since we only use additions and subtractions to solve for any linear systems, the unicity of the solution is guaranteed. The generalization to every rectangular size has also been found.

However, mathematical principles behind that algorithm come from number theory and a detailed proof is beyond the scope of this paper.

The correspondence between Fourier and Radon transforms extends the block splitting. The projection can also be quantized and only significant variations transmitted. When the projection is decomposed into low and high resolution parts with a dyadic scheme, its high-frequency part can be zero and has not to be transmitted.

2.3 The multiresolution inverse transform

The multiresolution analysis can drastically reduce the amount of transmitted samples. It also corresponds to the final psychovisual patch splitting which allows for a quantization using the local band-limited contrast.

From the full-resolution image (supposed square here for simplicity), the $N \times N$ pixels are transformed into projections by the direct transform. We assume here to compute enough projections to cover the psychovisual domain and give back the unique solution. Each projection is splitted into dyadic low-res(olution) and high-res parts by the use of a 1D scaling function and its wavelet counterpart (noted W_θ in the following). We use a cubic cardinal spline scaling function and its associated wavelet to perform this decomposition⁶. This biorthogonal quadrature mirror filter has two interesting properties: the associated scaling function is very close to the ideal sinc filter (as described in Figure 5) and the wavelet rapidly vanishes in the spatial domain. Notice also that the wavelet is defined for each projection since the cut-off frequency $F_m(\theta)$ is related to the angle $\theta_i = \text{atan} \frac{p_i}{q_i}$. Thus even if the number of bins depends on the considered projection, the dyadic decomposition can be applied.

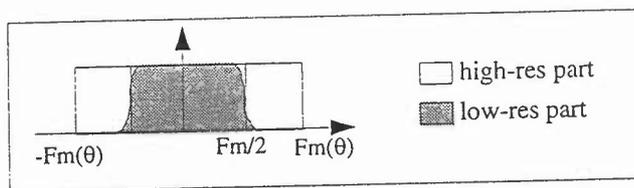


Figure 5 : Splitting the projection line using the cubic cardinal spline wavelet transform

The spline order rules the transition zone between two adjacent coronas and we obtain a 2-D dyadic splitting of the HVS-plane because of the relative cut-off frequency of each angle as described in Figure 6.

Now, the low-res projections can be examined as a whole corresponding to a low-res image at resolution $\frac{N}{2} \times \frac{N}{2}$. However, there are too many samples and some projections can be discarded without jeopardizing the unicity of the solution for the low-res image as shown in Figure 6. Interestingly enough, we are able to reconstruct these projections by using the inverse transform at low resolution and recomputing the transform at the discarded angles : these projection parts do not have to be transmitted since computable. This process can be recursively use to restrict the transmission to a pseudo-critical sampling while preserving the quasi-dyadic psychovisual scheme. For instance, as the sum of cells in each projection is constant since corresponding to the sum of pixels in the image or the central point in the Fourier domain, it is not necessary to transmit this value M times but only once.

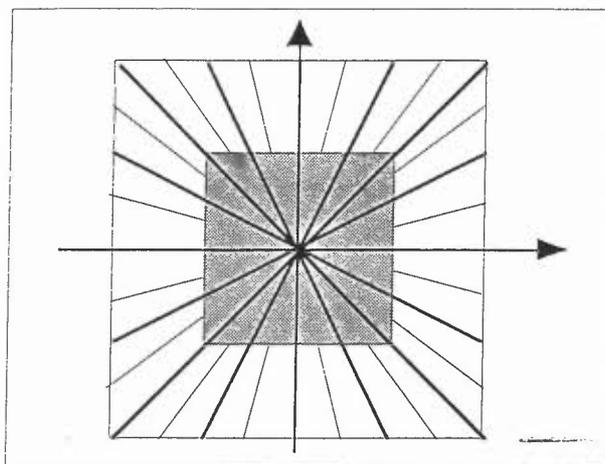


Figure 6 : The corresponding decomposition with 2 levels of frequencies.

The inverse transform with a multiresolution analysis is easily described. We start for instance with a 2×2 image and compute some new low-res projections. Each of these projections are separately grouped with their (4×4) high-res transmitted counterparts. On this new set the transform is recomputed giving the 4×4 image and the process goes on. In other words, the reconstruction of the image with a multiresolution scheme use both the direct and inverse transform at each resolution. The low computation order for both transforms is then a key point here.

3. PSYCHOVISUAL CODING SCHEME

3.1 The coding scheme

From the direct transform and the mono- and multi-resolution inverse transforms, the coding scheme is easily deduced. For sake of simplicity, we restrict here to only two squared coronas. The first step consists in obtaining the set of projections from the original image. Each projection is then splitted into high-res and low-res parts using the wavelet transform. The third step groups all the projection parts belonging to the same psychovisual patch (angular range, radial range). Then the psychovisual quantization takes place as explained in ². The local contrast can be directly derived here since all the computations are done in the spatial domain. The corresponding scheme is summarized in Figure 7.

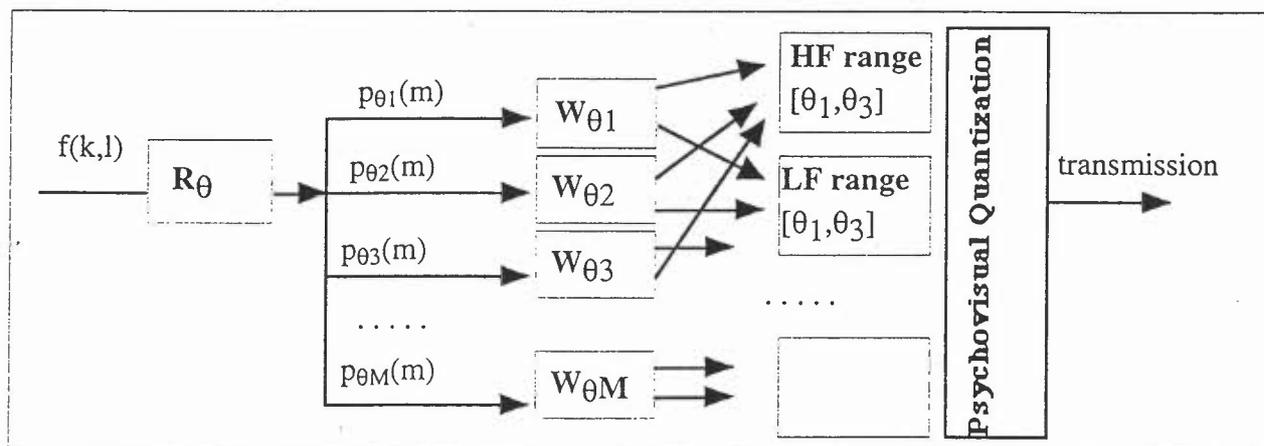


Figure 7 : The Radon-psychovisual coding scheme

The luminance transformation applied onto the original image in order to quantize an effective local contrast can also be done here.

3.2 The decoding scheme

As expected, the decoding scheme proceeds in the exact reverse way. Figure 8 describes the corresponding operations. When a patch has been received, the inverse quantization gives the (low-res and high-res) projection parts.

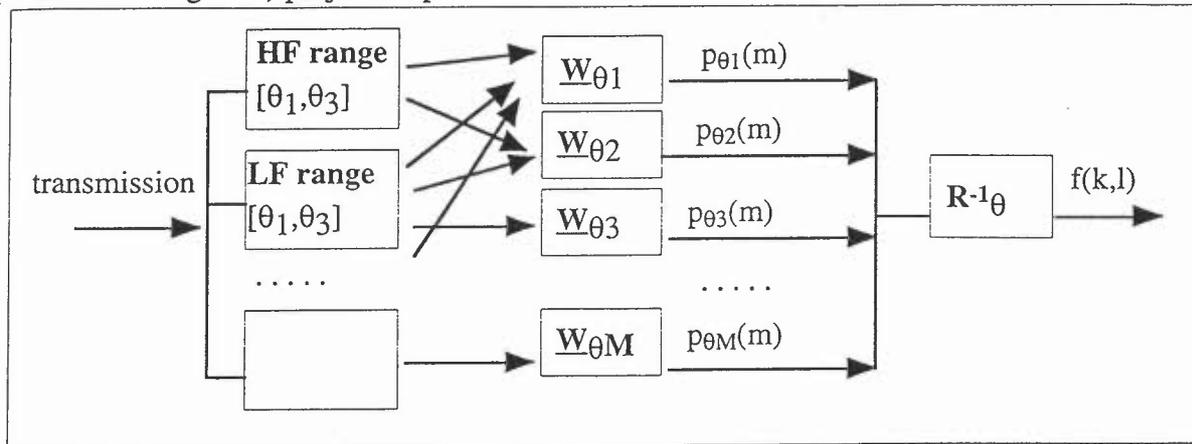


Figure 8 : The Radon-psychovisual decoding scheme

Each projection is then build up using the biorthogonal transformation. Like in the multiresolution inverse transform, the missing low-res projection parts are reconstructed using the direct transform of the low-res image. When all the projections are reconstructed, the inverse Radon transform is applied.

Associating portion of projections before quantization can be represented in the Fourier domain as shown in Figure 9. Each grey level corresponds to projection parts inside a patch which have been quantized using a different psychovisual quantizer law. The angular range is then the same as used for the previous decomposition 2, while the radial range is slightly different since the dyadic scheme on each projection induces a square partition instead of regular (disks) coronas.

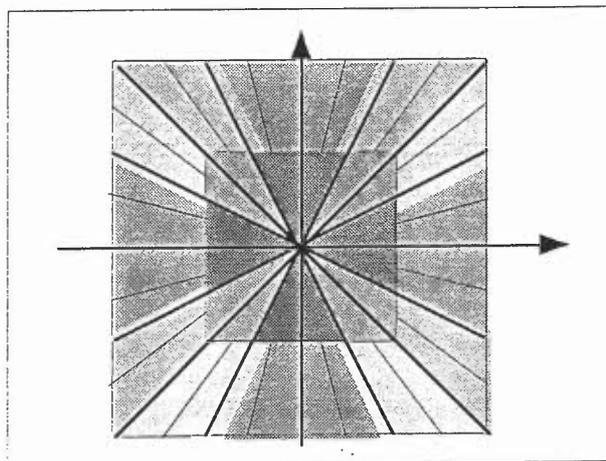


Figure 9 : The corresponding psychovisual decomposition with 2 coronas.

4. DISCUSSION

4.1 Results

The direct transform has been implemented and its rectangular generalization to any image size is obvious from the algorithm given in Appendix A. The inverse mono-resolution transform has been implemented for sub-images of size 4^2 , 8^2 , 16^2 , and 32^2 . We thus reconstructed images with the transform. Moreover, a generalization of the inverse transform for any image dimensions has been found and is currently under test. We actually split the images into 32^2 square sub-images and perform the transform and its inverse on each sub-image. This dimension is not larger enough to fully use the multiresolution inverse transform. By nature, the transform is exact and since no quantization has been made, its effects are not yet understood.

4.2 Differences with the classical psychovisual scheme

The first main difference between the psychovisual scheme described in ² and this one is the substitution of the disk support by a square one. There are no obvious reason for the eye directional isotropy not to be kept in the images. Moreover, the classical psychovisual scheme takes profit of this property (visual information not generally seen) for discarding the area between the square and the disk (high frequencies in the four corners). However, the present scheme can be used when no specific part of information has to be removed. This is precisely the case for medical images or image classes where a detection task has to be done on the image. If the signal to be detected belongs to the corner regions it could be detected with this kind of coding scheme.

The second difference between a classical psychovisual scheme and this one lies on the radial expression of the initial information for the latter. As a matter of fact, the classical transformation used for coding employs a 2D separable Fourier transform of the original image before reducing the frequency domain to the unitary disk. Unfortunately, to split the disk into several radial/angular patches a radial/angular grid can summarize the initial information in a better way. In the presented scheme, we eventually implement the exact reverse situation by keeping the square domain but partitioning it in a radial way. The obtained lattice is not a regular lattice but it perfectly matches the partition constraints and the dyadic scheme. By mixing the radial discrete grid with a still separable grid in the Fourier space this representation simplifies the 2D signal processing.

The third interesting point is that fast algorithms are obtained with the new transform. The previous representation can be implemented with a fast Fourier transform if the image domain has width and height sizes which are powers of two. This leads to algorithms of order $n \cdot \log_2 n$ which have to be implemented twice for the direct and inverse transform. For other cases, a prime number decomposition of the width and height has to be used. The Radon transform implementation can be done for every rectangular images without any changes in the algorithm. Moreover, the spatial implementation (Radon and wavelet transform) leads to fast algorithms.

An advantage remains for the classical psychovisual transform compared to this new scheme. This is the adequation of the Fourier plane cutting up with the psychovisual experiments and the associated quantization laws. We do not know if the square splitting will keep linear these quantization laws but it would be expected some degradations in the linearity.

4.3 Channel-adapted coding scheme

A very nice feature of this algorithm is the redundancy introduced during the transformation step. Contrarily to most transforms classically used, the number of projections can be higher than required for the inverse transform. In the context of high speed networks like B-ISDN, the redundancies classically introduced by the network protocol are mostly relaxed to meet real-time constraints on data exchanges. The higher protocol layers have to choose if these information protection mechanisms meet the real-time constraints. For image transport with this kind of coding, we are able to relax such constraints. In effect, if a projection has not been correctly transmitted, any other projection can replace it.

5. CONCLUSION

In this paper, we presented a new discrete transform and its inverse based on an exact discrete Radon transform. This transform is well adapted to decompose the visual information and do not need a Fourier transformation of the data. The direct transform only uses projection angles which tangent is an integer ratio. The inverse transform was presented both from a mono- and multi-resolution analysis scheme. The latter uses the direct transform to generate low-resolution projection which are not transmitted. The 2D multiresolution analysis is decomposed into M different 1D multiresolution decompositions. The psychovisual coding scheme is then obtained from a radial/angular information decomposition where the Radon (respectively the wavelet) transform provides the angular (resp. the radial) decomposition. By applying the Radon transform followed by the wavelet decomposition, the visual patches are built and can be quantized. The decoding scheme proceeds in the same way. This psychovisual coding scheme is also well-adapted to high-speed network transmission due to the redundancy directly introduced in the original transformation.

ACKNOWLEDGEMENTS

The authors thank Nicolas Normand and Laurent Bédât for their valuable suggestions in writing the manuscript.

APPENDIX : PROJECTION CODING ALGORITHM

Radon (N, p, q)

```
pix ← 0           /pixel index in the image/
bin ← 0           /bin index onto the projection/
if (p ≥ 0) then start ← 0       /if  $\theta \in [0, \pi/2[$  /
    else start ← p.(1-N)       /if  $\theta \in [\pi/2, \pi[$  /
for i ← 0 to (N-1) do
    bin ← start + i.p
    for j ← 0 to (N-1) do
        projection(bin) ← image(pix);   bin ← q;   pix ← 1
    end for
end for
end Radon
```

REFERENCES

1. A.B. Watson : 'Efficiency of a model human image code', JOSA, Vol.4, pp. 2401-2417, 1987.
2. A. Saadane, H. Sénane, D. Barba : 'On the design of psychovisual quantizers for a visual subband image coding', SPIE Proc. VCIP 94, Vol. 2308, pp. 1446-1453, 1994.
3. E. Peli, 'Contrast in complex images,' JOSA, Vol. 7, n°10, pp. 2032-2040, October 1990.
4. J. Radon, 'Über die Bestimmung von Functionen durch ihre Integralwerte langs gewisser Mannigfaltigkeiten,' Berichte Sachsische Academie der Wissenschaften, Leipzig, *Math. - Phys. Kl.*, vol 69 pp. 262-267, 1917.
5. M. Katz, *Questions of Uniqueness and Resolution in Reconstruction from projections*, Lecture Notes in Biomathematics, Springer-Verlag, Vol. 26, 1979.
6. M. Unser, A. Aldroubi and M. Eden, 'Polynomial spline signal approximations : filter design and asymptotic equivalence with Shannon's sampling theorem,' *IEEE Trans. Information Theory*, vol. 38, pp. 95-103, January 1992.
7. A. Saadane, H. Sénane, D. Barba : 'Masking and quantization laws in a visual subband coding scheme,' IEEE Conf. on image processing, Vol. II, pp.869-873 , November 1994.

