jet INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2010, VOL. 56, NO. 3, PP. 273-280
Manuscript received July 10, 2010: revised September, 2010. DOI: 10.2478/v10177-010-0036-x

Semantic Sparse Representation of Disease Patterns

Artur Przelaskowski

***Abstract*—Sparse data representation is discussed in a context of useful fundamentals led to semantic content description and extraction of information. Disease patterns as semantic information extracted from medical images were underlined because of discussed application of computer-aided diagnosis. Compressive sensing rules were adjusted to the requirements of diagnostic pattern recognition. Proposed methodology of sparse disease patterns considers accuracy of sparse representation to estimate target content for detailed analysis. Semantics of sparse representation were modeled by morphological content analysis. Subtle or hidden components were extracted and displayed to increase information completeness. Usefulness of sparsity was verified for computer-aided diagnosis of stroke based on brain CT scans. Implemented method was based on selective and sparse representation of subtle hypodensity to improve diagnosis. Visual expression of disease signatures was fixed to radiologist requirements, domain knowledge and experimental analysis issues. Diagnosis assistance suitability was proven by experimental subjective rating and automatic recognition.**

***Keywords*—Sparse representation, compressive sensing, information theory, semantic information, disease pattern.**

I. Introduction

THE problem of data representation is one of the most critical issue concerning the realization of intelligence machines which are able to solve real life problems [2]. Adjustment of signal or source data representation to application requirements is a clue problem of many theories and algorithmic solutions. General purpose is successful separation of a signal content to manipulate it effectively.

More advanced signal study looks for information representation intended to represent only such data which are useful for the user/receiver. Data redundancy relates in that case to signal components useless because of its irrelevant meaning according to semantic information theory. Generally, information means semantic content functional for epistemic purposes. Floridi [10] defined semantic information as well-formed, meaningful and veridical data.

Well-formed data often means sparse representation of signal content expressed by morphology with attached meaning and veridicy. Structured, hierarchical and scalable representation well-fitted to specific domain knowledge may be realized by sparsity of a whole signal representation or by sparse form of semantic components useful for data analysis. Systematic approach to the analysis of semantic features for spatial information is intensively searched. Well-formed image data with semantic sparsity of desired information is required for many computational intelligence applications.

The role of effective information representation is extremely significant for computer medical applications, especially computer-aided diagnosis [12]. Human need computers in radiology mostly because of limited accuracy of radiological diagnosis. Computer-based aiding tools are able to storage, communicate, retrieve, emphasize, recognize and distinctly visualize image-based diagnostic content. Because the key constituents of diagnosis are the accurate detection and defining of the disease, full understanding and specific assessment of image content including semantics of recognized objects, mutual relations and accessible information complements is a key issue of successful exploitation of imaging capabilities in diagnosis.

A. Przelaskowski is with the Institute of Radioelectronics, Warsaw University of Technology, Nowowiejska 15/19, 00-665 Warsaw, Poland (e-mail: arturp@ire.pw.edu.pl).

For radiological applications, sparse signatures of disease patterns, invariably identifying clue diagnostic information across imaging conditions, cases and progress form are searched. Enhanced pathology symptoms are recognized to be interpreted according to diagnostic rules and procedures.

One of important subjects is subtle or hidden signal extraction. Through expansions in local, scalable and adaptively adjusted bases capability for signal energy packing with preserved localization across scales and subbands is realized. Selection of specific decomposition atoms adjusted to crucial image features allow target content modeling and extraction through sparse data representation [9].

Sparse texture analysis is used for semantic component selection. Identification of dominant morphological ingredients is the most optimistic step for image analysis. Extracted texture characteristics may be useful for specific pattern recognition. Computer aided diagnosis often investigates subtle signatures of pathology in a context of general content characteristics. Numerical data analysis is used for sensing compliment of human content assessment and interpretation. Thus formalized medical knowledge is required for dominant component recognition. Moreover, empirical knowledge from reference database indexed by content is useful for comparative case study. Numerical descriptors of image content resonant to semantic image extent are designed according to structured knowledge and following expert requirements [15]. The expected results are computational models of image semantics formulating sparse representation of information.

The main contribution of this paper is outlined methodology of sparse disease patterns. The concepts of compressive sensing were adjusted to real diagnostic problem of acute stroke diagnosis. We proposed and characterized sensing rules based on accuracy, semantics and usefulness of sparse disease patterns. Realized detection method of sparse stroke signatures was verified by experimental subjective rating and automatic recognition. Concluded remark is that semantic sparse representation of medical images is useful for pathology extraction and recognition.

II. Sparse Data Representation – the Problem Definition

Common practice of signal acquisition follows the basic principle of the Nyquist-Shannon sampling theory defined for frequency bandlimited signals. It states that perfect reconstruction of a signal needs the number of signal samples or measures that is dictated by its bandwidth. The number of Fourier samples we need to acquire must match the desired resolution of the signal. However, it is possible to reconstruct signals of interests accurately and sometimes even exactly from a number of samples which is far smaller than the desired resolution of the signal, e.g. the number of acquired measures. A few years ago, *compressive sampling* (or sensing) theory has emerged which shows that super-resolved signals can be reconstructed from far fewer measurements than what is usually considered necessary [4]. Since we can sample signals at roughly the "information rate" rather than the Nyquist rate. Compressive sensing emphasizes sparsity as very attractive theoretical and practical signal property that has played and continues to play a fundamental role in many fields of signal processing applications.

Sparsity and, more generally, compressibility leads to efficient data source modeling and separation, estimation, coding, dimensionality reduction, feature extraction and pattern recognition. Nowadays, sparse signal representation plays extremely important role in many up-to-date theories and applications [22]. In the field of medical imaging (according to current call for papers of special issue IEEE Trans Medical Imaging[1]), compressive sensing allows accurate recovery of an image from far fewer measurements than the number of unknowns. Moreover, it does not require a close match between the sampling pattern and characteristic image structures giving sparse structure essence estimates. Compact and sensible signal appearance is hot topic of medical imaging because its transformative potential in major aspects of system design, algorithm development, and preclinical and clinical applications.

For instance, data representation influences the generalization error of kernel based learning machines like Support Vector Machines (SVM) for classification [1]. In case of sparse data representation, the generalization capacity of an SVM trained by using polynomial or Gaussian kernel functions is equal to the one of a linear SVM. It means that sparse data representations reduce the generalization error as long as the representation is not too sparse, as in the case of very large dictionaries.

A. Basics of Sparsity

Strictly, a signal is sparse in relation to source dimension if most of its entries are (approximately) zero. Let's consider discrete signal $\mathbf{f}$ in finite-dimensional subspace of $\mathbf{R}^N$ as a vector of N measurements: $\mathbf{f} = [f_1, f_2, \ldots, f_N]$. A vector is exactly **sparse** if a finite set of significant measurement indexes $\Lambda = \{1 \leq i \leq N : f_i \neq 0\}$ is of cardinality $\#\Lambda = M \ll N$ [23]. In that case we can say that $\mathbf{f}$ is M–sparse, i.e. $||\mathbf{f}||_0 = M$ (the number of nonzero measurements of x is called l_0 pseudo-norm of x: $||x||_0$).

[1] http://www.ieee-tmi.org/CallForPapers.html

Most real signals sampled or acquired are not sparse in source space but they can be sparse after being decomposed on a specific set of functions – elementary waveforms called atoms. Appropriate transform domain is searched to make signal representation as sparse as possible. Generally, it is realized by signal expansion over dictionary of atoms, i.e. possibly redundant collection $\mathcal{D} = \{\varphi_i\}_{i=1}^{I}$ of unit-norm vectors: $||\varphi_i||_2 = 1$ for all i such as span$\{\varphi_i\} = \mathbf{R}^N$. If $\{\varphi_i\}$ are linearly dependent, the dictionary is redundant.

Optimized expansion in $\mathcal{D}$ leads to compact representation of a large class of signals according to compressive sampling concept. Signal $\mathbf{f}$ is synthesized as a linear combination of M respectively adjusted atoms of $\mathcal{D}$ (according to ad-hoc or more less formal method), such that

$$\mathbf{f} = \sum_{j=1}^{M} a_j \varphi_{i_j} \tag{1}$$

Decomposed representation of $\mathbf{f}$ in $\mathcal{D}$ is a vector of coefficients $\mathbf{a} = [a_1, a_2, \ldots, a_M] \in \mathbf{R}^N$ and $\mathbf{a} =< \mathbf{f}, \varphi_{i_j} >$.

Flexible expansion atoms are adaptively adjusted to morphological signal content according to the prior knowledge and signal sparsity requirements or assumptions. Sparse signal expansion in specific $\Phi = \{\varphi_{i_j}\}_{j=1}^{M}$ means that only a few atoms of $\mathcal{D}$ are active to describe $\mathbf{f}$. The signal is modeled only with the atoms well approximating its investigated features.

Formally, exact (interpolated) sparse representation problem is defined as solution of

$$\min_{\mathbf{a}} ||\mathbf{a}||_0 \text{ subject to } \mathbf{f} = \sum_{j=1}^{M} a_j \varphi_{i_j} \tag{2}$$

what means finding the sparsest representation of $\mathbf{f}$ over $\mathcal{D}$ [13]. Optimization procedure design tries to answer how to construct such set of approximants when the approximated function is known.

In practice, many signals of interests are not exactly sparse in expansion over any basis because of any acquisition limitations, specificity of analyzed semantic information etc. Instead, they may be weakly sparse or approximately sparse that the sorted magnitudes of representation coefficients decay quickly according to different forms of power law.

A sparsest approximation of $\mathbf{f}$ that achieves error $\epsilon \geq 0$ (error sparse approximation) is found by solving

$$\min_{\mathbf{a}} ||\mathbf{a}||_0 \text{ subject to } ||\mathbf{f} - \sum_{j=1}^{M} a_j \varphi_{i_j}||_2 \leq \epsilon \tag{3}$$

Alternatively, best approximation problem with other constraints is formulated as

$$\min_{\mathbf{a}} ||\mathbf{f} - \sum_{j=1}^{M} a_j \varphi_{i_j}||_2 \text{ subject to } ||\mathbf{a}||_0 \leq M \tag{4}$$

The purpose is to find the sparsest representation of $\mathbf{f}$ using assumed number of M atoms (minimizing approximation error with assumption of M–sparse approximation).

Unfortunately, given an arbitrary redundant dictionary $\mathcal{D}$ and a signal $\mathbf{f}$, it is NP-hard to solve the sparse representation problem. But compressed sensing proposes strong theoretical and algorithmic support for methods that investigate sparse solutions. One of important issues is that natural and most of interest dictionaries are far from arbitrary. Orthogonal transform bases give unique solution to sparse problem. The coherence of a dictionary is a measure of dictionary usefulness. Large, incoherent (with small coherency) dictionaries make the solution of spare problem more predictable.

B. Proposed Methodology of Sparse Disease Patterns (SDP)

Disease patterns are searched as a descriptors of specific diagnostic information. Verified hypothesis is that sparse representation of image data is useful for pathology extraction and recognition. However, signal decomposition procedures should comply domain knowledge and specific requirements of considered application. Proposed scheme is partly general, partly adapted for recognition of diagnostic information hidden in source data.

We propose three conditions important for sparse representation efficiency, considering even (if possible) signal acquisition in "information rate" or postprocessing, according to compressive sensing rules:

- accuracy of sparse representation – ϵ approximated signal (noiseless setting is $\epsilon = 0$) representation form according to actual signal estimation criteria; generally noise or artifacts should be reduced through increased sparsity of the measurements;
- semantics of sparse representation – information extraction by morphological content analysis and adjustment of expansion coefficient distribution; diagnostically important components which are subtle or hidden can be extracted to increase information completeness;
- usefulness of sparse representation – determining of the receiver requirements and assurance of user-oriented output; domain knowledge is extremely useful to represent semantic information in a form optimized to application requirements.

III. Accuracy of Sparse Representation

Accuracy of sparse representation is useful as a first stage of noisy signal processing to estimate target content for farther analysis. Generally useful procedure is the sparsest approximation of $\mathbf{f}$ according to eq.(3) that achieves acceptable error ϵ. However, important question is how, knowing the regularity of $\mathbf{f}$ to be approximated, how to derive approximation error bounds for the best approximants within a class of bounded complexity.

To solve the best approximation problem defined by eq.(4), estimation of such approximant set is based on noisy data from the unknown target function to be approximated. The number of approximants should be fixed according to gathered knowledge.

Error bands, the number and kind of approximants may be fixed computationally with Orthogonal Matching Pursuit, Basis Pursuit, Basis Pursuit Denoising, Iterative Thresholding, Compressive Sampling Matching Pursuit, and many other techniques [17], [11]. However, formalized domain knowledge (ontologies) completed with heuristic and experimental procedures give semantic models advising choice of best parameters [19].

A. Nonlinear Approximation

The fundamental problem of approximation theory is to resolve a possibly complicated target function by simpler, easier to compute basis functions called the approximants [6], [16]. Formally, an approximation process can be simply defined in a Hilbert space with inner product $\langle .,. \rangle$ and norm $\|.\|_H$. Let $\{\varphi_i\}_{i=1,2,...}$ be an orthonormal basis for complete $\mathbf{H}$. Each signal $\mathbf{f} \in \mathbf{H}$ can be decomposed on this basis $\mathbf{f} = \sum_{i=1}^{+\infty} a_i \varphi_i$ with the coefficients of orthogonal expansion $a_i = \langle \mathbf{f}, \varphi_i \rangle$.

In case of nonlinear approximation we use the nonlinear space $\mathbf{A}_N$ for expression of $\tilde{\mathbf{f}} \in \mathbf{H}$ as $\tilde{\mathbf{f}}_{(N)} = \sum_{i \in \Lambda} a_i \varphi_i$, where $\Lambda \subset \mathbf{N}$ is a finite set of indexes with the cardinality $\#\Lambda = M \leq N$. M terms are chosen depending on the meaningful features of approximated $\mathbf{f}$.

Nonlinear approximation error $\tilde{\epsilon}_N^2(\mathbf{f}) = \sum_{i \notin \Lambda} |a_i|^2$ is minimal and decays as M increases if Λ corresponds to the M vectors that best correlate to $\mathbf{f}$, i.e. having the largest coefficients of the expansion $|a_i|$. For the set of indexes $\Lambda_r = \{i_j\}_{j=1,2,..,M}$ sorted according to decreasing order of the corresponding coefficients, $|a_j = \langle \mathbf{f}, \phi_{i_j} \rangle| \geq |a_{j+1}|$, $0 < j < M$, we have $\mathbf{f}$ approximation $\tilde{\mathbf{f}}_{(M)} = \sum_{j=1}^{M} a_j \varphi_{i_j}$ with the error $\tilde{\epsilon}_N^2(\mathbf{f}) = \sum_{j=M+1}^{+\infty} |a_j|^2$. The decay rate of $\tilde{\epsilon}_N^2(\mathbf{f})$ as M increases is a measure of approximation efficiency. Consequently, the basis functions of approximation space should represent the most important, precisely characterized and distinguishable features of the target function, according to knowledge platform.

B. Bases

Scalable local bases, especially wavelets are tailor-made for nonlinear approximation because of fast and simple computation, simplified characterization of approximation spaces based on almost unconditional function classes with controlled regularity and transparent strategies of basis selection and target function estimates. A wavelet multiscale representation focuses on localized signal structures with a zooming procedure that progressively reduces the scale parameter. Local signal regularity is characterized by the decay of the wavelet transform amplitude across scales. Singularities are detected and interpreted by following the wavelet domain local maxima at fine scales. Adaptive thresholding of wavelet coefficients makes such representation extremely useful for nonlinear approximation [14], [11].

A nonlinear approximation in a wavelet orthonormal or biorthogonal basis defines an adaptive grid that refines the approximation scale in the neighborhood of the signal singularities. If the target function is smooth in a region of interests a coarse scale of dyadic decomposition is used. For regions where the target function is less smooth the wavelet functions of higher resolution scales are used for the approximation. More accurate allocating terms in a nonlinear strategy depend on signal energy distribution across scales and subbands.

If $\mathbf{f}$ is piecewise regular then few wavelet coefficients are affected by isolated discontinuities and the error decay depends on the uniform regularity between these discontinuities. For smooth wavelet basis with p vanishing moments we have $\tilde{\epsilon}_N^2 = O(M^{-2\alpha})$ for uniformly Lipschitz function $\mathbf{f}$ with $\alpha < p$.

More efficient nonlinear image approximation may be constructed with scaled basis functions whose support shape can be adapted to the orientation and regularity of the object contours. It refers to non-separable wavelet kernels called 2D wavelets with anisotropic dilations, rotations and translations of mother function $\varphi_{m,n,\theta}(\cdot,\cdot)$ oriented by θ. The basic concepts of 2D wavelets use adaptive geometry-based approaches such as wedgelets (beamlets, platelets etc.), or directional frames such as ridgelets, curvelets, contourlets [8].

We found curvelets as natural and flexible extension of tenor wavelets to 2D domain. Curvelet decomposition is a multiscale pyramid corresponding to family of functions with many directions and positions at each length scale, and needle-shaped elements at fine scales. This pyramid contains elements with a very high degree of directional specificity. In addition, the curvelet transform is based on a established anisotropic scaling principle which is quite different from the isotropic scaling of wavelets.

First generation curvelets were based on ridgelets, i.e. continuous functions in the form of $\rho_{m,n,\theta}(x,y) = m^{-1/2}\psi((\cos(\theta)x+\sin(\theta)y-n)/m)$. Ridgelet decomposition is a form of wavelet image analysis in the Radon domain. It solves the problem of sparse approximation of smooth objects with straight edges. But for finer approximation of curved edges one can use a sufficient fine scale to capture curves as almost straight edges. Thus curvelet transform was based on multiscale ridgelets combined with a spatial bandpass filtering operations and subbands splitting into blocks. Second generation curvelets are defined directly in via frequency partitioning without ridgelets. Digital curvelet image decomposition is based on unequally-spaced fast Fourier transforms or the wrapping of specially selected Fourier samples [5].

Dictionary of such wavelet bases is really important for compressing sensing methodology.

IV. Semantics of Sparse Representation

Typical optimization criteria of signal expansions in atom dictionary are based only on l–norms and metrics in constructed solution space. However, medical applications pay special attention for semantic models of compactly distributed pathology signatures. Algorithms should be adaptively adjusted to predicted meaning of extension coefficients.

A. Estimation by Thresholding

Estimation of masked signal is often necessary condition for effective detection of diagnostic information. Basic scheme of nonlinear approximation applies thresholding function with zeroing the expansion coefficients of magnitude less than the threshold value τ. For the source signal $\mathbf{s} = \mathbf{f} + \eta$ (with masking background η), we can estimate target disease function $\mathbf{f}$ by selection of the coefficients $a_i^{(\mathbf{s})}$ with thresholding function $d(\cdot)$ as follows: $\tilde{\mathbf{f}} = \sum_{i=1}^N d(a_i^{(\mathbf{s})})\varphi_i$. Considering specific coefficient meaning dependent on context appearance c_i, thresholding is realized by more general formula

$$\tilde{\mathbf{f}} = \sum_{i=1}^{N} d(a_i^{(\mathbf{s})}, \Sigma_i)\varphi_i \tag{5}$$

where $\Sigma_i = \Sigma(c_i)$ is modeled according to domain knowledge.

The thresholds are matched adaptively considering coefficients distributed across scales, subbands and orientations keeping only expansion domain transients coming from the disease function, according to heuristic model of sparse disease signatures.

B. Patch Domain Modeling

To make representation effectively fixed to local semantics of diagnostic image, patch-based image processing is useful. Moreover, optimization is often simpler and more accurate because of more distinct and unique local criteria. Local patches are block contexts of each or selected pixels to be decomposed according to multi-component sparse criteria [18]. An image $\mathbf{f} \in \mathbf{R}^N$ of N pixels is processed by extracting patches $p(f_x)$ of size $\nu \times \nu$ around selected pixel position $x \in \{1,\ldots,X\}$, $X \leq N$:

$$\forall_l \in \mathcal{L} = \{-\nu/2+1,\ldots,\nu/2\}^2, p_l(f_x) = f_{x+t} \tag{6}$$

Thus, the patch $p(f_x) = \{p_l(f_x)\}_{l\in\mathcal{L}}$ is handled as a vector of size $n = \nu^2$. Next, linear modeling operator $\mathcal{P} : \mathbf{f} \to \{p(f_x)\}_x$ extracts all the patches from an image.

Each patch is approximated by M-sparse representation of local features as

$$\tilde{f}_x = \sum_{j=1}^{m} w_{j_k}\varphi_j \tag{7}$$

where each $w(j_k) \in \mathbf{R}$ is sorted coefficient associated to the atom from a dictionary $\mathcal{D} = \{\varphi_j\}$. Ordered patch-based approximation allows extraction of local image features important for diagnosis.

Patch domain data modeling is useful for image texture analysis, synthesis, inpainting and classification [18]. We decided to adjust patch-based image analysis for ischemic tissue recognition.

C. Multi-component Data Representation

One possible realization of data content analysis is morphological component analysis (MCA) taking into account semantics of represented information. MCA was designed to separate several components which have different morphologies through decomposition of the signal into building blocks [3]. MCA decomposition exploits morphological diversity of selected data features associated to respective atoms of used dictionary. Fundamental assumption is that for every signal target behavior to be separated, there exists in dictionary a basis enables its sparse and as sparse as possible representation.

Let a given signal $\mathbf{f}$ be a sum of K components having different morphologies μ_k and meanings Σ_k: $\mathbf{f} = \sum_{k=1}^{K} f_k(\mu_k, \Sigma_k)$ according to available domain knowledge. We assume that a dictionary of bases (sets of respective atoms) $\mathcal{D} = \{\Phi_1, \ldots, \Phi_K\}$ exists such that for each k component f_k is satisfactorily sparse in respective Φ_k. It means that $\forall j \neq k, ||\Phi_k^T f_k||_0 < ||\Phi_k^T f_j||_0$ and $||\Phi_k^T f_k||_0 << ||f_k||_0$, where $||f||_0$ denotes the l_0 pseudo-norm of the vector (*de facto* the number of nonzero coefficients of f).

To make the problem solution accurate and useful for exemplar stroke disease modeling, two semantic components of density distribution f_d and tissue texture characteristics f_t were assumed to be estimated.

We have $\mathbf{f} = f_d + f_t$ and assume heuristically and empirically determined basis Φ_d and Φ_t. The components f_d and f_t are estimated by solving the following constrained optimization problem:

$$\min_{f_d, f_t} \{||\Phi_d^T f_d||_1 + ||\Phi_t^T f_t||_1\} \text{ s.t. } ||\Sigma_{\mathbf{f}} - \Sigma_{f_d} - \Sigma_{f_t}||_2 < \Sigma_\sigma \quad (8)$$

where Σ_σ is acceptable approximation of $\mathbf{f}$ meaning $\Sigma_{\mathbf{f}}$ by sum of component meanings taking into account imaging modality and case conditioning. The algorithm of respective component estimation relies on general concept of an iterative and interactive (according to radiologist requirements) alternate matched decomposition and thresholding scheme originated in the method proposed by [22]. Current estimate of f_d at iteration m, $\tilde{f}_d^{(m)}$ is fixed by:

- nonlinear approximation of $f_d^{(m)}$ by
  - residual representation of $r_d^{(m)} = \mathbf{f} - \tilde{f}_t^{(m-1)}$,
  - thresholding of the expansion coefficients $\mathbf{a} = \Phi_d^T r_d^{(m)}$ according to semantic characteristics of density distribution Σ_{f_d}, and reconstruction of $\tilde{f}_d^{(m)} = \sum_i d(\mathbf{a}, \Sigma_{f_d}) \Phi_{di}$,
- iterative and alternatively interactive regulation of the threshold, i.e. approximation rate, to extract accurately useful features of disease.

V. Usefulness of Sparse Representation

Accurate and semantic representation is useful if the method of content extraction is adjusted to human perception conditioning and interpretation procedure with responsibility of making diagnostic decisions. It is highly specific condition of sparsity application.

A. Ischemic Stroke Diagnosis

Usefulness of sparse representation was confirmed by effective extraction of hidden diagnostic information in case of acute ischemia detection. Accurate early diagnosis of hyperacute ischemic stroke is critical due to limited timing of applicable thrombolytic therapy. However, clinical phenotype is today obligatory completed with neuroimaging. It should allow identification of patients with acute stroke and selection of suitable treatment. Computed tomography (CT) as an imaging method of first choice is used for efficient identification of patients with acute stroke. Consequently, it allows selection of suitable treatment, exclusion of intracerebral hemorrhage and determination of etiology as well as follow-up therapy and its possible complications.

A CT image of the brain in acute stroke patients is not self-evident. Reading of CT needs training and additional knowledge about the physical conditions of image contrast distribution with noise and artifacts-caused limitations [24]. Significant CT number instability masks very subtle hypodense changes within ischemic region making pathology detection extremely difficult for many cases of irreversible infarcts. Thus, a challenge for CAD applications is making hypodensity distribution more distinct to reveal the diagnostic content and improve accurate recognition of infarct signatures.

B. SDP Implementation

The sparse model of brain tissue density distributed across image was used to extract subtle, diagnostically important hypodensic changes. Proposed procedure computes sparse representation of ischemic stroke patterns to extract hidden pathology manifestation according to accuracy, semantics and usefulness conditions.

According to assumed CT scan model of disease content, two component data representation consists of mean density distribution extracted for hypodensity perception improvement and tissue texture distinction used for completed verification of ischemia in automatic procedure.

Accuracy of sparse hypodensity representation was provided by nonlinear approximation with heuristically selected curvelet signal expansion and adapted thresholding procedure of *waveshrink*.

Curvelets provide an essentially optimal representation of hypothetical density target function f which is C^2 (twice continuously differentiable) except for discontinuities along C^2 curves. The nonlinear approximation error obeys $\tilde{\epsilon}_N^2(f) = O(M^{-2}(\log M)^3)$ and is optimal in the sense that no other representation can yield a smaller asymptotic error with the same number of terms [6].

Satisfying results of disease estimation were achieved with curvelets *waveshrink* defined by $d_\tau^{(wavesh)}(a_i^{(s)}, \Sigma_i) = \frac{a_i^{(s)}}{\left\|a_i^{(s)}\right\|} \cdot (\left\|a_i^{(s)}\right\| - \tau_i)_+$ for complex coefficients with magnitudes $\left\|a_i^{(s)}\right\| > \tau_i$. Threshold $\tau_i = \tau(\Sigma_i)$ is fitted to scale and subband data characteristics. Moreover, τ_i is used for interactive selection of morphological content complexity (see Fig. 1).

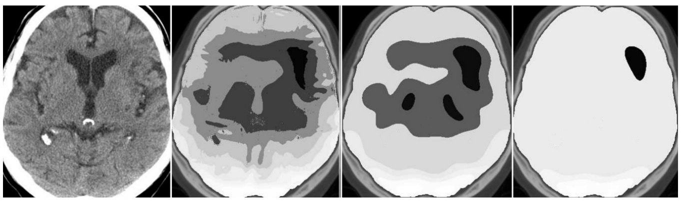

Fig. 1. Diagnostic content extracted from source CT scan (left). Hypodensity distribution across ROI was approximated progressively with 6, 4 and 3 components, respectively (left to right). The darkest component asymmetrically distributed across brain axis was considered to be potential indication of central region of ischemia.

Additionally, tensor wavelet-based perfecting of image reconstruction was used for smoothing approximated content that contains stroke-oriented components. Fundamental wavelet decomposition requires the filters to be finite impulse response and linearly phased to form orthogonal filter banks (FBs). However, only non-smooth Haar filter fulfill such requirements. We constructed orthogonal FB with softening perfect reconstruction (PR) condition controlling the distortion introduced in data processing to extract basic (lower frequency) signal content. Spline non–PR FB was defined by low pass filter $h = [1/4, 1/2, 1/4]$ and mirror high pass filter $g = [-1/4, 1/2, -1/4]$ [20]. As result, directional characteristics of extracted components was more continuous and smooth. Consequently, reconstructed information was assessed by observers to be nicer, more natural and reliable.

Automatic recognition of textural tissue features was applied as additional indicator of ischemia regions to increase usefulness of sparse patterns for real diagnosis aid. Stroke detection was verified on a set of selected CT scan regions susceptible to stroke.

Specific features were extracted in wavelet domain – nearly symmetrical symlets and two scales of decomposition were applied for patches of 50×50. Set of features contains an energy of approximation related to the energy of details, and distribution of detail energy and entropy across scales for maximum magnitude details of successive scales. Moreover, additional features were estimated including compactness of energy in sparse curvelet and contourlet domain (20% of hard thresholded coefficients) in relation to the energy distribution in source image domain. Supervised classification based on SVM with radial kernel, regularization and crossvalidation was used.

Contourlet image transform was defined in a discrete domain as multiresolution and multidirectional expansion with contour segments derived from non-separable, pyramidal directional filter banks [7]. Contourlets-based sparse representation for two-dimensional piecewise smooth signals that resemble images satisfy the anisotropy scaling relation for curves. Contourlets approximate signals which are C^2 with rate $\tilde{\epsilon}_N^2(f) = O(M^{-2}(\log M)^3)$, similarly to curvelets. Through more flexible adjustment of filter characteristics and scale-subband decomposition scheme, contourlet bases were fitted adaptively to specific textural characteristics of brain tissue.

To sum up, adapted and concreted method of stroke disease recognition is as follows:

1) initial region of interests (ROI) conditioning – segmentation of stroke-susceptible regions of brain tissues with locally adaptive region growing and thresholding algorithms with smooth complement of segmented diagnostic areas;
2) providing accuracy of sparse representation – nonlinear approximation of target density distribution by expansion of ROI in curvelet basis and thresholding with *waveshrink* procedure;
3) providing semantics of extracted content, i.e. density distribution:
   a) interactive regulation of progressive disease extraction (see Fig. 2) by *waveshrink* control and adjusting wavelet-based smoothing procedure with spline non–PR FB.
   b) providing usefulness of visualized information – hypodensity expression by display arrangement of processed regions and source scans with greylevel quantization and contrast enhancement according to observer suggestions and semantic content models;
4) complement of morphological content by ischemic tissue distinction in hypodense areas – automatic recognition of ischemic brain tissue with sparse texture classification; a texture dictionary of wavelets, contourlets and curvelets was used.

C. Experimental Verification

Computer-assisted tools realized according to SDP methodology occurred really useful for diagnosis support.

Experimental verification was realized as subjective diagnosis of ischemic stroke completed with automatic recognition of disease tissue. Test database consisted of 123 CT examinations including 105 patients aged 24-92 years (70 years in average) with proved infarction. No direct hypodense signs of hyperacute ischemia were found on positive data sets (follow-up confirmation).

Average diagnosis sensitivity of seven radiologists increased from 0.385 to 0.513 (+38%) with additional preview of extracted hypodensity. Respective specificity decrease was of 0.817 to 0.774 (-5%) because of difficulties in interpretation of increased perceptibility of tissue density changes for less experienced radiologists. In group of four more experienced experts, increase of sensitivity and specificity was 26% and 2%, respectively.

Automatic recognition of ischemic tissue based on sparse image representation was used as a compliment of visual

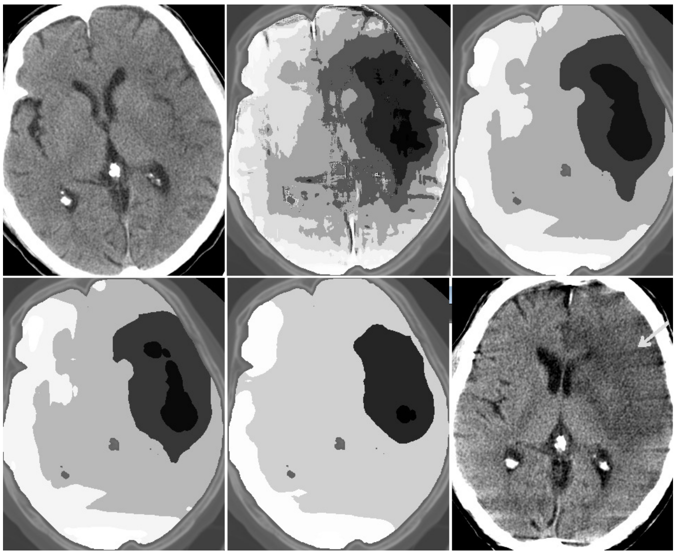

Fig. 2. Exemplary results of SDP realization. Sparse disease patterns were extracted from acute CT scan (top-left) and confirmed by follow-up CT indicating area of infarct (down-right); estimated density distribution (left to right, top-down) from sparse representation consisting of: 0.1%, 0.02%, 0.01% and 0.005% of nonzero coefficients, respectively.

hypodensity expression based on extracted compact disease signatures. Test set of ROIs segmented from 123 cases of stroke database consisted of 236 blocks including 82 positive cases.

To compare, initially only texture features defined in image domain, i.e. typical statistics (standard deviation, kurtosis, skewness, 0-order entropy, energy) and features based on co-occurrence matrix (joint entropy, contrast, correlation, energy, homogeneity), completed with Tamura textural features (coarseness, directionality, contrast) were considered. The recognition results for classified image features were limited to 0.57 of sensitivity and 0.88 of specificity.

The effectiveness of sparse texture characteristics, defined according to SDP paradigm to distinct disease patterns and verified experimentally was 0.72 (+26%) of sensitivity and 0.90 (+2%) of specificity, respectively.

More exhaustive description of stroke detection procedures and other results verifying their usefulness was presented in [21].

VI. Conclusions

Key problem of a semantic gap between the numerical descriptors and human interpretation of images is still challenging problem of assisted radiology. Low level descriptors are not uniquely and explicitly associated to specific meaningful label of abstractional description of medical, image-based knowledge. Sparse data representation, related to specific dictionary atoms selected and fitted according to semantical reasons and models, is potentially susceptible to be recognized as invariant disease patterns. Compact distribution of such scalable "disease image" means new possibility of pathology manifestation understanding for computer-aided diagnosis.

Suggested methodology is aimed at cognitive resonance of estimated sparse descriptors with formalized, structured and often heuristically established medical knowledge platform. Thanks that considered methodology includes estimation of semantic sparse representation of an image that is optimized with medical knowledge platform, the following classification of extracted semantic components is simplified and designed according to specified diagnostic categories.

Achieved results confirmed high potential of semantic sparse models for diagnostic content extraction and recognition of diseases. However, further improvement of proposed SDP methodology is necessary to make semantic sparsity criteria more formalized and computationally unique. Clue problems are: – more effective schemes of nonlinear approximation for target content estimation, – new atoms of multiscale, local, flexible image approximants adjusted invariantly to representative pathology patterns, – investigation of pursuit and thresholding algorithms to investigate optimal representation over overcomplete dictionaries according to diagnostic knowledge criteria. New applications of sparse representation for disease modeling and recognition are necessary to develop reliable tools.

References

[1] N. Ancona, R. Maglietta, and E. Stella, "Data representation in kernel based learning machines," in *Proceedings of IASTED International Conference on Artificial Intelligence and Soft Computing*, September 2004, pp. 129–136.

[2] A. Blum and P. Langley, "Selection of relevant features and examples in machine learning," *Artificial Intelligence*, vol. 97, pp. 245–271, 1997.

[3] J. Bobin, J.-L. Starck, J. M. Fadili, Y. Moudden, and D. L. Donoho, "Morphological component analysis: an adaptive thresholding strategy," *Image Processing, IEEE Transactions on*, vol. 16, no. 11, pp. 2675–2681, 2007.

[4] E. J. Candés, "Compressive sampling," in *Proceedings of the International Congress of Mathematicians*, 2006, pp. 1433–1452.

[5] E. J. Candés, L. Demanet, D. L. Donoho, and L. Ying, "Fast discrete curvelet transforms," Technical Report, Cal Tech, Tech. Rep., 2005.

[6] R. A. DeVore, "Nonlinear approximation," *Acta Numerica*, vol. 7, pp. 51–150, 1998.

[7] M. N. Do and M. Vetterli, "Contourlets," in *Beyond Wavelets*, G. V. Welland, Ed. Academic Press, 2003.

[8] G. V. W. (editor), *Beyond Wavelets*, ser. Studies in Computational Mathematics. Academic Press, 2003, vol. 10.

[9] H. Fhr, L. Demaret, and F. Friedrich, "Beyond wavelets: new image representation paradigms," in *Document and Image Compression*, M. Barni, Ed. CRC Press, 2006.

[10] L. Floridi, "In defense of the veridical nature of semantic information," *European Journal of Analytic Philosophy*, vol. 3, pp. 31–41, 2007.

[11] M. Fornasier and H. Rauhut, "Iterative thresholding algorithms," *Applied and Computational Harmonic Analysis*, vol. 25, no. 2, pp. 187–208, 2008.

[12] M. L. Giger, "Computer-aided diagnosis in medical imaging - a new era in image interpretation," World Markets Research Centre, Tech. Rep., Tech. Rep., 2000.

[13] A. C. Gilbert, M. J. Strauss, and J. A. Tropp, "Algorithms for simultaneous sparse approximation. part i: Greedy pursuit," *Signal Processing, Special issue on sparse approximations in signal and image processing*, vol. 86, pp. 572–588, 2006.

[14] R. C. Guido and J. C. P. (guest editors), "Wavelet-based algorithms for medical problems," *Computers in Biology and Medicine, Special issue*, vol. 37, no. 4, 2007.

[15] S. K. Kinoshita, P. M. de Azevedo-Marquesand, R. R. Pereira, J. A. Rodrigues, and R. M. Rangayyan, "Content-based retrieval of mammograms using visual features related to breast density patterns," *Journal of Digital Imaging*, vol. 20, no. 2, pp. 172–190, 2007.

[16] S. Mallat, *A wavelet tour of signal processing*, 2nd ed. Academic Press, 1999.

[17] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *Signal Processing, IEEE Transactions on*, vol. 41, no. 12, pp. 3397–3415, 1993.

[18] G. Peyre, "Sparse modeling of textures," *Journal of Mathematical Imaging and Vision*, vol. 34, no. 1, pp. 17–31, 2009.

[19] T. Podsiadly-Marczykowska and A. Guzik, "Mammography ontology, model structure, definitions and conception instances," *Bio-Algorithms and Medical Systems*, vol. 1, no. 1, pp. 247–252, 2005.

[20] A. Przelaskowski, P. Bargiel, K. Sklinda, and E. Zwierzynska, "Ischemic stroke modeling: multiscale extraction of hypodense sign," *Lecture Notes in Artificial Intelligence*, vol. 4482, pp. 171–181, 2007.

[21] A. Przelaskowski, K. Sklinda, G. Ostrek, R. Jozwiak, and J. Walecki, "Computer aided diagnosis in hyper-acute ischemic stroke," in *Progress in Neuroradiology 2009*, J. Walecki, Ed. International Scientific Literature, 2009, pp. 69–78.

[22] J.-L. Starck, M. Elad, and D. L. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *Image Processing, IEEE Transactions on*, vol. 14, no. 10, pp. 1570–1582, 2005.

[23] J.-L. Starck, F. Murthag, and J. M. Fadili, *Sparse image and signal processing*. Cambridge University Press, 2010.

[24] R. von Kummer, "The impact of computed tomography on acute stroke treatment," in *Thrombolytic therapy for stroke*, P. Lyden, Ed. Humana Press, 2005, pp. 249–278.