SEMANTIC SUBSPACE LEARNING FOR MENTAL SEARCH IN SATELLITE IMAGES

Dinh-Phong Vo, Hichem Sahbi

{vo,sahbi}@enst.fr
LTCI Telecom ParisTech
46 rue Barrault 75013 Paris

1. INTRODUCTION

With the exponential growth of satellite image collections, mapping services are nowadays emerging (Google Maps, GeoPortail, Bing, etc.). These tools are valuable to explore areas of interest for professional as well as large public purposes (land management, risk assessment, tourism, etc.). Most of these tools are coupled with search engines, based on existing meta-data or precomputed visual indices, and rely on add-hoc functionalities such as zooming and panoramic navigation. When using large scale and complex images, provided by high resolution satellite sensors[1], these basic functionalities become helpless in order to efficiently explore these images. Indeed, navigation becomes systematic and often tedious as the user spends his effort in zooming (in/out) and shifting from one area to another with the only navigation criterion being "geographic proximity" ; this burden is further amplified when meta-data are scarce or unintuitive to the user.

Content based visual search [1, 2] is another paradigm that allows the user, to search for his mental target in a satellite image, using visual queries. This search paradigm includes the query-by-example [3] and relevance feedback [4] both of them require existing (or hand-sketched) examples which may be difficult to obtain.

Image database visualization [1, 5] is an alternative, that relies on mapping data from high to low dimensional spaces where data can be easily spot and explored by the user. Popular mapping techniques include linear principle component analysis, as well as non linear manifold learning (ISOMAP [6], Locally Linear Embedding [7], Laplacian Eigenmaps [8], t-SNE [9], etc.). In spite of their popularity, the success of these mapping techniques in image database visualization is limited to datasets with well defined semantics such as poses in faces, or deformations in digits [6, 7, 8, 9, 10, 11, 12]; as these techniques are totally unsupervised, their application to generic databases [5, 13, 14] produces "semantic-less" dimensions which are difficult to interpret and explore.

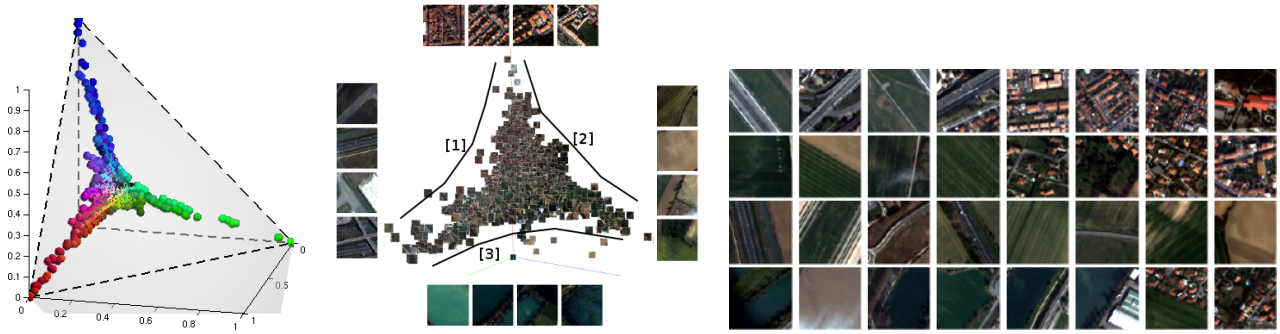
In this paper, we address this issue and we introduce a novel mental search algorithm, for satellite images, based on semantic subspace learning. The latter is designed using a novel principle, that unmixes semantics from satellite image regions and maps them from an initial ambient space (related to low level visual features including texture, color and shape) to an output space spanned by a well defined semantic basis. As will be shown, we cast this problem as convex quadratic programming (QP) optimization, constrained in a simplex spanned by few (pure) semantic endmembers. The advantages of the proposed approach are twofold; on the one hand, it significantly reduces the dimensionality of the input space (which is difficult to explore/visualize), and on the other hand, it learns features which are semantically interpretable, i.e., their values are highly correlated with the defined semantics. Thereby, searching for a mental target simply reduces to scanning and targeting data according to their coordinates in the learned semantic subspace.

In the remainder of this paper, we consider the following notation: $\mathbf{A}_k$ (resp. $\mathbf{A}_{.k}$) denotes the k^{th} row (resp. column) of a given matrix $\mathbf{A}$ while $\mathbf{A}_{ki}$ denotes the i^{th} entry of its k^{th} row. We also denote $\mathbf{A}'$ as transpose of $\mathbf{A}$. Other notations will be introduced as we go through different sections of the paper which is organized as follows: Section 2 describes in detail our proposed method as well as its efficient optimization procedure. Section 3 shows its application to interactive visualization and exploration of satellite images. We conclude in Section 4 while providing a possible extension for a future work.

2. SEMANTIC SUBSPACE LEARNING

In this contribution, we are interested in learning a particular low dimensional subspace. The underlying assumption is: the probability distribution generating the input data admits a density with respect to the canonical measure on a sub-manifold of the Euclidean input space. The goal of our method is to define a mapping that preserves the local distances while capturing a global topology. The latter is defined by the dynamic of variation of data (through different intrinsic dimensions) which should be consistent with the semantics assigned to these dimensions (for instance progressive variation from rural to urban areas, from sea to land, etc.; see Fig. 1(c)).

[1] with more than $10,000 \times 10,000$ pixels.



(a) The learned embedding of RGB toy data, with three pure endmembers in the simplex.

(b) This figure shows embedding of satellite image patches in a subspace of 3 (semantic) dimensions corresponding to *building*, *road*, *vegetation*. Examples used to learn the 3 dimensions as well as the 4^{th} one (i.e., *water*) are shown.

(c) Each row in this figure shows a progressive variation of the "semantic" of image patches from: "*road* to *building*" (row 1 corresponding to path 1 in Fig. 1(b)), "*field* to *building*" (row 2 corresponding to path 2 in Fig. 1(b)), "*road* to *field*" (row 3 corresponding to path 3 in Fig. 1(b)) and "*water* to *building*" (row 4, not shown in Fig. 1(b)).

Fig. 1. The learned embedding of toy data and real satellite image using our proposed method

Let $\mathbf{X} \in \mathbb{R}^{d\times n}$ be a matrix of input data corresponding to n non-overlapping image patches taken from a given satellite image and $\mathbf{Y} \in [0,1]^{K\times n}$ its underlying membership matrix; here K corresponds to the number of semantics[2] fixed to $K = 4$ in practice. A given entry $\mathbf{Y}_{ki}$, of a membership vector $\mathbf{Y}_{.i}$, is strictly positive iff the k^{th} semantic is present into $\mathbf{X}_{.i}$ according to some ground truth (see Fig. 1(b)). When only one entry of $\mathbf{Y}_{.i}$ is positive and also equal to 1, then $\mathbf{Y}_{.i}$ is referred to as *pure* membership vector (also called pure endmember). In the remainder of this paper, we assume that only the pure membership vectors in $\mathbf{Y}$ are known while the other vectors in $\mathbf{Y}$ are unknown and set to $\mathbf{0}$; the all-zeros vectors.

Our goal is to learn an embedding $\mathbf{Z} \in [0,1]^{K\times n}$ where each entry $\mathbf{Z}_{ki}$ corresponds to a mapping of $\mathbf{X}_{.i}$ into the k^{th} semantic dimension; a high value of $\mathbf{Z}_{ki}$ indicates that the k^{th} semantic is present into $\mathbf{X}_{.i}$ with a high probability and vice-versa. In order to find $\mathbf{Z}$, we introduce the following optimization problem

$$\begin{array}{ll}\min\limits_{\mathbf{Z}\in\mathbb{S}} & \frac{1}{2}\mathrm{tr}\left(\mathbf{Z}\mathbf{L}\mathbf{Z}'\right) \\ \text{s.t} & \mathbf{Z}\mathbf{C} = \mathbf{Y}\end{array}, \tag{1}$$

here tr stands for the matrix trace operator. In the above objective function, elements in $\mathbf{Z}$ are taken from a unit $(K-1)$-simplex, that belongs to the positive orthant, and spanned by the canonical basis of $\mathbb{R}^K$, i.e., $\mathbb{S} = \{\mathbf{Z} \in \mathbb{R}^{K\times n},\ \mathbf{Z} \geq \mathbf{0}_{K\times n},\ \mathbf{Z}'\mathbf{1}_K = \mathbf{1}_n\}$ with $\mathbf{1}_K$, $\mathbf{1}_n$ (resp. $\mathbf{0}_{K\times n}$) being the all-ones vectors (resp. all-zeros matrix). This simplex condition $\mathbf{Z} \in \mathbb{S}$ is essential; given a patch, this model measures the abundance of each semantic into that patch so these abundances should be positive and their sum equal to 1.

[2]A semantic is any pure concept such as "road", "vegetation", "building", "water", "mountain", "desert."

The main term, in (1), is a regularizer that ensures similar embedding for neighboring data in $\mathbf{X}$ and $\mathbf{L}$ is a normalized graph Laplacian [15]. The latter is defined as $\mathbf{L} = \mathbf{I}_n - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$, with $\mathbf{A} \in \mathbb{R}^{n\times n}$ being a symmetric similarity matrix, with each entry $\mathbf{A}_{ij} = \exp\left(-\|\mathbf{X}_{.i} - \mathbf{X}_{.j}\|_2^2/\sigma^2\right)$, and $\mathbf{D} = \mathrm{diag}(\mathbf{A}\mathbf{1}_\mathrm{n})$. The constraint $\mathbf{Z}\mathbf{C} = \mathbf{Y}$, in (1), guarantees that the embedding $\{\mathbf{Z}_{.i}\}_i$ related to known (pure) membership vectors coincides with the vertices of the simplex; in this expression $\mathbf{C}$ is the diagonal $n \times n$ matrix for which the i^{th} diagonal element is fixed to 1 for a known membership vector and 0 for an unknown one. Notice that, per construction, different pure membership vectors (i.e., vertices of the simplex), as well as the underlying semantics should be mutually uncorrelated.

It is easy to see that the objective function in (1) can be rewritten as $\frac{1}{2}\sum_k \mathbf{Z}_k\mathbf{L}\mathbf{Z}_k'$ which is the sum of constrained QP problems each one is convex (as $\mathbf{L}$ is positive semi-definite and constraints in Eq. 1 are linear), and can be solved by chunking according to Algorithm 1. This algorithm proceeds by randomly selecting, for each iteration, "active" and "inactive" variables and *very efficiently* solving the problem w.r.t the "active" variables while keeping the "inactive" ones unchanged. As the problem is globally convex, convergence to a unique global solution is guaranteed.

We run the proposed algorithm on toy data which consist in 1,000 points randomly sampled from a 3D RGB cube, of unit side, including three pure endmembers; red $(1\,0\,0)'$, green $(0\,1\,0)'$ and blue $(0\,0\,1)'$. The learned embedding, in Fig. 1(a), shows gradual change of color, in the span of the three endmembers of the simplex (i.e., red, green, blue).

At each iteration of this algorithm, active variables $\alpha = (\mathbf{Z}_{ka}\ \mathbf{Z}_{\ell a})$ should be jointly optimized in order to guarantee the simplex condition in Eq. 1 (i.e., $\mathbf{Z} \in \mathbf{S}$); if one updates $\mathbf{Z}_{ka}$, $\mathbf{Z}_{ka}$ separately, then this condition cannot be satisfied.

Input: Patches in $\mathbf{X}$ and membership vectors in $\mathbf{Y}$.
Output: Learned semantic maps $\mathbf{Z}^*$.

Set $\mathbf{Z}^{(0)} = \frac{1}{K}\,\mathbf{1}_K \times \mathbf{1}'_{\mathbf{n}}$ and $t \leftarrow 0$; // This corresponds to an initial feasible solution.

repeat

Pick (k, ℓ) randomly from $\{1, \ldots, K\} \times \{1, \ldots, K\}$, with $k \neq \ell$.

Pick randomly a subset of "active" indices $a \subseteq \{1, \ldots, n\}$, and let $\bar{a} = \{1, \ldots, n\} \backslash a$, with $|a| \ll |\bar{a}|$.

Given a matrix $\mathbf{A}$, define a sub-matrix $\mathbf{A}_{ka} = (\mathbf{A}_{ki})_{i \in a}$. Let $m = |a|$ and $\alpha = (\mathbf{Z}_{ka}\ \mathbf{Z}_{\ell a})$.

Solve the following QP // using **quadprog** in MATLAB Optimization Toolbox

$$\begin{array}{lll} \alpha^* \leftarrow \underset{\alpha \geq \mathbf{0}_{2m}}{\arg\min} & \frac{1}{2}\alpha(\mathbf{I}_2 \otimes \mathbf{L}_{aa})\alpha' & + (\mathbf{Z}^{(t)}_{k\bar{a}}\mathbf{L}_{\bar{a}a}\ \ \mathbf{Z}^{(t)}_{\ell\bar{a}}\mathbf{L}_{\bar{a}a})\alpha' \quad ; \text{// } \otimes \text{ is the Kronecker tensor product.} \\ \text{s.t} & \alpha(\mathbf{1}_2 \otimes \mathbf{I}_m) & = \mathbf{1}'_m - \sum_{p \neq k,\ell} \mathbf{Z}^{(t)}_{pa} \quad ; \text{// } \mathbf{I}_m \text{ is the } m \times m \text{ identity matrix.} \\ & \alpha(\mathbf{I}_2 \otimes \mathbf{C}_{aa}) & = (\mathbf{Y}_{ka}\ \mathbf{Y}_{\ell a}) \end{array} \tag{2}$$

update

$$\begin{array}{lll} (\mathbf{Z}^{(t+1)}_{ka}\ \mathbf{Z}^{(t+1)}_{\ell a}) & \leftarrow & \alpha^* \quad \text{and} \quad (\mathbf{Z}^{(t+1)}_{k\bar{a}}\ \mathbf{Z}^{(t+1)}_{\ell\bar{a}}) \leftarrow (\mathbf{Z}^{(t)}_{k\bar{a}}\ \mathbf{Z}^{(t)}_{\ell\bar{a}}) \\ \mathbf{Z}^{(t+1)}_{pi} & \leftarrow & \mathbf{Z}^{(t)}_{pi}, \quad \forall p \in \{1, \ldots, K\} \backslash \{k, \ell\}, \ \forall i \in \{1, \ldots, n\} \end{array} \tag{3}$$

Set $t \leftarrow t + 1$

until $\left\|\mathbf{Z}^{(t+1)} - \mathbf{Z}^{(t)}\right\|_2^2 \leq \epsilon$;

$\mathbf{Z}^* \leftarrow \mathbf{Z}^{(t)}$

Algorithm 1. Optimization algorithm

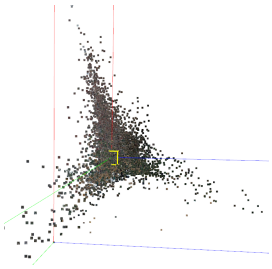
3. SATELLITE IMAGE NAVIGATION WITH SPACIOUS

We have developed an interactive visualization software called Spacious (Fig. 2(b)), used to plug the embedding of patches in the semantic subspace learned by our method, and to help analysts exploring a given large satellite image efficiently. Spacious provides dual view to arbitrary satellite image: the *map view* shows the image itself and the *manifold view* is the embedding produced by Eq. 1. While the *map view* provides standard look to some geographical region on Earth, the *manifold view* exhibits a mental picture of the map and allows the user to define his visual objects of interest in order to interact with. Spacious will then synchronize the two views so that the user can rapidly target objects of interest. Beneath Spacious is the graphics engine of PartiView [3] allowing the software smoothly visualize ten thousands of image patches. In the following, we describe the processing pipeline in order to launch interactive visualization with Spacious.

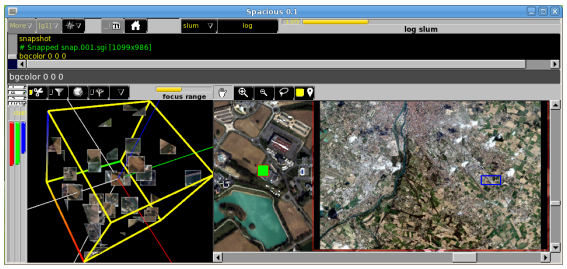
Initially, a given satellite image is divided into patches using a rectangular grid with appropriate resolution, i.e., the patch size should be adjusted according to the image resolution. In our experiment, the original satellite image, including $6,876 \times 7,265$ pixels, is partitioned into approximately $12,000$ patches of 64×64 pixels. As mentioned earlier, we define $K = 4$ usual semantics ("building", "road", "vegetation" and "water") and we select 15 patch examples per semantic (see Fig. 1(b)). The underlying membership vectors in $\mathbf{Y}$ are set accordingly; again we assume that each selected example includes only one semantic so its corresponding membership vector is pure. Selected examples, even though few, are representative enough and cover at some extent the diversity of these four semantics. Prior to run our subspace learning algorithm, we process each patch in order to extract its visual features (i.e., matrix $\mathbf{X}$) and compute the underlying graph Laplacian $\mathbf{L}$. In practice, we combine color histograms as well as WLD features [16] in order to describe each patch; the good performance of this combination was also observed in neighboring tasks in land cover mapping.

Using the learned semantic subspace, the user can explore the manifold view using supporting tools including different navigation modes (orbital, flight, rotate, translate), subset selection, subset filtering, etc. Additionally, the functionality *dimension switching* is useful in case the embedding has more than three dimensions, i.e $K > 3$. When plugged into Spacious, the learned subspace meets our design criteria mentioned in Section 1. Fig. 1(b) shows a mental picture of the input satellite image in which 3 (out of 4) semantics are illustrated. As we traverse patches, in the learned dimensions, we observe a gradual variation of the semantic from: *vegetation* to *building*, from *building* to *road*, and from *road* to *vegetation* (see. Fig. 1(c)) whereas in the span of these dimensions, patches mix several semantics. For instance, a mental query like "find a location with the same proportions of vegetation, road, and building" can be easily expressed by sliding the coordinates of a selector (bounding box center) towards the barycenter of the simplex (see example in Fig. 2). As soon as the selector (yellow cube in Fig. 2(b)) moves to the desired

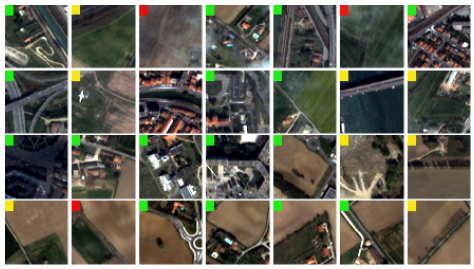
[3] http://www.haydenplanetarium.org/universe/partiview



(a) A mental query like "find a location with the same proportions of vegetation, road, building" is expressed by sliding the center of the selector towards the barycenter of the simplex.



(b) Screenshot of Spacious at the time of querying in Fig. 2(a). The left-hand side corresponds to clipped & zoomed *manifold view* that clearly shows visually similar patches, and on the right is the zoomed *map view* with one candidate shown in green square dot and its absolute location in the map preview. It is easy to see that this location simultaneously contains *building*, *field*, and *road*.



(c) A list of image patches within the cube of Fig. 2(b) in which a green dot means all three semantics appear; a yellow dot means two out of three semantics appear; a red dot means just one semantic appears. From these results, the semantics in these patches become disproportionate as we move away from the center of the selector.

Fig. 2. This figure shows an example of a mental query and screenshot of our interactive visualization software "Spacious".

location(s) on the manifold view, the map view will promptly update those location(s) on the original satellite image, giving the user a possibility to further refine the search. A video demonstration of this mental query example is also available online at `http://perso.telecom-paristech.fr/~vo/page2.html`.

4. CONCLUSION

In this paper we introduced a new approach, for mental satellite image search and visualization. Throughly, we addressed the difficulties when visualizing high dimensional data, and we proposed a semantic subspace learning approach for effective exploration of large scale satellite image data. Our formulation is based on a quadratic programming algorithm that easily extend to large scale data. By plugging the embedding solution of this algorithm into Spacious, users can specify objects of interest, as mixtures of predefined semantics in the learned subspace, and retrieve their targets.
As a future work, we are currently investigating the use of other low level features as well as the combination of our semantic subspace learning method with relevance feedback.

5. REFERENCES

[1] Daniel Heesch, "A survey of browsing models for content based image retrieval," *Multimedia Tools Appl.*, vol. 40, no. 2, pp. 261–284, 2008.

[2] Jiang Li and R.M. Narayanan, "Integrated spectral and spatial information mining in remote sensing imagery," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 42, 2004.

[3] M. Zloof., "Query by example," *AFIPS*, vol. 44, 1975.

[4] I-J. Cox, M. Miller, T-P. Minka, T. Papathomas, and P. Yianilos, "An optimized interaction strategy for bayesian relevance feedback," in *IEEE Conf. Computer Vision and Pattern Recognition*, 1998.

[5] G. Schaefer, "A next generation browsing environment for large image repositories," *Multimedia Tools and Applications*, vol. 47, pp. 105–120, 2010.

[6] J-B. Tenenbaum, V. de Silva de Silva de Silva, and J-C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[7] S-T. Roweis and L-K. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," *Science*, vol. 290, pp. 2323–2326, 2000.

[8] M. Belkin and P. Niyogi, "Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering," in *NIPS*, 2001, pp. 585–591.

[9] Laurens V D Maaten and Geoffrey Hinton, "Visualizing Data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.

[10] John A. Lee and Michel Verleysen, *Nonlinear Dimensionality Reduction*, Springer, 2006.

[11] F.S. Tsai, "Comparative study of dimensionality reduction techniques for data visualization," *Journal of Artificial Intelligence*, vol. 3, 2010.

[12] C. Shan, S. Gong, and Peter W Mcowan, "Appearance Manifold of Facial Expression," .

[13] Y. Rubner and C. Tomasi, *Perceptual Metrics for Image Database Navigation*, Springer, 2001.

[14] Xiaofei He and Partha Niyogi, "Locality preserving projections," 2002.

[15] Alex J. Smola and Risi I. Kondor, "Kernels and regularization on graphs," in *COLT*, 2003, pp. 144–158.

[16] J. Chen, "Wld: A robust local image descriptor," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, 2010.