MULTI-RESOLUTION SPATIAL UNMIXING FOR MERIS AND LANDSAT IMAGE FUSION

J. Amorós-López, L. Gómez-Chova, L. Guanter, L. Alonso, J. Moreno and G. Camps-Valls

Image Processing Laboratory (IPL). Universitat de València, Spain.
julia.amoros@uv.es

1. INTRODUCTION

Some current and planned Earth Observation missions provide global coverage from medium and high spatial resolution instruments with typical revisiting times around $2-3$ days or $8-16$ days. However, users continuously demand higher spatial resolution images with better temporal and/or spectral resolution that cannot be provided by a single instrument. Image fusion techniques allow us to combine information from different sensors into a single image product. In particular, high spatial resolution images can be merged with data from high temporal resolution systems to provide the continuous time-series at high spatial resolution that are required by many operational applications such as land cover mapping, crop phenology, emergency monitoring, or natural resource monitoring.

In this context, we propose an image fusion approach based on a multi-resolution and multi-source unmixing. The methodology proposed obtains a composite image with the spatial resolution of the higher resolution image (*downscaling*) while retaining the spectral and temporal characteristics of the medium spatial resolution image. The approach is tested in the specific cases of MERIS/ENVISAT and Landsat/TM cases, but is general enough to be applied to other sensor combination.

2. PROPOSED DOWNSCALING METHODOLOGY

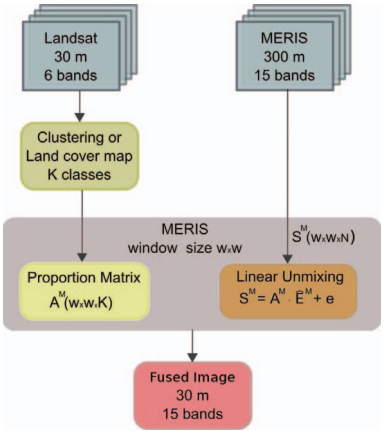


Fig. 1: Proposed scheme for MERIS and Landsat image fusion.

The general idea of the proposed method is to combine the concepts of image fusion and spatial unmixing. The goal is to essentially generate images with the spatial resolution of Landsat 5 TM (30m, 7 channels in $0.45-2.35\mu$m) and the MERIS spectral resolution (FR 300m, 15 channels in $412.5-900$ nm). The processing scheme is illustrated in Fig. 1, and can be summarized as follows. The proposed fusion algorithm is based on the work in [1,2], but conveniently modified to accommodate some important characteristics of the particular problem. First, the Landsat image is segmented into a number of classes K by using an unsupervised clustering algorithm, and the membership of each Landsat pixel to the clusters is also calculated (i.e. Landsat pixels are considered as mixed pixels rather than as pure land covers). Next, these posterior probabilities at Landsat resolution are used to get the abundance or proportion for each MERIS pixel. Then, each MERIS pixel is unmixed in a window by the inversion of a system of linear mixture equations, i.e. spatial instead of spectral unmixing. A regularization term is added to the cost function to limit the large deviations of the unmixed pixels from natural spectral prototypes for each class. Finally, the fused image is obtained by assigning the estimated spectra to the corresponding Landsat pixels, which are estimated from the MERIS endmembers of each land cover weighted by the Landsat posterior class (land-cover) probabilities. The use of these posteriors from the Landsat image guarantee the spectral variability of the land cover classes within each MERIS pixel footprint in the fused image. In the following, we analyze the main steps of the algorithm in further detail.

2.1. Soft clustering of the high resolution image

The first step in the algorithm consists of an unsupervised classification of the Landsat image. We selected the algorithm of self-organizing maps (SOM) [3] due to several reasons: 1) it carries out a topology preserving mapping from the D-dimensional

This paper has been partially supported by the Spanish Ministry for Education and Science under projects TEC2006-13845, CSD2007-00018 and AYA2008-05965-C04-03.

input space onto a two-dimensional grid ; 2) it gives better initialization than the k-means algorithm for mixture density modeling since it gives better generalization to the test data; and 3) it produces probabilistic (soft) outputs to be used in the next steps of the proposed method. Other clustering algorithms could be equally used, such as Gaussian mixture models or graph-cuts though. The method requires the definition of a number of clusters or classes K to be found in the image.

Hereafter, superscript (M) refers to MERIS pixels and superscript (L) to Landsat pixels. The soft clustering approach yields a relative membership function to each pixel in the clustering solution. We modeled this as a monotonic decreasing function [4], $A_{ik}^{(L)} = \frac{(\|S_i^{(L)} - \boldsymbol{\mu}_k\|_{\boldsymbol{\Sigma}}^2)^{-2/(m-1)}}{\sum_{k=1}^{K}(\|S_i^{(L)} - \mu_k\|_{\boldsymbol{\Sigma}}^2)^{-2/(m-1)}}$, where $S_i^{(L)}$ is the Landsat pixel i, $\boldsymbol{\mu}_j$ is the cluster centroid k, m is a parameter in the range $[1, \infty)$ but good results are typically obtained in $[1.1, 5]$, and Σ is a positive definite covariance matrix implementing the Mahalanobis distance.

2.2. Sliding-window spatial unmixing of the low spatial resolution image

The proportions of land-cover classes in a MERIS pixel is estimated as the average of the previously estimated class proportions of Landsat pixels contained in the MERIS pixel footprint $\mathcal{P}$: $A_k^{(M)} = \frac{1}{|\mathcal{P}|}\sum_{i\in\mathcal{P}} A_{ik}^{(L)}$. The MERIS image is processed in windows of size $w \times w$ centered in pixel $S_i^{(M)}$. In each MERIS window, we run a linear spatial unmixing by solving the linear equations system in matrix notation,

$$\underbrace{\mathbf{S}^{(M)}}_{w^2\times N} = \underbrace{\mathbf{A}^{(M)}}_{w^2\times K} \cdot \underbrace{\hat{\mathbf{E}}^{(M)}}_{K\times N} + \underbrace{\mathbf{e}}_{w^2\times N} \tag{1}$$

where $\mathbf{S}^{(M)}$ are the actual MERIS spectra in the window $\mathcal{W}$, $\hat{\mathbf{E}}^{(M)}$ are the estimated pure spectra in the window for each class, K is the number of classes, N is the number of MERIS bands, and $\mathbf{e}$ are the errors or residuals of the linear model. Note that, unlike standard spectral unmixing, we are estimating endmembers rather than abundances. Therefore, this problem is solved individually for each MERIS spectral band by optimizing:

$$\hat{\mathbf{E}}^{(M)} = \min_{\hat{\mathbf{E}}} \left\{ \sum_{i\in\mathcal{W}} \left[S_i^{(M)} - \sum_{k=1}^{K} A_{ik}^{(M)} \hat{E}_k^{(M)} \right]^2 \right\} \tag{2}$$

2.3. Regularized linear unmixing

Note that the unmixing process could be carried out for the whole image in one shot, but this would make all pure spectra of each class to collapse to the average one, hence disregarding the spectral variability. Besides, in this case, the number of classes needed for characterizing the scene adequately would largely increase ($K > 100$ classes are generally needed for heterogeneous images). The sliding window approach allows us to obtain more variability for a given land-cover but, unfortunately, adds one more free parameter to study, the optimal window size. The window should contain enough MERIS pixels for the sake of a robust system inversion, but also, the window should allow us to correctly characterize the pure spectra present in the MERIS pixel for the sake of good fusion quality. On the contrary, the size of the window should be small enough to reflect the variability of the low resolution image (Landsat) to avoid the aforementioned spectral collapse.

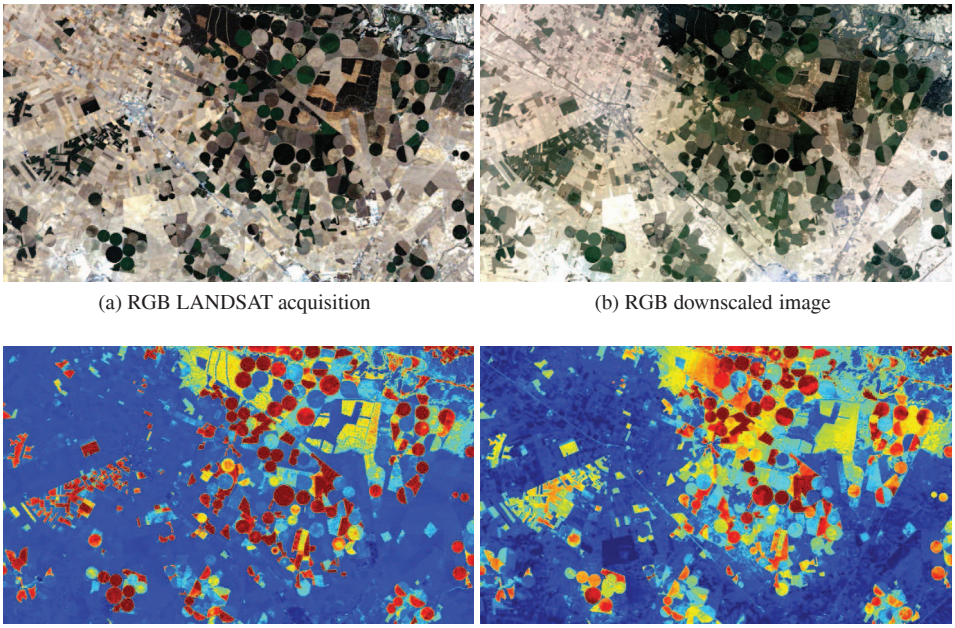
An additional problem appears from the band-by-band unmixing approach, which may give rise to unrealistic estimated spectra, as the spectral shape is not imposed in any way. Furthermore, the purest spectra (extreme values of the vertex) are rarely selected as they increase the unmixing error. These problems are alleviated here by including a regularization term in the cost function [1]:

$$\hat{\mathbf{E}}^{(M)} = \min_{\hat{\mathbf{E}}} \left\{ \sum_{i\in\mathcal{W}} \left[S_i^{(M)} - \sum_{k=1}^{K} A_{ik}^{(M)} \hat{E}_k^{(M)} \right]^2 + \lambda \frac{w^2}{K} \sum_{k=1}^{K} \left[\hat{E}_k^{(M)} - \bar{S}_k^{'(M)} \right]^2 \right\}, \tag{3}$$

where $\bar{S}_k^{'(M)}$ are a set of 'predefined' endmembers *per* class selected from the set of MERIS pixels giving higher abundance levels for each class, and λ is a regularization parameter that avoids that found endmembers in a window differ too much from the expected class spectra.

2.4. Spectral assignment

Once the class endmembers in each window $\mathcal{W}$ are estimated, one could simply assign the estimated endmember to each Landsat class pixels in the window. This approach is problematic because all pixels of the same class in the analyzed window would have



(a) RGB LANDSAT acquisition (b) RGB downscaled image

(c) NDVI LANDSAT acquisition (d) NDVI downscaled image

Fig. 3: RGB composition and NDVI for the original Landsat acquisition on 17/07/2004 (left) and the downscaled image with the proposed method (right).

the same signature, and thus no spectral variability would be obtained. An alternative approach consists on using the posteriors from the soft clustering, $\mathbf{A}^{(L)}$, to generate the high resolution fused pixels in the evaluated MERIS pixel, $\hat{\mathbf{S}}^{(L)}_{fused} = \mathbf{A}^{(L)}\hat{\mathbf{E}}^{(M)}$, where $\mathbf{A}^{(L)}$ contains class proportions of all the Landsat pixels in the MERIS pixel footprint $\mathcal{P}$.

3. EXPERIMENTAL RESULTS

This section presents the obtained experimental results and analyzes the impact of the free parameters on the solution.

3.1. Data collection and experimental setup

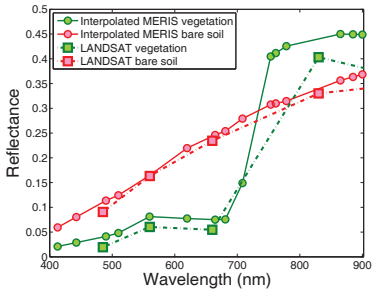


Fig. 2: Vegetation and soil spectra of original and downscaled images.

A temporal series of Landsat TM and MERIS acquired over Barrax (Spain) site in the frame of the ESA SPARC 2004 campaign is used to illustrate the capabilities of the proposed method. A 32×20km area of the Landsat images acquired on July 10th, 17th, and 26th are used. The intermediate image is not used during the downscaling process to allow us validating the results. The free parameters of the algorithm were tuned as follows: number of clusters $K = \{10, 16, 21, 24, 30\}$, the window size $w \in [5, 15]$, and the regularization parameter $\lambda \in [0, 0.6]$.

3.2. Evaluation of the obtained results

To assess the obtained fusion quality we used both qualitative and quantitative distortion measurements between the fused and the actual Landsat images. The qualitative assessment consists of the inspection of typical products delivered to the final user: the RGB composition and the Normalized Difference Vegetation Index (NDVI). Figure 3 compares the obtained results for the target date 2004/07/17. We can observe that the obtained and expected products are visually very similar.An additional assessment of the obtained products consists in analyzing the obtained spectra for different cover classes. Figure 2 shows the interpolated and actual spectra for the labeled classes 'bare soil' (red lines) and 'vegetation' (green lines). Note that Band 4 of LANDSAT is strongly affected by water absorption band so differences with downscaled spectra (MERIS resolution) are expected in this range.

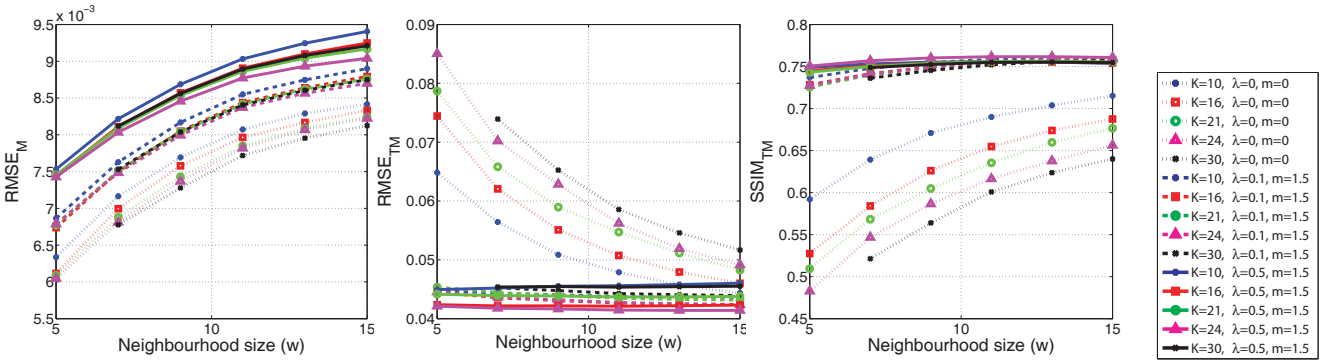


Fig. 4: RMSE and SSIM results as a function of the parameters w, λ, m and K.

3.3. Analysis of free parameters

The quantitative assessment was conducted using standard distortion metrics, such as the root-mean-square-error (RMSE) and the structural similarity index (SSIM) [5]. The RMSE criterion gives an estimation of the Euclidean distortion and treats all errors equally. We compute two version of RMSE: the RMSE_M is computed at MERIS resolution, thus accounting for the spectral distortion, while RMSE_{TM} is computed at Landsat resolution and better reflects the spatial distortion. SSIM is a perceptual criterion for assessing image similarity, and is well-known in image coding and denoising communities as it reflects adequately the perceived distortion by a human observer.

Figure 4 shows the obtained results for different combination of the parameters and for the used distortion metrics. When looking at the RMSE_M, it is observed that the error increases with the window size for all parameter combination. The proposed method ($\lambda \neq 0$, $m \neq 0$) performs worse than when no regularization or abundance estimation is included. Note, however, that results are numerically very similar. The proposed method improves when regularization and abundance estimation is included in the case of RMSE_{TM}, thus suggesting that a much better spatial interpolation is obtained. Regularization makes the solution almost independent on both the window size and the number of classes. This may be due to the fact that with small windows the linear inversion is not well-posed. The same result is observed for the SSIM criterion.

4. CONCLUSIONS

Results of the proposed methodology are illustrated in ENVISAT MERIS and Landsat TM Images. These are excellent candidate sensors for the proposed synergistic approach since they provide medium and high spatial resolution images, respectively, with a global coverage and a spectral overlap in the visible and near-infrared region. It is worth noting that other current and future sensors could also be used for the same purpose. Future GMES and contributing satellite missions like EnMAP or SEOSAT are suitable to increase the temporal resolution by means of downscaled images. In particular, the imaging spectrometers on board the Sentinel-2 and Sentinel-3 systems would be a perfect example of a high and medium resolution instruments with complementary temporal resolution to be combined.

5. REFERENCES

[1] B. Zhukov, D. Oertel, F. Lanzl, and G. Reinhackel, "Unmixing-based multisensor multiresolution image fusion," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 37, no. 3, pp. 1212–1226, May 1999.

[2] R. Zurita-Milla, J. Clevers, and M.E. Schaepman, "Unmixing-based landsat TM and MERIS FR data fusion," *Geoscience and Remote Sensing Letters, IEEE*, vol. 5, no. 3, pp. 453–457, July 2008.

[3] T. Kohonen, *Self-Organizing Maps*, Springer Series in Information Sciences, Vol. 30, 3rd extended edition, 2001.

[4] Ethem Alpaydin, "Soft vector quantization and the em algorithm," *Neural Networks*, vol. 11, no. 3, pp. 467–477, 1998.

[5] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612, 2004.