

# Wavelet-based image fusion in multi-view three-dimensional microscopy

Jose L. Rubio-Guivernau<sup>1</sup>, Vasily Gurchenkov<sup>2</sup>, Miguel A. Luengo-Oroz<sup>3</sup>, Louise Duloquin<sup>4</sup>, Paul Bourguin<sup>5</sup>, Andres Santos<sup>6</sup>, Nadine Peyrieras<sup>7</sup> and Maria J. Ledesma-Carbayo<sup>8</sup>

## ABSTRACT

**Motivation:** Multi-view microscopy techniques such as Light-Sheet Fluorescence Microscopy (LSFM) are powerful tools for 3D + time studies of live embryos in developmental biology. The sample is imaged from several points of view, acquiring a set of 3D views that are then combined or fused in order to overcome their individual limitations. Views fusion is still an open problem despite recent contributions in the field.

**Results:** We developed a wavelet-based multi-view fusion method that, due to wavelet decomposition properties, is able to combine the complementary directional information from all available views into a single volume. Our method is demonstrated on LSFM acquisitions from live sea urchin and zebrafish embryos. The fusion results show improved overall contrast and details when compared with any of the acquired volumes. The proposed method does not need knowledge of the system's point spread function (PSF) and performs better than other existing PSF independent fusion methods.

**Availability and Implementation:** The described method was implemented in Matlab (The Mathworks, Inc., USA) and a graphic user interface was developed in Java.

The software, together with two sample datasets, is available at <http://www.die.upm.es/im/software/SPIMFusionGUI.zip>

temporal and spatial resolution, tracking of every single cell and reconstructing the cell lineage tree becomes possible (Olivier *et al.*, 2010; Swoger *et al.*, 2010). Typical optical microscopy techniques (like confocal or two-photon laser scanning microscopy) provide images with resolution along the optical axis considerably worse than lateral resolution. Moreover, image quality gets progressively worse as light travels deeper inside the specimen, meaning that for relatively large specimens (such as zebrafish embryos or larvae) it is impossible to get good images of the whole embryo.

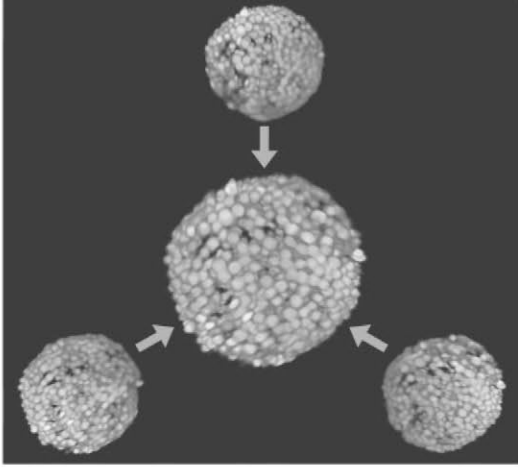
To overcome this kind of limitations other microscopy techniques have been developed. Spinning-disk microscopy (Graf *et al.*, 2005), for instance, provides much higher imaging speed compared to scanned techniques. In theta-microscopy (Stelzer, 1994), separate illumination and detection optical paths along orthogonal directions are used in order to improve the axial resolution. In a more recent approach called Light-Sheet-based Fluorescence Microscopy (LSFM) (Huisken *et al.*, 2004; Huisken and Stainier, 2009) wide-field detection is combined with a lateral light-sheet illumination along the focal plane of the detection objective. LSFM is claimed to offer several advantages over conventional confocal laser scanning microscopy (CLSM) technique:

- Unlike CLSM where a large part of the specimen is illuminated when recording a single point or plane, LSFM illuminates only the imaged plane, greatly reducing the overall photo-damage of the specimen.
- The system's point spread function (PSF) is the combination of the light-sheet shape and the detection objective's PSF. For this reason, when low numerical aperture detection lenses are used, LSFM achieves significantly better axial resolution than other techniques such as confocal or two-photon fluorescence microscopy.
- In LSFM wide-field detection is used, meaning that a full plane of the volume is acquired at once by an array of detectors (typically a CCD camera). In contrast, laser scanning microscopy (either confocal or two-photon) acquires one pixel at a time. This means that, for identical frame rates, LSFM has much longer per-pixel measurement time, which translates in a higher number of collected photons per-pixel, higher signal-to-noise ratio and dynamic range.

## 1 INTRODUCTION

### 1.1 Multi-view 3D microscopy on live embryos

3D+time images of fluorescently labeled cells in live model organisms are essential to developmental biology (Dzyubachyk *et al.*, 2010; Muzzey and van Oudenaarden 2009; Truong and Supatto 2011). When this kind of images is acquired with enough



**Fig. 1.** A rendering of the fusion of the three different views of a sea urchin embryo, acquired at  $\sim 20$ h post-fertilization (hpf) is shown at the center. Renders of the individual views are shown in a smaller scale around the fusion.

Another important feature of typical LSFM implementations is their ability to perform multi-view imaging, i.e. to obtain several volumes of the sample with different orientations. This is usually achieved by mounting the specimen under study on a rotating stage, and it takes advantage of the large working distance of water dipping lens objectives or of relatively low NA objectives, which makes the implementation of the rotation stage easier. Each view is a 3D volume formed as a stack of planes acquired sequentially. General schemes of LSFM can be found in the literature (Huisken and Stainier, 2009) and in Supplementary Figure S1. Such multi-view acquisition capability is desirable because, despite the previously mentioned advantages of LSFM over other fluorescence microscopy techniques, single views still show a series of problems (see Supplementary Fig. S2), like an axial resolution still worse than in plane resolution, slices increasingly dim and blurred with depth, etc., which will be further described below.

## 1.2 Multi-view fusion techniques

The multi-view imaging capability of LSFM provides extra information with respect to a single-view approach, as those regions of the specimen that are acquired with lower quality in one view will appear sharper and brighter in a different view.

To be fully useful, the information of the sample distributed among several volumes by multi-view imaging should be combined into a single volume (Fig. 1). Several techniques have been recently proposed for this task (Krzic, 2009; Preibisch *et al.*, 2008 ; Swoger *et al.*, 2007; Temerinac-Ott *et al.*, 2011).

One approach is to pose the problem as a multi-view deconvolution, which can be solved using extensions of classic iterative deconvolution algorithms like Richardson–Lucy (Krzic, 2009) or Maximum A Posteriori (Swoger *et al.*, 2007) to the multi-view situation, by updating the estimate using one view at a time.

This kind of algorithms performs fine when the actual PSF is well approximated and used in the deconvolution process. Its main drawback is precisely the need for a good estimate of the PSF, which

is particularly difficult to obtain when the PSF is spatially variant as observed in the LSFM imaged volume due to the shape of light sheet and thickness of specimens. In (Temerinac-Ott *et al.*, 2011) a spatially variant Richardson–Lucy is proposed, where the PSF is dynamically estimated based on fluorescent beads embedded in the sample. Finally, it is worth mentioning that PSF-based algorithms require an extremely precise registration of the views.

A different approach called content-based fusion (Preibisch *et al.*, 2008) was recently proposed in which the fused volume is the weighted average of all the available views, using the local entropy of each view as weights. This method is very fast and the results provided show clear improvement with respect to each of the acquired views.

We propose here an alternative method for multi-view fusion on the Discrete Wavelet Transform (DWT) (Mallat, 2008) space. An advantage of such approach when compared to multi-view deconvolution is the fact that it does not require a PSF estimate.

Moreover, due to the use of multi-band transforms that decompose each acquired view onto several scales and orientations, our method is able to select the best view independently along each band and can thus be considered as a multi-scale and multi-orientation content-based fusion.

This enhanced orientation discrimination feature is very valuable for LSFM data, as in a given region of the sample two or more views often provide useful information, each one along a different orientation (an example of such situation is provided in Supplementary Figure S3). In this kind of situation, the image-space weighted average performed by the content-based fusion method is unable to independently discriminate directional information components on each view, while the proposed method can do so by working in the wavelet-space.

## 2 METHODS

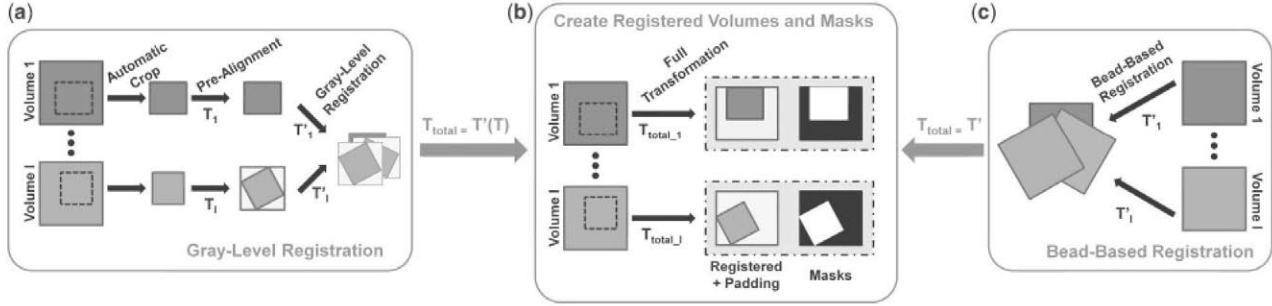
### 2.1 Problem description

The proposed method is designed to deal with the multi-view volume fusion problem in LSFM. There are mainly two ways in which different 3D fluorescence microscopy acquisitions can be complementary in a multi-view setting (Swoger *et al.*, 2007):

- Even though LSFM offers a relatively small axial PSF, due to the use of orthogonal illumination and detection axes, axial resolution remains usually worse than in-plane resolution. The in-plane and axial resolution throughout the specimen's volume change roles while it rotates around an axis orthogonal to the detection axis. Two orthogonal views offer the most complementary information in terms of PSF.

For highly diffusive samples (which applies to living tissues even when relatively transparent) excitation and emission light far from the excitation and detection objective respectively gets severely attenuated/absorbed as it travels through the sample, which results in increasingly degraded image quality deeper into the sample. For large specimens such as the zebrafish embryo or larva, each view will contain useful information for less than half of the sample. The complementarity of views taken from different angles to compensate for this effect is quite clear.

The goal of our method is to take a set of volumes, each one corresponding to a different view of the same sample (e.g. live embryo or larva of chosen animal models), and combine the complementary information available in all of them to create a new *fused* volume of a better overall quality than any original view alone.



**Fig. 2.** Overview of the pre-processing workflow. On the left (a) and on the right (b) we find the two possible registration methods, while the final step is depicted in the middle (c). For gray-level registration (a) two additional steps are performed: automatic cropping and pre-alignment, while bead-based registration (b) is directly performed on the acquired volumes. With either registration, the output of this step is a total transformation matrix for each of the input volumes, which is then used (c) to create the registered volumes and masks that will be passed to the fusion method. Both the registered volumes and the masks are padded as needed so that all of them have the same dimensions and all their information is preserved.

## 2.2 Pre-processing

Several pre-processing steps might be performed before applying the fusion method to the data, e.g. cropping the volumes and, for very large datasets, down-sampling the in-plane resolution to lower the computation time. However, the only necessary step that must be performed before the fusion is the alignment or registration of all the acquired views, because any miss-registration will lead to artifacts in the final fused volume.

The rotation angle of the sample mount is known from the acquisition setup. But it only provides a coarse estimation of the actual transformation between views. Several artifacts including the lack of precision and accuracy in the rotation calibration and operation (Krzic, 2009) make a fine registration necessary.

Fine registration of LSFM images is quite challenging, as the depth-dependent blurring means that the registration procedure is complicated by information differences between the views. Furthermore, for large specimens, some regions can be totally obscured because of light scattering as described above. To face these problems, more views might be required to ensure sufficient overlap of usable information and external cues such as fluorescent beads in the mounting medium might be required (Preibisch *et al.*, 2010). Two different approaches have been used during this work, depending on the characteristics of the dataset:

- A bead-based affine registration algorithm (Preibisch *et al.*, 2010) was used for datasets with fluorescent beads embedded in the mounting medium around the specimen.
- A gray-level affine registration algorithm (Thevenaz *et al.*, 1998) for datasets with no fluorescent beads, which uses just the information from the image intensity values.

For small embryos such as sea urchin, bead-less sample preparation and thus gray-level registration is preferred. However, 3D volumes acquired from large embryos like zebrafish show extensive blurring and eventually little overlap between views, and we confirm that the bead-based registration proposed previously (Preibisch *et al.*, 2010) is a useful strategy. The detailed pre-processing scheme used for each of the registration approaches is depicted in Figure 2.

For the gray-level registration, each of the acquired volumes is first automatically cropped in order to reduce as much as possible the data size and thus the time and memory consumption of the registration. Due to the low noise level of LSFM images, their background is quite uniform and such automatic cropping works well. Each cropped volume is then pre-aligned according to the rotation step configured during the acquisition, and the pre-aligned volumes are then passed to the gray-level affine registration algorithm. The transformation matrices corresponding to pre-alignment and

fine registration are then composed to produce the total transformation matrix of each input volume.

On the other hand, when bead-based registration is selected no initial cropping is used, mainly because it would not make sense to remove the beads needed for the registration algorithm to perform its task. No pre-alignment step is needed either for this kind of registration, as it performs well directly on the acquired volumes. Once the total transformation matrix has been computed for each of the views (by composing pre-alignment and fine registration matrices in the gray-level case, or directly obtained from the bead-based registration otherwise) they are applied to the original views in order to obtain the registered volumes, which are padded as needed in order to make all of them have the same dimensions. In parallel to this transformation, binary masks are also computed that indicate which voxels of each registered volume contain actual acquired data. These masks will be later used to prevent border artifacts, as it will be further explained later.

In both approaches, the first view is arbitrarily chosen as a reference, and either registration method is used to compute the affine transform matching it with the other views. When all the views have been brought to the reference frame coordinate system, the fusion process can start.

## 2.3 3D wavelet fusion

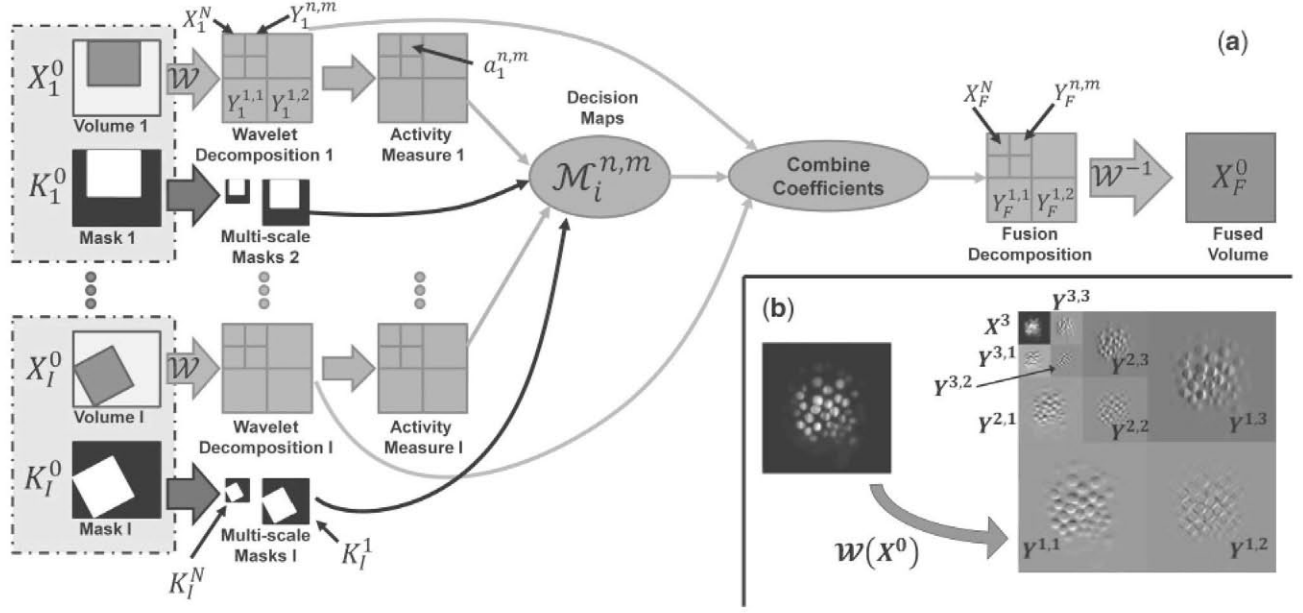
There is a family of techniques named image fusion (Piella, 2003), typically used on 2D images, which can effectively combine information from different sources into a single composite image.

Among these techniques, wavelet-based image fusion (Li, 1995) is one of the most widely used, and has already found some applications in biomedical imaging (Rajpoot *et al.*, 2009).

The main idea behind this approach is to take advantage of the properties of multi-band image decomposition schemes like the discrete wavelet transform. In this kind of decompositions, an input image (or volume)  $X^0$  is decomposed onto several bands (each one corresponding to a specific scale and orientation, for instance)  $\mathcal{W}(X^0) = \{Y^1, \dots, Y^K, X^K\}$ , where  $Y^k, k = 1 \dots K$  are the different bands in which the image is decomposed and  $X^K$  is the residual low-pass approximation of  $X^0$ .

Salient features in the input image  $X^0$  become high-energy coefficients in at least one of the decomposition bands  $Y^k$ , while smooth regions become low- or zero-energy coefficients. So, roughly speaking, by applying a common multi-band decomposition to all the input images available, and then choosing the highest energy coefficients, a fused image combining salient features from all input images is obtained.

The detailed process is outlined in Figure 3a and the different steps are further described below. Before explaining the different steps, we recall the



**Fig. 3.** (a) Overview of the proposed fusion method. The inputs are a set of volumes containing the registered views  $X_i^0 (i=1 \dots N)$  and a corresponding set of masks  $K_i^0 (i=1 \dots N)$ . Each of the registered views is then passed to the wavelet transform of choice, generating the wavelet decomposition bands  $Y_i^{n,m}$  (where  $n=1 \dots N$  is the scale index and  $m=1 \dots 7$  the orientation index) and the low-pass approximation residuals  $X_i^N$ . In parallel the masks are also processed to get the multi-level masks  $K_i^n (n=1 \dots N)$  needed later. An activity measure  $a_i^{n,m}$  is computed for each of the wavelet coefficients, and by looking at those activity measures together with the multi-level masks the decision maps  $\mathcal{M}_i^{n,m}$  are computed. Finally, the wavelet coefficients are combined according to the decision maps, giving place to the fusion decomposition  $\{Y_F^{1,1}, \dots, Y_F^{N,7}, X_F^N\}$ , and by applying the inverse transform we get the fused volume  $X_F^0$  as output. (b) Wavelet decomposition example of a 2D slice, using three scales and three orientations, resulting in nine bands ( $N=3$  and  $K=9$ ). Although in this work we use undecimated wavelet transform, for illustration purposes this figure represents the classical decimated wavelet transform.

inputs of the fusion method:

- A set of  $I$  volumes  $X_i^0 (i=1 \dots I)$ , each of them containing a different 3D view of the specimen, which have already been registered, using the first volume  $X_1^0$  as reference.
- The corresponding set of binary masks  $K_i^0$ , each of them indicating which voxels on each of the volumes  $X_i^0$  contain relevant information. These masks are used, for instance, to tell which voxels of a registered volume come from the actual transformed volume, and which ones are just padding, allowing us to avoid border artifacts due to incomplete specimen coverage in some of the views.

**Transform input volumes:** we first apply the wavelet transform of choice to each of the input volumes:  $\mathcal{W}(X_i^0) = \{Y_i^1, \dots, Y_i^K, X_i^K\}$ , where  $i=1 \dots I$  is the view index. Depending on the specific wavelet transform chosen, there will be a different number and arrangement of the bands. If 3D DWT is selected, we get 7 different orientations on each scale, so when  $N$  scales are used during the decomposition, we end up with  $K=7 \cdot N$  bands. In this case it is more practical to index the bands by scale ( $n=1 \dots N$ ) and orientation ( $m=1 \dots 7$ ):

$$\mathcal{W}(X_i^0) = \{Y_i^{1,1}, \dots, Y_i^{n,m}, Y_i^{N,7}, X_i^N\}$$

Figure 3b shows an example of the decomposition of a single slice of a LSFM Volume, using a 2D DWT with three scales ( $N=3$ ) and three orientations on each scale, resulting in  $K=3 \cdot N=9$  bands.

The size of the bands will depend on the precise wavelet transform used. For instance, when using decimated DWT, the bands' size is halved on each level, so that the total number of elements in the whole decomposition  $\mathcal{W}(X_i^0)$  is always equal to that of  $X_i^0$ . On the other hand, undecimated implementations of DWT, keep the size of each  $Y_i^{n,m}$  and  $X_i^N$  equal to  $X_i^0$ 's size, so each additional level increases the total number of elements

in the decomposition. Undecimated implementations are usually preferred for fusion applications due to its shift-invariance property (Amolins *et al.*, 2007; Gonzalez-Audicana *et al.*, 2004; Redondo-Tejedor, 2007).

**Generate multi-scale masks:** parallel to the transformation of the input volumes, the corresponding masks are processed in order to prepare adequate multi-scale masks:  $K_i^n (n=1 \dots N)$ . These multi-scale masks are designed to ensure that only valid voxels from the input volumes are taken into account when generating the fused volume.

For this purpose we define the *support* of any given wavelet coefficient ( $Y_i^{n,m}(\cdot)$  or  $X_i^N(\cdot)$ ), as the set of voxels in the corresponding input volume which get involved in the computation of said coefficient. The size of the *support* will depend on the scale  $n$  and on the wavelet family used. When using *haar* wavelets, for instance, the support will be a  $2 \times 2 \times 2$  sub-volume for coefficients in scale 1, a  $4 \times 4 \times 4$  sub-volume for scale 2, and so on.

If we denote by  $S(X_i^n(\cdot))$  the support of the coefficient  $X_i^n(\cdot)$ , the multi-scale masks are built so that  $K_i^n(\cdot)$  takes the value 1 if the original mask  $K_i^0$  is equal to 1 at all the voxels  $v$  belonging to  $S(X_i^n(\cdot))$ , and otherwise it takes the value 0, that is:

$$K_i^n(\cdot) = \begin{cases} 1 & \text{if } K_i^0(v) = 1, \forall v \in S(X_i^n(\cdot)) \\ 0 & \text{otherwise} \end{cases}$$

Supplementary Figure S4 shows an example of how the use of these multi-level masks prevents the appearance of artifacts when some of the acquired views do not fully cover the specimen.

**Measure activity:** the next step is to compute an activity or saliency measure at every location of each band from all the wavelet decompositions. This will later be used to decide which of the wavelet coefficients are related to salient features in the input volumes, so that they should be included in the fused volume. A common and simple choice, providing very good results,

is to use the energy or the module of wavelet coefficients. In this context, the activity measure for any given wavelet coefficient  $Y_i^{n,m}(\cdot)$  is:

$$a_i^{n,m}(\cdot) = |Y_i^{n,m}(\cdot)|$$

*Computation of decision maps:* this is the most important step in the process depicted in Figure 3a, and it might be considered the core of the fusion method. We mean by decision maps the weights that define how, for each location in every band of the wavelet decompositions (including the residual low-pass approximation), the coefficients from all the input volumes are combined to create the fused volume.

As already mentioned, these kind of methods aim at keeping the most salient features, at different scales and orientations, among the input volumes. Having the activity measure from the previous step, a straightforward way to implement such behavior is to create decision maps  $\mathcal{M}_i^{n,m}$  that, for each location, scale and orientation, select the maximum activity measure from the different input volumes. The multi-level masks are taken into account in order to avoid selecting coefficients whose support includes non-valid voxels from the input volumes. This can be achieved with the following decision map equation:

$$\mathcal{M}_i^{n,m}(\cdot) = \begin{cases} 1 & \text{if } a_i^{n,m}(\cdot) \cdot K_i^n(\cdot) = \max_{j=1 \dots I} [a_j^{n,m}(\cdot) \cdot K_j^n(\cdot)] \\ 0 & \text{otherwise} \end{cases}$$

However, this definition is not adequate for using on every decomposition band. For instance, depending on the size of the objects present in the input volumes, some scales of their wavelet decomposition might be dominated by noise, so selecting the maximum coefficients on those bands would lead to increased noise in the fused volume. As an alternative for bands featuring noisy components, selecting the minimum activity coefficients instead of the maximum leads to better results. Supplementary Figure S5 shows a comparison demonstrating how using minimum activity selection on some bands can improve the results.

In order to take full advantage of the orientation and scale discrimination capability of the wavelet transform, it is necessary that the decision maps are able to select, in a given region of the image space, information from different views for different decomposition bands. In this sense, the local entropy maps used in content-based fusion, as seen in figure 6 from (Preibisch *et al.*, 2008), are too smooth. For this reason, in this work more *local* decision maps like the ones we just defined are preferred. In order to support this statement, in Supplementary Figure S6 we show the result of combining wavelet decomposition with local entropy maps, which gives results almost identical to content-based fusion itself.

With respect to the low-pass approximations corresponding to each input volume, they should contain similar information after extracting the relevant features into the different bands. For this reason it is usually preferred to average all the low-pass approximations, taking into account only those coefficients whose associate multi-level mask equals 1, which can be implemented by defining a different decision map equation, which is simply:

$$\mathcal{M}_i^{n,m}(\cdot) = \frac{K_i^n(\cdot)}{\sum_{j=1}^I K_j^n(\cdot)}$$

Other possible strategies are easy to incorporate into the framework by just creating alternative definitions of the decision maps.

*Combination of wavelet coefficients:* once the decision maps are ready, the next step is to combine the coefficients from all the input volume decompositions, generating the wavelet decomposition of the fused volume:

$$Y_F^{n,m}(\cdot) = \frac{\sum_{j=1}^I (\mathcal{M}_j^{n,m}(\cdot) \cdot Y_j^{n,m}(\cdot))}{\sum_{j=1}^I \mathcal{M}_j^{n,m}(\cdot)}$$

As for the residual low-pass approximation of the fused volume  $X_F^N$ , it is computed in the same way by averaging the low-pass approximations of all the available volumes  $X_i^N, i=1 \dots I$ .

*Inverse transform:* finally, the wavelet decomposition generated in the previous step is inverted to get the final fused volume.

$$\mathcal{W}^{-1} \left( \left\{ Y_F^{1,1}, \dots, Y_F^{n,m}, Y_F^{N,7}, X_F^N \right\} \right) = X_F^0$$

**Table 1.** Performance evaluation of both registrations and both fusion methods

Step	Method	Time (min/view)	Fixed Time (min)
Registration	Gray-level	24.5	0
	Bead-based	1.8	0
Fusion	Wavelet	64.7	19.2
	Content-based	1.5	0

All tests were performed on a Intel® Xeon® E5506@2.13 GHz with 48 GB RAM.

**Table 2.** Acquisition parameters for the two datasets used for this work

Animal [Dataset ID]	Objective	Volume size (voxels)	Voxel size ( $\mu\text{m}$ )	No. of views [angle]	No. of time steps
Sea urchin [090916eS]	Zeiss 40 $\times$ /1.0NA W	972 $\times$ 972 $\times$ 82	0.185 $\times$ 0.185 $\times$ 1.85	3 [120°]	406
Zebrafish [100728aS]	Zeiss 10 $\times$ /0.3NA W	600 $\times$ 600 $\times$ 111	1.48 $\times$ 1.48 $\times$ 5.55	5 [72°]	281

## 2.4 Algorithm implementation and performance

The whole algorithm has been implemented in MATLAB®, using a custom 3D extension of its undecimated discrete wavelet transform implementation (swt and iswt functions).

Table 1 summarizes the time consumed by the two main parts of the overall process: registration and fusion. Times are provided for the two registrations used and for both the proposed fusion method and the content-based fusion included in the ‘SPIM Registration’ plug-in of Fiji (<http://www.fiji.sc>).

The results are averages based on several timesteps from each dataset, and to allow comparison despite different volume sizes, times have been normalized to a common volume size of 600 $\times$ 600 $\times$ 600. For wavelet-based fusion, the final inverse transform step is independent of the number of views, so the time has been split in a per-view time plus a fixed time.

The current implementation of the proposed method is quite more time-consuming than content-based fusion. We expect a 5- to 10-fold reduction in computation time when the method is recoded in C. Besides all the process is based on filters implemented as convolutions with rather small kernels, so the code should greatly benefit from parallelization schemes.

## 2.5 Dataset description

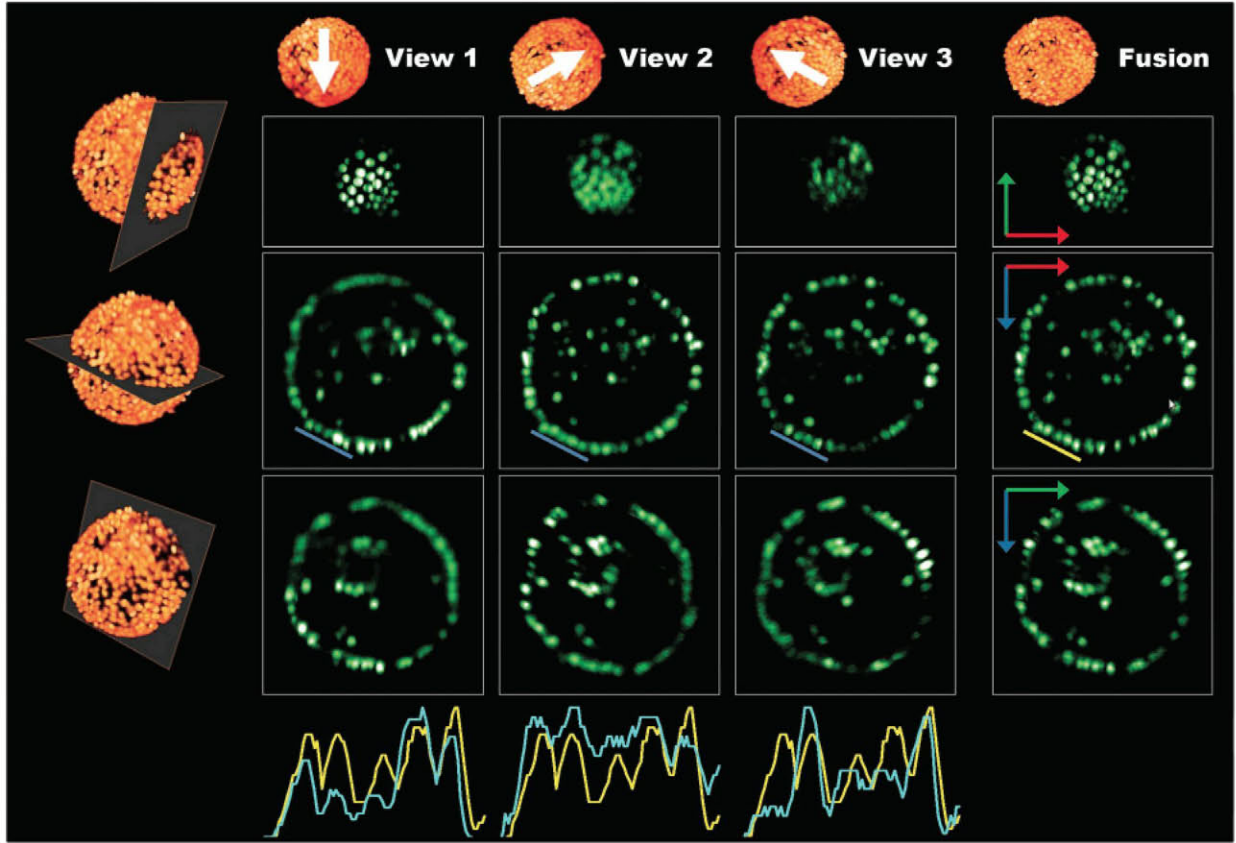
Datasets used in this work were acquired using a Digital Scanned Light-Sheet (DSL.M) version of the LSF.M microscope (Keller *et al.*, 2008) where the light sheet is generated by scanning a light beam instead of using a cylindrical lens as in the Selective Plane Illumination Microscope (SPIM) (Huisken *et al.*, 2004). Our method was applied to various kinds of datasets. The results presented here correspond to two datasets obtained from developing sea urchin (*Paracentrotus lividus*) and zebrafish (*Danio rerio*) embryos. Acquisition parameters are summarized in Table 2.

## 3 RESULTS

### 3.1 Live sea urchin embryo

A live sea urchin embryo was imaged for 20 h, from 5 h post fertilization (hpf) until 25 hpf. The whole acquisition time was divided in 180 s intervals, leading to over 400 time steps. At each time step 3 volumes (i.e. views) were acquired, using a 120° rotation of the sample between consecutive views. On all the acquired volumes, signal intensity relates to the local concentration of fluorescent protein accumulated in the cell nucleus.





**Fig. 4.** Results of the proposed method on a sea urchin embryo, imaged at 20 h post-fertilization (hpf). Each column shows three orthogonal slices from a single volume. Columns 1–3 contain slices from each of the three acquired views, while column 4 contains slices from the volume obtained by our fusion method. Rows 1–3 show slices along XY, XZ and YZ planes, respectively. At the bottom of the figure, three line profiles compare each of the individual views with the fusion. On each profile, the blue plot represents intensity values on the corresponding view, while the yellow line follows intensity on the fused volume. Line segments have been overlaid to the XZ slices (middle row), in parallel to the group of five cells represented by the profiles.

The sample did not contain fluorescent beads in this study, so the image registration method used before the fusion is the aforementioned gray-level affine registration (Thevenaz *et al.*, 1998).

After views registration, we applied the wavelet fusion method as described in Section 2.3. A 3D undecimated DWT was used as  $\mathcal{W}(X_t^0)$ , with *haar* wavelet filters and three scales ( $N=3$ ). Regarding the different possibilities for computation of decision maps (as discussed in Section 2.3), the following configuration parameters were used for this dataset:

- Minimum activity selection for bands in scale 1 (the finest) which is dominated by noise.
- Maximum activity selection for bands in scales 2 and 3, which contain the most relevant detail information.
- Averaging of the low-pass approximations.

Figure 4 shows an example of the fusion performance, by comparing slices and profiles of the fused volume and the acquired views corresponding to a time step around 20 hpf.

Each of the three views provides an incomplete picture of the embryo, with some regions showing sharp and bright nuclei, while

others exhibit heavy blurring and even some missing nucleus. Nevertheless, slices from the fused volume present overall sharp nuclei with good contrast. At the bottom of the figure line profiles show that the fusion provides the best representation of cell populations, because even though each view might show increased brightness and contrast for some cells, they might completely miss other ones that are indeed recovered in the fused volume.

The whole dataset (406 time steps) was processed similarly. A complete view of the results for this dataset is showed in the video displaying the evolution of the embryo during the 20 h of the imaging procedure (Supplementary Movie S11)<sup>1</sup>. Supplementary Movie S11 shows the improvement of individual cell identification in the fused volume that becomes even more obvious as the embryo grows and the individual views are unable to properly capture the details of the whole embryo.

### 3.2 Live zebrafish embryo

The second example is a live zebrafish embryo imaged for 12 h, starting at 5 hpf. The whole acquisition time was divided in

<sup>1</sup>Direct download available at [http://www.die.upm.es/im/videos/SPIM/Movie\\_S11.avi](http://www.die.upm.es/im/videos/SPIM/Movie_S11.avi) (last accessed date November 27, 2011).

155 s intervals, leading to over 280 time steps. At each time step five volumes (i.e. views) were acquired, using a  $72^\circ$  rotation of the sample between consecutive views. In all the acquired volumes, signal intensity correlates with the fluorescent protein concentration in the cell nucleus.

Fluorescent beads were added to the mounting medium, so that the bead-based registration method could be used. This means that the five acquired views show little overlapping information as shown in Supplementary Figure S7. For this reason, standard gray-based registration would be very challenging for this dataset, and bead-based registration is preferred.

The last column in Supplementary Figure S7 shows the result of fusing the five available views (columns 1–5). The configuration parameters used for this dataset, including wavelet family, number of scales and decision maps computation scheme, were identical to the ones described in Section 3.1 for the sea urchin dataset.

The fusion represents the full embryo better than any of the individual volumes, and, while some regions of the embryo are brighter in one specific view, the fusion captures all the relevant information, keeping good contrast whenever such contrast exists in any of the views (Supplementary Movie S12)<sup>2</sup> It is worth mentioning that, due to the size of this embryo, each view appears highly blurred in some regions, but the fusion method is able to deal with this situation and gives preference to non-blurred information from other views.

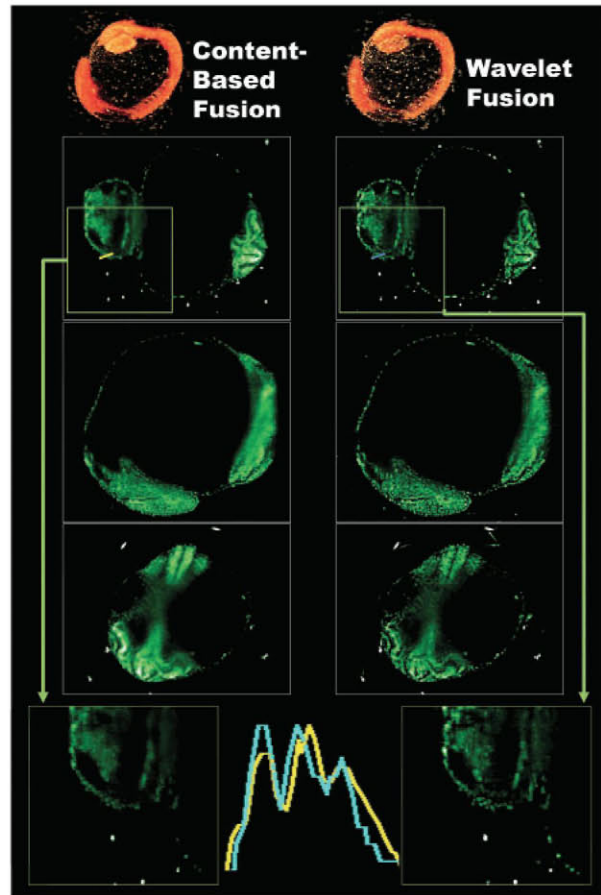
Supplementary Figure S8 shows snapshots of the renders available in the video, with the five original views around the resulting fused volume. For both Supplementary Figure S8 and Movie S12, a hard threshold was applied to the original and fused volumes in order to reduce the number of visible beads in the renders.

### 3.3 Comparison with existing methods

For both datasets, we have also compared our results with those of the previously mentioned content-based weighted averaging (Preibisch *et al.*, 2008), which is also PSF independent and freely available as part of Fiji open-source image processing package.

The comparison between our method and Preibisch *et al.*'s method is shown in Supplementary Figure S9 for the sea urchin embryo, and in Figure 5 for the zebrafish. The results of the wavelet-based fusion method look less blurred, and more individual nuclei can be distinguished. The latter is supported by the line profile comparison through a group of three nuclei in a blurred region (Fig. 5). For the sea urchin, the difference is less striking, but the line profiles in Supplementary Figure S9, reveals slightly better peak contrast with our wavelet fusion method (in blue) compared to the content-based fusion (in yellow).

Additionally, a quantitative comparison was carried out on two different timesteps for both datasets, by measuring the contrast of several line profiles, each one passing through two adjacent nuclei. An overall 31% improvement was measured for the proposed method in comparison to content-based fusion. The details are provided in Supplementary Figure S10.



**Fig. 5.** Comparison of content-based fusion (left column) and our fusion method (right column) for the zebrafish embryo, same slices as in Supplementary Figure S7. At the bottom, a line profile comparison of both (content-based fusion in yellow and our fusion in blue) across a group of three cells in a blurred region, surrounded by insets showing more detail of the areas where the line profile was obtained.

## 4 DISCUSSION

We propose a novel methodology to fuse multi-view images, such as those obtained from light-sheet-based fluorescence microscopy, but not limited to that particular technique. Almost any imaging technique in which several volumes of the same object are acquired, each of one providing incomplete and complementary information, could benefit from the proposed fusion scheme, including recent light-sheet techniques like DSLM using structured illumination (Keller *et al.*, 2010).

Our method is based on wavelet multi-scale and multi-orientation decomposition of the input volumes, and does not rely on a priori knowledge of the system PSF. While PSF-aware methods have more potential for reducing PSF-related artifacts, the proposed method can be used on those situations in which such knowledge is not available as is the case with LSFM acquisition where the PSF varies over the volume due to light sheet shape and to distortion caused by the specimen itself.

With respect to the other PSF-independent method mentioned in this work, content-based weighted averaging, the new method

<sup>2</sup>Direct download available at [http://www.die.upm.es/im/videos/SPIM/Movie\\_S12.avi](http://www.die.upm.es/im/videos/SPIM/Movie_S12.avi) (last accessed date November 27, 2011).



has improved orientation and scale discrimination as a direct consequence of performing the fusion in the wavelet domain instead of the intensity space. This improved discrimination is useful for instance in regions of the sample where two views contribute sharp details in different orientations.

This work focuses on the fusion method itself, and does not consider the effect of other complementary methods like denoising techniques which could be used in a pre-processing step. For instance wavelet denoising (Chang *et al.*, 2000; Donoho and Johnstone 1994) and bilateral filtering (Paris *et al.*, 2009; Yang *et al.*, 2009) seem good choices for future tests. Some variations (cross/joint and dual bilateral filtering) have been used for image fusion in certain situations (Bennett *et al.*, 2007), but they are not adequate in situations where different views do not cover the full specimen, as uniform regions in the image being filtered will stay uniform regardless of the information provided by the other views.

A possible way to further improve the usefulness of the proposed method would be to automatically adapt the wavelet fusion parameters, e.g. number of scales and fusion scheme, to the content of the input volumes.

Moreover, the general scheme can be accommodated to any kind of wavelet transform (different base wavelet basis, non-separable decompositions, etc.), making the proposed method highly flexible and thus adaptable to any kind of input volumes.

## ACKNOWLEDGEMENTS

ccIN2P3, Pascal Calvat and Jean-Yves Nief for data management, BioEmergences project and Uros Kzric for the SPIM/DSLIM setup installation in Gif sur Yvette, Miguel Carmona for the SPIM/DSLIM software development.

*Funding:* Ministry of Science and Innovation (TEC2008-06715-C02-02), Spain; Agence Nationale de la Recherche, France; Agence pour la Recherche sur le Cancer (to N.P.), France; Région Ile de France, France; Centre National de la Recherche Scientifique, France; FP6 New Emerging Science and Technology EC program (Embryomics and BioEmergences projects), European Union; European Fund for Regional Development (FEDER), European Union.

*Conflict of Interest:* none declared.

## REFERENCES

- Amolins, K. *et al.* (2007) Wavelet based image fusion techniques. *ISPRS J Photogramm.*, **62**, 249–263.
- Bennett, E.P. *et al.* (2007) Multispectral bilateral video fusion. *IEEE T. Image Process.*, **16**, 1185–1194.
- Chang, S.G. *et al.* (2000) Adaptive wavelet thresholding for image denoising and compression. *IEEE T. Image Process.*, **9**, 1532–1546.

- Donoho, D.L. and Johnstone, I.M. (1994) Threshold selection for wavelet shrinkage of noisy data. In *Proceedings of 16th Annual International Conference of the IEEE EMBS*. IEEE, Baltimore, MD, pp. A24–A25.
- Dzyubachyk, O. *et al.* (2010) Automated analysis of time-lapse fluorescence microscopy images: from live cell images to intracellular foci. *Bioinformatics*, **26**, 2424–2430.
- Gonzalez-Audicana, M. *et al.* (2004) Fusion of multispectral and panchromatic images using improved IHS and PCA mergers based on wavelet decomposition. *IEEE T. Geosci. Remote*, **42**, 1291–1299.
- Graf, R. *et al.* (2005) Live cell spinning disk microscopy. *Adv. Biochem. Eng. Biotechnol.*, **95**, 57–75.
- Huisken, J. and Stainier, D.Y.R. (2009) Selective plane illumination microscopy techniques in developmental biology. *Development*, **136**, 1963–1975.
- Huisken, J. *et al.* (2004) Optical sectioning deep inside live embryos by selective plane illumination microscopy. *Science*, **305**, 1007–1009.
- Keller, P.J. *et al.* (2008) Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy. *Science*, **322**, 1065–1069.
- Keller, P.J. *et al.* (2010) Fast, high-contrast imaging of animal development with scanned light sheet-based structured-illumination microscopy. *Nat. Methods*, **7**, 637–642.
- Krzic, U. (2009) Multiple-View Microscopy with Light-Sheet based Fluorescence Microscope. PhD Thesis, University of Heidelberg, Heidelberg.
- Li, H. (1995) Multisensor Image Fusion Using the Wavelet Transform. *Graph. Model Im. Proc.*, **57**, 235–245.
- Mallat, S. (2008) *A Wavelet Tour of Signal Processing*. 3rd edn. Academic Press, Burlington, MA.
- Muzzey, D. and Oudenaarden, A. van (2009) Quantitative time-lapse fluorescence microscopy in single cells. *Annu. Rev. Cell Dev. Biol.*, **25**, 301–327.
- Olivier, N. *et al.* (2010) Cell lineage reconstruction of early zebrafish embryos using label-free nonlinear microscopy. *Science*, **329**, 967–971.
- Paris, S. *et al.* (2009) Bilateral Filtering: Theory and Applications. *Foundations and Trends® in Computer Graphics and Vision* Vol. 4, No 1, pp. 1–73.
- Piella, G. (2003) A general framework for multiresolution image fusion: from pixels to regions. *Inform. Fusion*, **4**, 259–280.
- Preibisch, S. *et al.* (2008) Mosaicing of single plane illumination microscopy images using groupwise registration and fast content-based image fusion. *Proc. SPIE*, **6194**, 69140E 1–69140E 8.
- Preibisch, S. *et al.* (2010) Software for bead-based registration of selective plane illumination microscopy data. *Nat. Methods*, **7**, 418–419.
- Rajpoot, K. *et al.* (2009) *Multiview RT3D Echocardiography Image Fusion*. Springer, pp. 134–143.
- Redondo-Tejedor, R. (2007) New Contributions on Image Fusion and Compression Based on Space-Frequency Representations. PhD Thesis, Universidad Politécnica de Madrid, Madrid.
- Stelzer, E. (1994) Fundamental reduction of the observation volume in far-field light microscopy by detection orthogonal to the illumination axis: confocal theta microscopy. *Optics Commun.*, **111**, 536–547.
- Swoger, J. *et al.* (2007) Multi-view image fusion improves resolution in three-dimensional microscopy. *Optics Express*, **15**, 8029.
- Swoger, J. *et al.* (2010) 4D retrospective lineage tracing using SPIM for zebrafish organogenesis studies. *J. Biophotonics*, **4**, 122–134.
- Temeriac-Ott, M. *et al.* (2011) Spatially-variant Lucy-Richardson deconvolution for multiview fusion of microscopical 3D images. In *IEEE International Symposium on Biomedical Imaging 2011*, IEEE, Chicago, IL, pp. 899–904.
- Thevenaz, P. *et al.* (1998) A pyramid approach to subpixel registration based on intensity. *IEEE T. Image Process.*, **7**, 27–41.
- Truong, T.V. and Supatto, W. (2011) Toward high-content/high-throughput imaging and analysis of embryonic morphogenesis. *Genesis*, **49**, 555–569.
- Yang, Q. *et al.* (2009) Real-time O(1) bilateral filtering. In *IEEE Computer Vision and Pattern Recognition, 2009*, IEEE, Miami, FL, pp. 557–564.