# BIOLUMINESCENCE ENHANCEMENT THROUGH FUSION OF OPTICAL IMAGING AND CINEMATIC VIDEO FLOW

*Mickael Savinaud*[1,2,3] *Aristeidis Sotiras*[1,2] *Serge Maitrejean*[3] *Nikos Paragios*[1,2]

(1) Laboratoire MAS, Ecole Centrale Paris, Châtenay-Malabry, France
(2) Equipe GALEN, INRIA Saclay - Ile-de-France, Orsay, France; (3) Biospace Lab, Paris, France

## ABSTRACT

Optical imaging is an efficient mean to measure biological signal. However, it can suffer from low spatial and temporal resolution while animal deformable displacements could also degrade significantly the localization of the measurements. In this paper, we propose a novel approach to perform fusion of cinematic flow and optical imaging towards enhancement of the biological signal. To this end, fusion is reformulated as a population (all vs. all) registration problem where the two (being spatially aligned) signals are registered in time using the same deformation field. Implicit silhouette and landmark matching are considered for the cinematic images and are combined with global statistical congealing-type measurements of the optical one. The problem is reformulated using a discrete MRF, where optical imaging costs are expressed in singleton (global) potentials, while smoothness constraints as well as cinematic measurements through pair-wise potentials. Promising experimental results demonstrate the potentials of our approach.

***Index Terms*—** Biomedical optical imaging, bioluminescence, group-wise registration, MRF, discrete optimization

## 1. INTRODUCTION

Non-invasive visible light imaging is now a widely accepted technology allowing researchers to follow many biological processes in healthy and diseased animal models [1]. The detection of the light emitted by a probe provides functional information and localization of the processes to be studied. One limitation of this modality, is the impact of the anesthetic agents used for animal handling, specially in functionals experiments [2, 3]. Optical imaging devices are now able to image these processes in freely moving animals [4]. However, localization of the signal can be degraded during the monitoring period (due to motion). In some applications, the signal is too weak to be significant, so measurements are accumulated and being averaged on several frames. Therefore, substantial movements during a time window T blur the data and degrade the localization of the signal (Fig. 1). Motion compensation, throughout the acquisition, can address the above mentioned limitation.

Prior work trying to tackle this problem includes both group-wise and pair-wise registration techniques. In [5] a variational method recovering the deformations between consecutive frames based on silhouette and landmark information is proposed. A group-wise approach that minimizes an objective function based on the minimum description length principle is introduced in [6]. This method is based on points of interest and their associate texture descriptors, and is restricted to the body. Both techniques, despite their differences, consider only information extracted from the cinematic video.

This work was supported by ANRT (grant 987/2006), Biospace Lab and EMIL (European Molecular Imaging Laboratories) network. The authors thank O. Levrey (Biospace Lab) and E. Roncali for their assistance. Moreover, the authors would like to acknowledge the contribution of R. Boisgard (CEA-LIME and INSERM U803, Paris, France) for *in vivo* experiments.

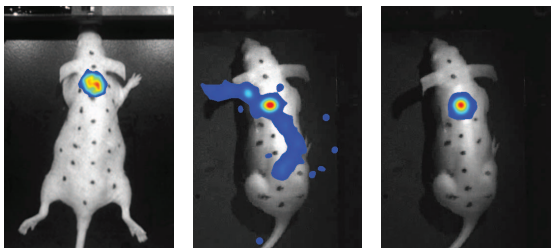


**Fig. 1**. Images of anesthetized and freely moving animal for a time window of 1 s. On the right motion compensation result.

The aim of this paper is to improve the localization of the bioluminescent (BL) signal by incorporating information not only from the cinematic flow but also from the optical imaging data. The approach followed is the one of the group-wise registration where a batch of multi-channel images is considered. In this context, standard pixel-based similarity measurements are considered between the frames of the video sequence, while a stack-wise statistical compactness criterion is considered for the case of BL signal. Free Form Deformations (FFD) are used, enabling for smooth and invertible deformation fields under specific conditions. The group-wise registration problem is formulated as a Markov Random Field (MRF) permitting the use of efficient discrete optimization techniques.

## 2. METHODS

The group-wise registration is based on an approach proposed in [7] and is now applied to a population of multi-channel images $\{I_1, \cdots, I_n\}$. The multi-channel image $I_i = \{V_i, O_i\}$, consists of the information obtained by the video acquisition of the moving object $V_i$ as well as the biological data $O_i$ that are simultaneously recorded. The goal of our approach is to bring the population of the images to the same pose, through mutual deformation, by taking advantage of the information provided by both modalities. In order to estimate the deformation field $\mathbf{T} = \{T_1, \cdots, T_n\}$ that would bring the population to a common domain, the problem will be casted as an energy minimization one.

### 2.1. Deformation Model

In order to decrease the dimensionality of the deformation model, cubic $B$-splines FFD are considered. The displacement field $D(\mathbf{x})$

is now given as a weighted average of the displacements applied to the nodes of the deformation grid. In a general way, it can be computed as $D(\mathbf{x}) = \sum_{\mathbf{p}\in G} \eta(|\mathbf{x}-\mathbf{p}|)\mathbf{d}_{\mathbf{p}}$, where $\eta(\cdot)$ is a weighting function that measures the contribution of each control point to the displacement field. In this particular case,

$$D(\mathbf{x}) = \sum_{r=0}^{3}\sum_{c=0}^{3} B_r(u)B_c(v)\mathbf{d}_{\mathbf{p}^{u_0+r,v_0+c}} \quad (1)$$

where $u_0 = \lfloor x/\delta_x \rfloor, v_0 = \lfloor y/\delta_y \rfloor, u = x/\delta_x - \lfloor x/\delta_x \rfloor$ and $v = y_i/\delta_y - \lfloor y_i/\delta_y \rfloor$. $B_r$ represents the $r$th basis function of the $B$-spline and $\delta_x = \frac{S_x}{K-1}, \delta_y = \frac{S_y}{L-1}$, ($S_x, S_y$ are the dimensions of the image along $x$- and $y$-axis respectively), denote the control point spacing. This deformation model allows for an explicit way to impose the diffeomorphism of the deformation. It suffices that the maximum displacement is constrained to be $0.4$ times the distance between the nodes of the grid [8].

As the goal is to deform the population of images towards the common domain at the same time, a deformation grid $\mathbf{G} = \{G_i, \cdots, G_n\}$, where each $G_i : [1, K] \times [1, L]$ is superimposed onto each image $I_i$ (the same grid is superimposed onto both components of the image), is considered. In other words, the goal is to deform the grids simultaneously (with a 2D displacement vector $\mathbf{d}_{\mathbf{p}_i^k}$ for each control point $k$ belonging to grid $G_i$) so that the images are coregistered. In this case, the transformation of a pixel $\mathbf{x}_i \in \Omega_i$, where $\Omega_i$ is the domain of image $I_i$ can be written as $T_i(\mathbf{x}_i) = \mathbf{x}_i + D_i(\mathbf{x}_i)$. The dimensionality of the parameter space and the non-convexity of the objective function makes challenging the estimation of the optimal deformation.

### 2.2. Discrete Population Fusion

Discrete optimization methods are quite popular in medical imaging and computer vision. This is due to their ability to provide guarantees on the quality of the obtained solution and their computational performance. Furthermore, they are modular with respect to the model (parameters to be estimated) and the model-to-data association (relation between a set of parameters and the observations). Therefore, such a framework is a natural approach to deformable population fusion as suggested in [7, 9]

Given the chosen deformation model, what should be done is to quantize the displacements the grid nodes can perform by sampling along the principal horizontal and vertical directions. The choice of labels entails a compromise between the quality of the solution and the computational effort. In order to keep the number of labels reasonable, the optimization will be performed in an iterative way introducing a notion of time $t$.

Let us consider a discrete set of labels $L = \{l_1, ..., l_q\}$ being associated with a quantized version of the deformation field $\{\mathbf{d}^{l_1}, \cdots, \mathbf{d}^{l_q}\}$. Then, by assigning a label $l_{\mathbf{p}}$ to a grid node $\mathbf{p}$ the node is displaced by the corresponding vector $\mathbf{d}^{l_{\mathbf{P}}}$. Thus, deformable population fusion consists of associating a label per node, such that by applying the resulting displacement field the images are aligned. The problem can be now formed as a multi-labeling one.

Mathematically, the problem can be formulated with the use of MRF. Typically, such a model can be represented as a graph $\mathbf{G}$ where the set of nodes represents the variables and the set of the edges represents the interaction between the variables. It can be written as:

$$E_{MRF}(\mathbf{l}) = \sum_{\mathbf{p}\in\mathbf{G}} V_{\mathbf{p}}(l_{\mathbf{p}}) + \sum_{\mathbf{p}\in\mathbf{G}}\sum_{\mathbf{q}\in\mathcal{N}(\mathbf{p})} V_{\mathbf{pq}}(l_{\mathbf{p}}, l_{\mathbf{q}}) \quad (2)$$

where $\mathcal{N}(\mathbf{p})$ represents the neighboring nodes to the node $\mathbf{p}$.

In our framework the previous energy can be rewritten as:

$$\begin{aligned} E_{MRF}(\mathbf{l}) =& \alpha \sum_{i}^{n} \sum_{\mathbf{p}_i^k \in G_i} V_{\mathbf{p}_i^k}(l_{\mathbf{p}_i^k}) \\ & \beta \sum_{i}^{n} \sum_{j=i+1}^{n-1} \sum_{\mathbf{p}_i^k \in G_i} \sum_{\mathbf{q}_j^k \in G_j} V_{\mathbf{p}_i^k\mathbf{q}_j^k}(l_{\mathbf{p}_i^k}, l_{\mathbf{q}_j^k}). \quad (3) \\ & \gamma \sum_{i}^{n} \sum_{\mathbf{p}_i^\lambda \in G_i} \sum_{\mathbf{q}_i^\mu \in \mathcal{N}(\mathbf{p}_i^\lambda)} V_{\mathbf{p}_i^\lambda\mathbf{q}_i^\mu}(l_{\mathbf{p}_i^\lambda}, l_{\mathbf{q}_i^\mu}) \end{aligned}$$

### 2.3. MRF For Optical And Cinematic Video Fusion

The MRF-model consists of singleton and pair-wise terms (Eq. 3). The first term (singleton) accumulates pixel-based measurements across the co-registered image stack. The second term (referred as inter pair-wise potentials) imposes the similarity between pairs of images while the third one (referred as intra pair-wise potentials) the smoothness of the deformation.

*Bioluminescence Statistical Compactness:* The idea is that as the images are getting aligned, the distribution of the BL signal values that correspond to respective places along the stack of images will become more and more compact. Here, standard deviation is used to measure the compactness of the distribution. Ideally, this measure should depend on deformations applied to all images, which would imply the use of high-order cliques. Though it may be possible, the use of more complex node interactions will decrease the efficiency of the model. Hence, an approximation will be made by considering that for a given node $\mathbf{p}_i \in G_i$, the rest of the images will not move during the current iteration.

$$\begin{aligned} V_{\mathbf{p}_i^k}(l_{\mathbf{p}_i^k}) = \int \cdots \int_{\Omega_1 \cup, \cdots, \cup \Omega_n} \eta_s^{-1}(\mathbf{x}_i, \mathbf{p}_i^k)\psi(T_1^{t-1}(\mathbf{x}_1), \cdots, \\ T_i^t(\mathbf{x}_i), \cdots, T_n^{t-1}(\mathbf{x}_n))\mathcal{STD}(O_1(T_1^{t-1}(\mathbf{x}_1), \cdots, O_i(T_i^t(\mathbf{x}_i), \cdots, \\ O_n(T_n^{t-1}(\mathbf{x}_n))d\mathbf{x}_1 \cdots d\mathbf{x}_n. \end{aligned} \quad (4)$$

$\psi$ is a Dirac-type function whose role is to identify which pixels have been mapped to corresponding positions in the common domain. It is defined as follows: $\prod_{j=1}^{n} \delta(|T_i^t(\mathbf{x}_i) - T_j^{t-1}(\mathbf{x}_j)|)$. The weighting function $\eta^{-1}$ computes the influence of the image point $\mathbf{x}_i$ to the control points $\mathbf{p}_i$. For example, if we consider the case of closest-neighbor interpolation, then a given pixel in the image will only contribute to the closest control point with a coefficient equal to one. The inverse function takes the following form: $\eta_s^{-1}(\mathbf{x}_i, \mathbf{p}_i^k) = \frac{\eta(|\mathbf{x}_i - \mathbf{p}_i^k|)}{\int_{\Omega_i} \eta(|\mathbf{y}_i - \mathbf{p}_i^k|)d\mathbf{y}_i}$. The previous relation gives the cost of deforming image $I_i$ through the application of a label $l_{\mathbf{p}_i^k}$ to the node $\mathbf{p}_i^k$, which belongs to $G_i$ and is situated at the position $k$, by taking into account only the optical imaging information.

*Similarity Between Pairs Of Video Frames:* The second term measures the similarity between all pairs of the video frames after having been deformed through the application of a pair of labels. It can be defined as:

$$\begin{aligned} V_{\mathbf{p}_i^k\mathbf{q}_j^k}(l_{\mathbf{p}_i^k}, l_{\mathbf{q}_j^k}) \approx \int_{\Omega_i \cup \Omega_j} \delta(T_i(\mathbf{x}_i), T(\mathbf{x}_j)) \\ \eta_p^{-1}(\mathbf{x}_i, \mathbf{p}_i, \mathbf{x}_j, \mathbf{q}_j)\rho(T_i(\mathbf{x}_i), T_j(\mathbf{x}_j))d\mathbf{x}_i d\mathbf{x}_j \end{aligned} \quad (5)$$

where $\rho$ is an appropriately defined function that provides for a dissimilarity measure and will be detailed in the continuation. The role of the $\delta(T_i(\mathbf{x}_i), T(\mathbf{x}_j)) = \delta(|T_i(\mathbf{x}_i) - T_j(\mathbf{x}_j)|)$ Dirac function is to determine which pixels $\mathbf{x}_i \in \Omega_i$ and $\mathbf{x}_j \in \Omega_j$ correspond to the same pixel in the domain where the comparison takes place. The inverse function takes the following form:

$$\eta_p^{-1}(\mathbf{x}_i, \mathbf{p}_i, \mathbf{x}_j, \mathbf{q}_j) = \eta_p^{-1}(|\mathbf{x}_i - \mathbf{p}_i|, |\mathbf{x}_j - \mathbf{q}_j|) = \frac{\eta(|\mathbf{x}_i - \mathbf{p}_i|)\eta(|\mathbf{x}_j - \mathbf{q}_j|)}{\int_{\Omega_i \cup \Omega_j} \delta(T_i(\mathbf{y}_i), T_j(\mathbf{y}_j))\eta(|\mathbf{y}_i - \mathbf{p}_i|)\eta(|\mathbf{y}_j - \mathbf{q}_j|)d\mathbf{y}_i d\mathbf{y}_j}. \quad (6)$$

Providing a similarity metric between the video frames is not evident due to the lack of texture and the important occlusions and dis-occlusions created by the fixed sensor view-point. To overcome these limitations, we defined $\rho$ from silhouette and landmarks representations in the domain $\Omega_i \cup \Omega_j$.

In order to design a robust criterion, we embed the 2D silhouettes, $\mathcal{S}_i$ and $\mathcal{S}_j$ in a higher dimensional space using the Euclidean distance transform [10]. The distance transform of the shape $\mathcal{S}$ is denoted $\phi_{\mathcal{S}}$.

$$\phi_S(\mathbf{x}) = \begin{cases} -\mathcal{D}(\mathbf{x}, \mathcal{S}), & \mathbf{x} \in \mathcal{S} \\ \mathcal{D}(\mathbf{x}, \Omega - \mathcal{S}), & \mathbf{x} \in \Omega - \mathcal{S} \end{cases} \quad (7)$$

where $\mathcal{D}(\mathbf{x}, \mathcal{S})$ refers to the minimum distance between the point $\mathbf{x}$ and the shape $\mathcal{S}$. This representation has the nice property to be invariant to translation.

To increase the ability to measure dissimilarity, we can extend the objective function within a band defined by isosurfaces of $\phi_S$. Therefore, $\rho_S$ is proposed as a robust quadratic similarity criterion:

$$\rho_S(T_i(\mathbf{x}_i), T_j(\mathbf{x}_j)) = w_\mu(\phi_S(T_i(\mathbf{x}_i)))\Big(\phi_S(T_i(\mathbf{x}_i)) - \phi_S(T_j(\mathbf{x}_j))\Big)^2, \quad (8)$$

where $w_\mu$ is a smooth Dirac function weighting positively the isosurfaces close to the zero level-set. The use of silhouettes to align the successive positions of the mouse would provide meaningful geometric correspondences between their shapes. However, their anatomical significance as well as the validity of the transformation within the silhouette are questionable. The use of landmarks can overcome this limitation and provide additionally constraints to align the entire structure.

These constraints are obtained by finding correspondences between points from the two shapes. Let us consider that $m$-landmarks $(\lambda_1^i, ..., \lambda_m^i)$ can be extracted from $V_i$ (drawn onto the mouse surface), and $n$-landmarks $(\lambda_1^j, ..., \lambda_n^j)$ from $V_j$. In our case, establishing correspondences is not straightforward given that the landmarks visibility depends heavily on the pose of the mouse and the patterns are very similar. Therefore to establish robust correspondences, we embedded the landmark information into an implicit criterion:

$$\phi_L(\mathbf{x}; i) = min_{k \in [1,m]} \mathcal{D}(\mathbf{x}, \lambda_k^i) \quad (9)$$

Then, alignment of shapes can be achieved through the alignment of their implicit representations within the mouse:

$$\rho_L(T_i(\mathbf{x}_i), T_j(\mathbf{x}_j)) = w_\nu(\phi_S(T_i(\mathbf{x}_i)))\Big(\phi_L(T_i(\mathbf{x}_i)) - \phi_L(T_j(\mathbf{x}_j))\Big)^2, \quad (10)$$

with $w_\nu$ is a smooth Heaviside function accounting for landmark correspondences only within the mouse. The use of landmarks will improve the performance of the method with respect to the alignment of the interior parts of the mouse. The above mentioned criteria can be integrated in $\rho$ which unifies silhouette and landmark constraints: $\rho = \rho_S + \kappa\rho_L$.

*Smoothness Deformation Prior:* The unary and the inter pair-wise potentials constrain the deformation to be consistent with both the visual support provided by the cinematic flow and the BL signal measurements. Thus, it is possible to localize the BL signal source with high accuracy. What is left to detail is the pair-wise potentials whose goal is to impose smoothness to the deformation fields applied to each image $I_i$. This can be achieved by penalizing the difference between the value of the labels that neighboring nodes can take. A simple distance function can be defined by computing the magnitude of vector differences:

$$V_{\mathbf{p}_i\mathbf{q}_i}(l^k_{\mathbf{p}_i}, l^\xi_{\mathbf{q}_i}) = |\mathbf{d}^{l^k_{\mathbf{p}_i}} - \mathbf{d}^{l^\xi_{\mathbf{q}_i}}| \quad (11)$$

where $d^{l^k_{\mathbf{p}_i}}$ is actually the displacement that corresponds to assigning a label $l^k$ to the node $\mathbf{p}_i \in G_i$.

## 3. EXPERIMENTAL VALIDATION

Our experiments were realized with an innovative device allowing to record simultaneously $O_i$ and $V_i$ at 45 $fps$. The scene video is acquired under near IR lighting and BL signal is recorded by an intensified CCD (Photon Imager, Biospace Lab) [4]. To validate our approach, we compare it with previous methods focusing on the motion correction of the BL signal. The evaluation is based on the BL signal because the aim is to provide an accurate and robust method for cinematic analysis in optical imaging. The method was tested on 3 acquisitions of 1300 frames. The acquisitions were subdivided in successive populations of 15 multi-channel images. Afterwards, the biological data were filtered with different time windows T. For all experiments, $\alpha, \beta$ and $\gamma$ (Eq. 3) were set equal to 2, 1 and 18 respectively. The weights were experimentally optimized for the sequence described in Sec. 3.2 and the case $\alpha = 0$ was also considered.

### 3.1. Group-wise Registration Evaluation

First, to visually assess the performance of our method, the result on a stack of images is presented in Fig. 2. In this figure, we present the mean image of the stack as well as the stack-wise sum of the biological data before and after registration. The fact that the mean image is far more sharp than the one before registration implies the success of the method on all the parts of the body contrary to [6]. Moreover the biological data seem well localized, without added noise.

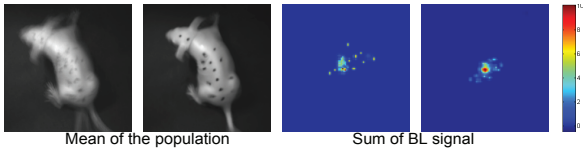
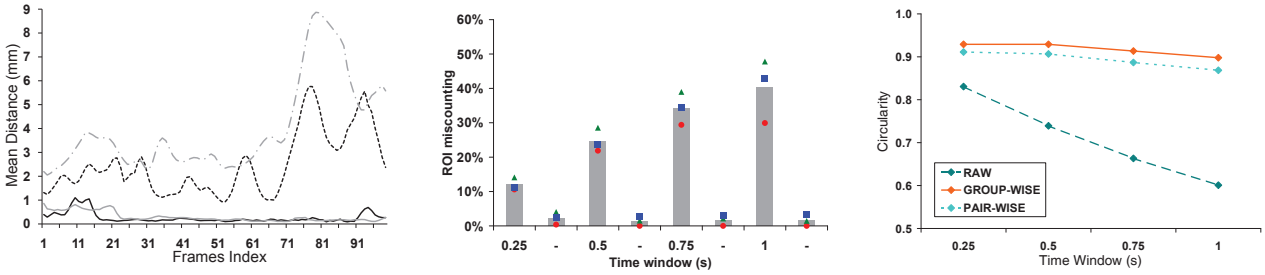


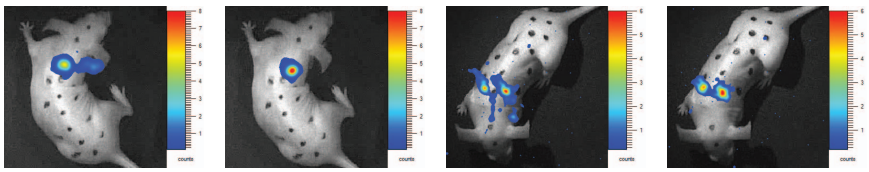

**Fig. 2**. Group-wise registration results: on the left the mean of the population, on the right sum of the BL signal (respectively before and after registration).

### 3.2. Impact On Signal Dispersion

We used a test sequence of 100 images in which we drew, on each frame a spot as a ground truth. We computed a statistical criteria




(a) Evolution of $D_T(t)$ (b) Evolution of ROI Miscounting (c) Circularity in function of $T$



(d) 1 source raw (e) 1 source corrected (f) 2 sources raw (g) 2 sources corrected

**Fig. 3**. (a) presents localization improvement for temporal fusions 0.5 s (black) and 1 s (gray). Dashed and bold lines indicate respectively raw and corrected data. (b) and (c) show the decrease of ROI miscounting and $\bar{c}$ for 3 sequences. In (b) bars indicate the mean of ROI miscounting criterion. (d) and (e) ; (f) and (g) show pair of images enhanced by our algorithm.

$D_T(t)$ to measure the signal dispersion for a time window T in frame $t$. Along this particular sequence, we have reduced the $D_T(t)$ by at least 85% which is better than the pair-wise method introduced in [5] (Table 1). Fig. 3(a) shows results throughout l00 frames. We used synthetic data to compare the intrinsic performances because biological signal changes strongly with respect to the emitting surface normal during acquisition.

| $D_T$ | 0.25 s | 0.35 s | 0.5 s | 0.75 s | 1 s |
|---|---|---|---|---|---|
| raw (mm) | 1.51 | 1.98 | 2.47 | 3.37 | 3.99 |
| corrected pw (mm) | 0.3 | 0.32 | 0.37 | 0.44 | 0.48 |
| corrected gw (mm) | 0.24 | 0.24 | 0.26 | 0.28 | 0.29 |

**Table 1**. Mean distance as function of T. Comparison between pair-wise and group-wise methods (pw and gw respectively).

### 3.3. Impact On Cinematic Analysis

We study now the biological signal throughout a sequence through visual assessment and cinematic analysis. Fig. 3 show two examples of our results with one and two small luminescent sources embedded on the mouse and temporal fusion of 1 s. We notice that the localization and the meaning of the biological data have been improved by our method as expected. To check this point throughout a sequence, we automatically compute for each frame regions of interest (ROI). We compare for each frame this number to the number of embedded sources to obtain the ROI miscounting. Results are summed up in the figure 3(b) for 3 sequences. As expected ROI miscounting decrease strongly with our method and ROI detection is better than with the pair-wise method. We also determine the circularity $c$ of the signal before and after registration as in [5]. In Fig. 3(c) the evolution of $\bar{c}$ between the group-wise and the pair-wise approach is compared.

## 4. CONCLUSIONS

In this paper, we have proposed a novel approach to enhance the BL signal by formulating the problem with the use of MRF theory. To the best of our knowledge, we are the first to consider together multi-channel images provided by different modalities. The fusion of different type of information enables us to further refine signal localization. We studied the impact of our method on cinematic acquisitions in optical imaging with temporal resolution of 1 s. This novel modality enables a quantification of a biological process even in awake animal. Our method compares favorably to the state of art.

## 5. REFERENCES


[1] R. Weissleder, "Molecular imaging in cancer," *Science*, 2006.

[2] I. Hildebrandt, H. Su, and W. Weber, "Anesthesia and other considerations for in vivo imaging of small animals," *ILAR Journal*, 2008.

[3] M. Keyaerts and al., "Isoflurane anesthesia inhibits firefly luciferase in a dose-dependent way: a pitfall for in vivo bioluminescence imaging," in *Proceedings of WMIC*, 2009.

[4] E. Roncali, M. Savinaud, and al., "A new device for real time bioluminescence imaging in moving rodents," *JBO*, 2008.

[5] M. Savinaud, N. Paragios, and S. Maitrejean, "Motion-based enhancement of optical imaging," in *IEEE ISBI*, 2009.

[6] Y. Han, G. Langs, and N. Paragios, "Group-wise mdl based registration of small animals in video sequences," in *IEEE ISBI.*, 2008.

[7] A. Sotiras, N. Komodakis, B. Glocker, J.-F. Deux, and N. Paragios, "Graphical models and deformable diffeomorphic population registration using global and local metrics," in *MICCAI*, 2009.

[8] D. Rueckert, L.I. Sonoda, C. Hayes, D.L.G. Hill, M.O. Leach, and D.J. Hawkes, "Diffeomorphic registration using B-splines," in *MICCAI*, 2006.

[9] Shu Liao and Albert C.S. Chung, "Non-rigid image registration with uniform spherical structure patterns," in *IPMI*, 2009.

[10] D. Cremers, "Dynamical statistical shape priors for level set-based tracking," *Transactions on PAMI*, 2006.