Model-Based Multi-view Fusion of Cinematic Flow and Optical Imaging*

Mickael Savinaud[1,2,3], Martin de La Gorce[1],
Serge Maitrejean[3], and Nikos Paragios[1,2]

[1] Laboratoire MAS, École Centrale Paris, France
[2] Equipe GALEN, INRIA Saclay - Île de France, Orsay, France
[3] Biospace Lab, Paris, France

Abstract. Bioluminescence imaging (BLI) offers the possibility to study and image biology at molecular scale in small animals with applications in oncology or gene expression studies. Here we present a novel model-based approach to 3D animal tracking from monocular video which allows the quantification of bioluminescence signal on freely moving animals. The 3D animal pose and the illumination are dynamically estimated through minimization of an objective function with constraints on the bioluminescence signal position. Derived from an inverse problem formulation, the objective function enables explicit use of temporal continuity and shading information, while handling important self-occlusions and time-varying illumination. In this model-based framework, we include a constraint on the 3D position of bioluminescence signal to enforce tracking of the biologically produced signal. The minimization is done efficiently using a quasi-Newton method, with a rigorous derivation of the objective function gradient. Promising experimental results demonstrate the potentials of our approach for 3D accurate measurement with freely moving animal.

1 Introduction

Non-invasive visible light imaging is now a widely accepted technology allowing researchers to follow many biological processes in animals [1]. The detection of the light emitted by a probe provides functional information and localization of the processes to be studied. The main limitation of such a modality is the difficulty to localize the signal in 3D especially in bioluminescence imaging techniques (BLI). Indeed photons emitted by bioluminescent cells are strongly scattered in the tissue of the subject and light propagation is diffusive by nature. Therefore different devices and reconstruction methods have been considered to solve this problem [2] but they all require surface acquisition. Furthermore, most of the existing techniques assume that animals have been anesthetized or are immobile, limiting the interest of this modality in functional experiments [3]. However new

* The authors thank Dr. S. Bonzom (Biospace Lab) for his technical assistance and would like to acknowledge the contribution of Dr. R. Boisgard (CEA-LIME, Orsay, France) for *in vivo* experiments.

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 668–675, 2010.

optical imaging devices are now able to image and quantify these processes in freely moving animals in 2D case [4,5].

Prior work trying to tackle these problems includes different techniques of computer vision and various hardware configurations. In Kuo et al. [6], mouse surface topography is obtained using a structured light combined with a single view detector. However this technique does not support freely moving animals because the hardware configuration does not enable cinematic acquisition. The use of temporal information involves either animal tracking or registration of the surface for different poses. In Papademetris et al. [7] a framework that capture articulated movement of the subparts in serial x-ray CT mouse images is proposed. This work has been enhanced to the whole body of the mouse with the use of a skeleton atlas [8] but is restricted to x-ray modality which provides intrinsically 3D information.

Pose estimation and tracking are well known problems in the computer vision community. Discriminative methods aim to recover pose from a single frame through classification or regression techniques [9]. However the high dimensionality of the space spanned by all possible pose restricted these methods to recognition of a limited set of predefined poses. Model-based methods are good candidates for continuous tracking over consecutive frames with small or predictable inter-frame displacements [10,11]. Another interesting aspect of model-based methods is that multi-view data can be handled without solving any correspondence problem between the images. Moreover the matching errors with 2D features on all the cameras can simply be summed to define a single error that should to be minimized.

The aim of this paper is to estimate the animal pose during a cinematic acquisition with a freely moving animal while providing accurate bioluminescence measurement. Our approach is a model-based one where the multi-channel flows are considered as an observation of the scene. In this context an articulated skeleton has been designed to deform the surface mesh towards producing different poses. The estimation of the 3D pose is solved through the optimization of on objective function that aims to generate the observed views from the model while detecting a consistent optical imaging signal across time. We propose a robust derivation of the criteria with respect to scene parameters using a classical gradient optimization.

2 Model Based Articulated Tracking

The proposed approach is inspired from [11] and is extended to the multi-view and multi-channel context. The multi-channel data $I_i = \{V_{i,j}, O_{i,j}\}$, consists of the information obtained by the video acquisition of the moving object $V_{i,j}$ in the different views j as well as the biological data $O_{i,j}$ that are simultaneously recorded on the same views. The goal of our approach is to evaluate the 3D pose with the population of the images by taking advantage of both channels and multi-views. In order to estimate the 3D pose that would correspond to the different observations, the problem will be cast as an energy minimization one.

2.1 Multi-views Pose Estimation

The mouse surface is deformed according to pose changes of an underlying articulated skeleton using Skeleton Subspace Deformation (SSD) [12,13]. The skeleton comprises 20 bones with 64 degrees of freedom (DOF). Each DOF corresponds to an articulation angle whose range is bounded to avoid unrealistic poses of the mouse. The mouse pose is fully determined by a vector $\mathbf{\Theta} = [\mathbf{w}, \mathbf{t}, \mathbf{q}]$ that comprises 57 articulation parameters vector $\mathbf{w}$, the 3D translation vector $\mathbf{t}$ and a quaternion $\mathbf{q}$ that specifies the global position and orientation of the mouse body with respect to the world's coordinate frame. In order to adapt the size of the mouse, three additional morphological parameters are introduced for each bone. These scale factors are added to the $\mathbf{\Theta}$ parameters that are optimized while fitting the model to the observations in the first frame and are kept constant for the subsequent frames.

The lighting is modeled as four point sources placed at an infinite distance and an ambient light. It is parameterized using three directional components for each light and with an additional ambient component, which produces a vector $\mathbf{L}$ of 13 parameters. The complexity of the lighting conditions is enforced by the fact that in our experiments light sources produce localized light spots due to high directivity of the light at output of the optical fibers. The mouse skin surface is assumed to be Lambertian. The mouse is white and we can assume the albedo to be constant over its entire surface. Thus we do not require the use of a texture mapped onto the surface due to the small variations of the albedo.

For a given pose $\mathbf{\Theta}$ and an illuminant $\mathbf{L}$, we define $V_{\mathrm{syn},j}(\mathbf{x};\mathbf{\Theta},\mathbf{L})$ to be the RGB intensities of the corresponding synthetic image comprising the mouse and the background evaluated at the point location $\mathbf{x}$ from the j^{th} camera. This is formulated using a classical perspective projection, the hidden surface removal and the Gouraud shading model. The tracking process attempts to recover for each successive frame the pose parameters $\mathbf{\Theta}$ and the illuminant $\mathbf{L}$ that produce the three synthesized images that best match the three observed ones, denoted by $V_{\mathrm{obs},j}$, with $j = 1, \ldots, 3$ the index of the camera. In the following objective function:

$$E_V(\mathbf{\Theta},\mathbf{L}) = \sum_{j=1}^{3} \int_{\Omega} \underbrace{\rho\big(V_{\mathrm{syn},j}(\mathbf{x};\mathbf{\Theta},\mathbf{L}) - V_{\mathrm{obs},j}(\mathbf{x})\big)}_{R_j(\mathbf{x};\mathbf{\Theta},\mathbf{L})} d\mathbf{x}, \tag{1}$$

the main term is defined by summing the residual errors $R_j(\mathbf{x};\mathbf{\Theta},\mathbf{L})$ between the synthetic images and the observed images V for each of the three cameras.

2.2 Bioluminescence Position Constraints

In order to take advantage of the information provided by the BL images $O_{\mathrm{obs},i,j}$, we compute in the first image the 3D position of the bioluminescence by automatic detection of the BL spot in each view. The 3D position X_{obs}^{O} of the light source can be estimated using a standard triangulation method. We do not adopt complex bioluminescence tomography methods because the tumors position is not expected to be far from the mouse surface. In case of tumors, we assume that this point is rigidly fixed to its nearest bone in the model. For each frame i, we detect automatically the position of the bioluminescence spot $P^O_{obs,i,j}$ in each view if possible. We aim at minimizing the sum of retroprojection error between these points and X^O_{obs}. We are now able to compute the expected position of X^O_{obs} given any new candidate mouse pose parameter vector $\mathbf{\Theta}$.

$$E_O(\mathbf{\Theta}) = \sum_{j=1}^{3} \|\Pi_j(X^O_{\text{obs}}(\mathbf{\Theta})) - P^O_{\text{obs},i,j}\|^2 \tag{2}$$

where Π_j corresponds to the operation of 3D to 2D projection using the j^{th} BL detector. E_O sums over the three views the 2D distances between the projection of the predicted bioluminescence source position X^O_{obs} and the actual observation extracted in the new $O_{i,j}$ image. This new term enforces the pose estimation of the mouse with respect to the BL signal during the tracking and enables to exploit in minimization process the biological information provided by secondary camera.

2.3 Tracking with Energy Minimization

During the tracking we determine, for each frame, the pose and the illumination parameters by minimizing an objective function which combines the two previous formulas. A factor β weights the two energies and is chosen empirically to be the squared inverse of the maximum expected deviation between the observed signal and the one fixed to the model. The minimization is done efficiently using a quasi-Newton method that requires the gradient of the objective function E_V. The gradient with respect to the lighting parameters is obtained by using the differentiation chain rule on the residual intensities. The gradient with respect to the pose $\mathbf{\Theta}$ is not straightforward to derive due to discontinuities in the residual image along the occlusion boundaries when $\mathbf{\Theta}$ varies. The adequate treatment of these discontinuities when computing the gradient is done using the proposed occlusion forces in [11].

3 Experimental Validation

Experiments were conducted using an innovative device capable of recording simultaneously scene video and optical data at 43 *fps*. The scene video V is acquired under near IR lighting and the BL signal O is recorded by an intensified CCD (Photon Imager, Biospace Lab). The two signals are simultaneously recorded and spatially registered [4]. Towards acquiring simultaneously different views of the animal and the BL signal emitted without large hardware modifications, we have considered two mirrors. Mirrors are defined by planes which are placed on the device stage with a angle of 90 degrees somewhere in the V camera field of view. The image of the mouse seen in each mirror can be interpreted as the image seen from a virtual camera, whose position and orientation are obtained by reflection with respect to the corresponding mirror plan (Fig. 1-C).

The parameters of the cameras are determined using the calibration toolbox. Mirror parameters are manually optimized with a known object to provide virtual camera positions and orientations. Illumination is provided by four optical fibers placed at the top of the the scene and at each extremity of the scene. The mouse can move in an area of 5 cm by 18 cm.

The mouse model used for the pose estimation is composed of a skeleton of 20 bones manually segmented from static micro-CT acquisitions (Skyscan 1178, Skyscan) and guided by a anatomical book [14](Fig. 1-A). The mouse surface is modeled as a three dimensional, closed and orientable triangulated surface of 1252 facets (Fig. 1-B). The mesh of the mouse was created with the micro-CT surface and elements of computer graphic project on mouse animation. The extremities of legs have not been modeled because it appeared through experiment that tracking these parts of the mouse is difficult given the quality of our observations while not being useful for our application (tumor cells embedded on the top of the mouse).

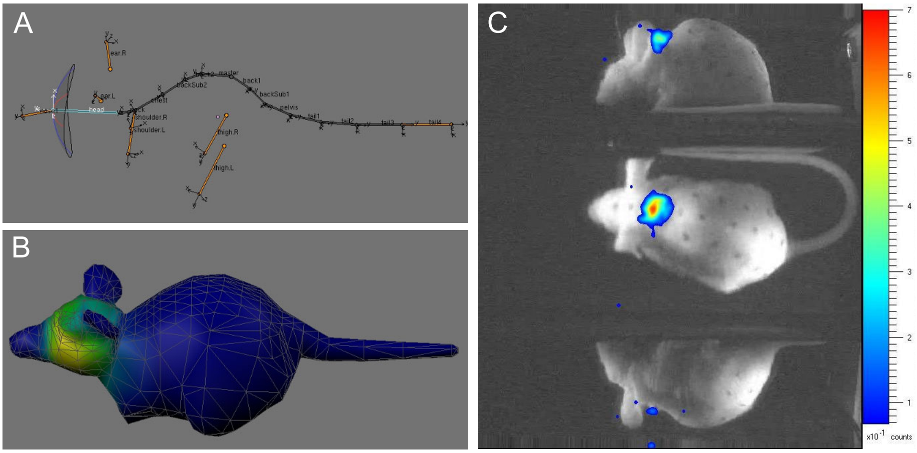

Fig. 1. Model and observations. On the left the skinned mouse model. On the right, fusion of the observed video and bioluminescence signal in the multi-view device.

This framework was applied to image a freely moving mouse (NMRI) bearing a PC12 tumor injected ten days before experiments in the dorsal part of neck (10000 cells in 0.5μL). In addition, we have drawn onto the surface of the mouse landmarks to measure locally the 3D position of the mouse surface. To validate our approach, we tested our method on 4 acquisitions which represent a total of 580 frames. Visual assessment and 3D cinematic analysis of the bioluminescence signal are used to demonstrate the interest of the method for measurement on freely moving animal.

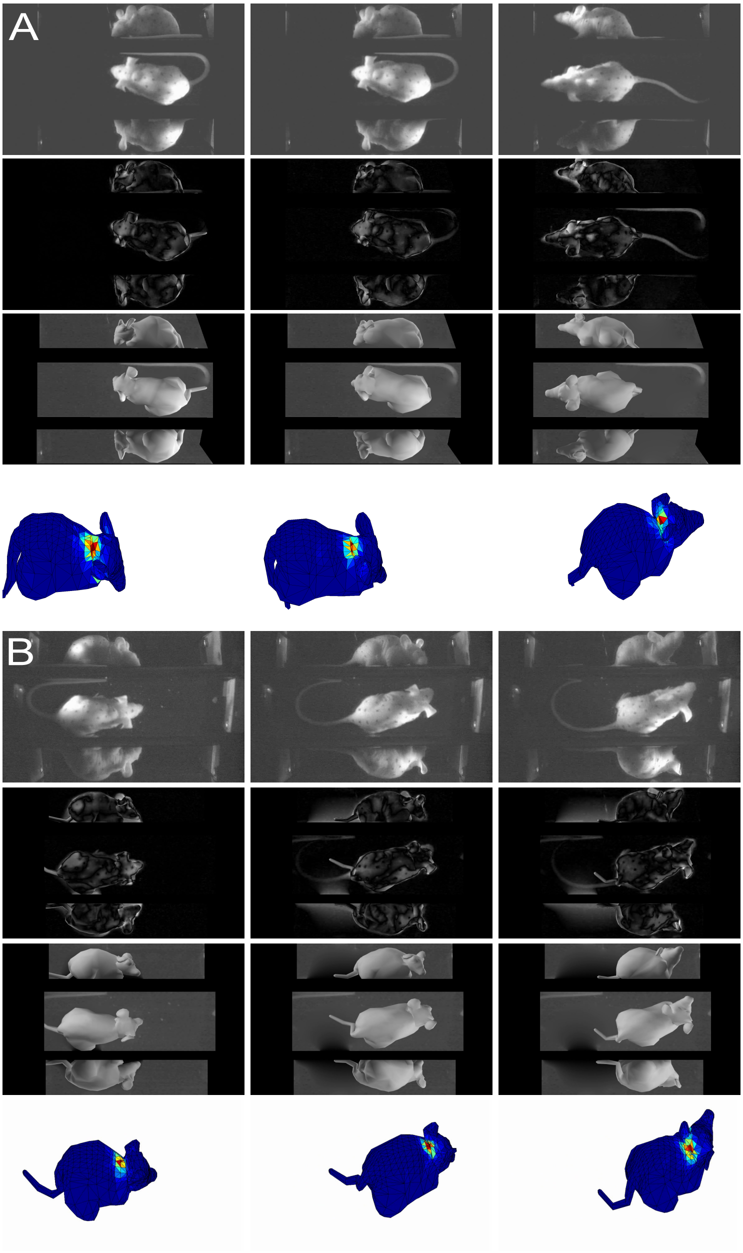

Fig. 2. Two sequences (A and B) processed with our tracking method. Each row corresponds to: the observed image, the final residual image, the synthetic final image and the model with the final pose and the bioluminescence backprojected to the surface.

3.1 Comparison with Triangulated Data

In order to evaluate our estimation of the pose, we manually annotated throughout the first sequence 800 couples of points in two different views. For each of these couples, we assigned a 3D position on the mesh with its first annotation. In the other frames, the new position of this point on the mesh was estimated with a triangulation method. Computing a 3D distance between the position of the reference with the corresponding pose provides a way to estimate the error produced by our method. In the first sequence, this error was about 5 mm with our 800 visual correspondences.

3.2 Visual Assessment

In order to check the usability of our pose estimation, we studied the biological signal throughout a sequence using visual assessment. Fig. 2 shows 6 frames extracted from the videos provided as additional material (respectively sequence 1 and 3). The low residual implies the correspondence between synthetic data and observations. Residual artifact are due to difficulty to render light spots generated by optical fibers with our illumination model. The backprojection of the BL signal on the surface is computed with all the views recorded by the camera O with a temporal smoothing of 5 frames and a spatial smoothing of 3 mm. In the great majority of frames the signal of interest is registered to the corresponding place of the emission surface. To our knowledge, this type of measurement is compatible with optical imaging experiments.

3.3 3D Cinematic Analysis

To evaluate the possibility to perform studies on freely moving animals with this new tool we computed for the first frame a region of interest (ROI) based on the faces which corresponds to the tumors position. In each following frame we compared the signal measured on these faces with the reference one to evaluate the stability and robustness of the pose estimation regarding to the biological data. Along our 4 sequences more than 75% of the signal was kept on the right faces (Table 1).

Table 1. ROI tracking: the two first lines indicate the characteristics of the first ROI while the last evaluates the quantity of the signal following the ROI throughout the sequence

	SEQ 1	SEQ 2	SEQ 3	SEQ 4
Number of faces:	61	41	42	49
Size of ROI (cm^2):	1.83	1.27	1.60	1.65
Mean of ROI intensity similarity:	88%	81%	83%	75%

4 Discussion

In this paper we have proposed a novel approach for multi-view fusion of cinematic flow and optical images of mice. The method explores an analysis-by-synthesis approach where a model involving articulations, surface properties and appearance properties is optimized with respect to the different views. Such optimization is done jointly on the visual/optical image space through the certain constancy hypothesis on the bioluminescence imaging. Promising results demonstrate the ability of the method to deal with freely moving animals and enhance the optical imaging signal towards improved preclinical exploitation. Future work consists of introducing explicit modeling of the bioluminescence sources, and a continuous manner on incorporating constancy on the optical imaging space.

References

1. Weissleder, R.: Scaling down imaging: molecular mapping of cancer in mice. Nature Reviews Cancer 2, 11–18 (2002)
2. Gibson, A.P., Hebden, J.C., Arridge, S.R.: Recent advances in diffuse optical imaging. Physics in Medicine and Biology 50(4), R1–R43 (2005)
3. Hildebrandt, I.J., Su, H., Weber, W.A.: Anesthesia and other considerations for in vivo imaging of small animals. ILAR Journal 49(1), 17–26 (2008)
4. Roncali, E., Savinaud, M., Levrey, O., Rogers, K.L., Maitrejean, S., Tavitian, B.: A new device for real time bioluminescence imaging in moving rodents. Journal of Biomedical Imaging 13(5), 054035 (2008)
5. Rogers, K.L., Picaud, S., Roncali, E., Boisgard, R., Colasante, C., Stinnakre, J., Tavitian, B., Brulet, P.: Non-invasive in vivo imaging of calcium signaling in mice. In: PLoS ONE (October 2007)
6. Kuo, C., Coquoz, O., Troy, T.L., Xu, H., Rice, B.W.: Three-dimensional reconstruction of in vivo bioluminescent sources based on multispectral imaging. Journal of Biomedical Optics 12(2), 024007 (2007)
7. Papademetris, X., Dione, D.P., Dobrucki, L.W., Staib, L.H., Sinusas, A.J.: Articulated rigid registration for serial lower-limb mouse imaging. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 919–926. Springer, Heidelberg (2005)
8. Baiker, M., Milles, J., Vossepoel, A., Que, I., Kaijzel, E., Lowik, C., Reiber, J., Dijkstra, J., Lelieveldt, B.: Fully automated whole-body registration in mice using articulated skeleton atlas. In: IEEE ISBI 2007, pp. 728–731 (April 2007)
9. Favreau, L., Reveret, L., Depraz, C., Cani, M.P.: Animal gaits from video. In: ACM SIGGRAPH Symposium on Computer Animation (2004)
10. Gall, J., Stoll, C., de Aguiar, E., Theobalt, C., Rosenhahn, B., Seidel, H.P.: Motion capture using joint skeleton tracking and surface estimation. In: IEEE CVPR 2009, pp. 1746–1753 (June 2009)
11. de LaGorce, M., Paragios, N., Fleet, D.: Model-based hand tracking with texture, shading and self-occlusions. In: IEEE CVPR 2008, pp. 1–8 (June 2008)
12. Magnenat-Thalmann, N., Laperrière, R., Thalmann, D.: Joint-dependent local deformations for hand animation and object grasping, pp. 26–33 (1988)
13. Lewis, J.P., Cordner, M., Fong, N.: Pose space deformation: a unified approach to shape interpolation and skeleton-driven deformation. In: ACM SIGGRAPH, pp. 165–172 (2000)
14. Cook, M.J.: The Anatomy of the Laboratory Mouse. Elsevier, Amsterdam (1965)