Group 5 Thesis

Us

April 3, 2016

Contents

1 Introduction

Imagine an object recorded by a camera and represented in a sequence of 2-dimensional, coloured images. The object may change size due to changes in position (closer to the camera) or even by nature (an inflating balloon). The object may enter a shadow or areas with different lighting causing its brightness and colour to change between images. If the object or the camera moves and rotates, then the location of the object, its shape, and even the background of the object will alter over the sequence of images. Nonetheless, a human could view this sequence of images and associate a region in each image as one object being manipulated.

Region tracking in a sequence of images seeks to create a computer program that, given a region in an initial image, can identify the region over subsequent images. The tracking is desired to be achieved even if some characteristics of the region change. This report proposes two distinct applications for such a program: hand-gesture tracking and tracking of anatomical structures of CT/MRI scans. While hand-gesture recognition will process images whose sequence represents change in time, the CT/MRI scan will process images whose sequence represents change along a dimension of the 3-dimensional object that was scanned.

1.1 Application: Medical Imaging

The main imaging modalities chosen were X-ray Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) scans. These modalities were chosen due to their diagnostic performance and ubiquity of use. For surgeons these scans are an integral portion of planning and preparing for procedures. The ability to automatically segment tissues and render models of them in 3D aids in this process and can also be used to produce implants that fit the patient's anatomy to a greater degree.

1.2 Application: Gesture Tracking

Following a hand through a video sequence can provide individuals with a touchless interface like the Kinect. Using the cheap an ubiquitous web cam the ability to follow gestures can allow a user to interact more fluidly with technology and prevent interruption of some tasks.

2 Prob Def

Produce an algorithm that efficiently tracks regions in medical and video sequences with low error.

3 Math

3.1 Problem Formulation

One of the most important design choices is the initial framing and structureing of the problem. By applying a riggerous mathimatical model to the design problem, a rich body of theory and tools were able to be used and implemented. The framework chosen was that of calculus of variation. This construction ultimatly turns the the problem of region tracking into a funtional minimization problem with convergance garenties.

To veiw region tracking under the the light of calculus of variation, some deffintions are needed first. let $\Omega = [0.1] \times [0,1]$ and let every image I_i in the sequence $i \in \{1,2,...,n\}$ be a mapping $\Omega \to \mathbb{R}$. A region $R_i \subset \Omega$ is an element of 2^Ω the power set of Ω. The problem is now structured as finding R_{l+1} given I_l, I_{l+1} and R_l. To make this problem a functional minimization task, let $E : 2^\Omega \to \mathbb{R}$ such that R_{l+1} is the minimizer of E.

3.2 Mathematical Tools

3.2.1 Euler-Lagrange Necessary Conditions

The machinery of calculus of variation bestows the ability to potentially find a minimizer $\gamma^* = \underset{\gamma \in C}{\operatorname{argmin}} \; E[\gamma]$. let C be the set of simple closed curves defined as:

$$C = \{\gamma : [0,1] \to \mathbb{R} | \gamma \in \mathbf{C}^2, \gamma(0) = \gamma(1), \dot\gamma(0) = \dot\gamma(1)\}$$

Now define E as a mapping $C \to \mathbb{R}$ and express it in the form of:

$$E[\gamma] = \int_0^1 L(s, \gamma(s), \dot\gamma(s))ds \tag{1}$$

From optimization it is clear that the first order condition must be satisfied for an x to be the minimizer of some function $f : \mathbb{R}^n \to \mathbb{R}$. The Euler-Lagrange necessary conditions are the functional analysis equivalent.

Theorem 1. *Let $\gamma \in C$ be a minimizer of E in the form of equation* **??** *then* $\forall s \in [0,1]$:

$$0 = \left.\frac{\partial L}{\partial \gamma}\right|_{(s,\gamma(s),\dot\gamma(s))} - \frac{d}{ds}\left(\left.\frac{\partial L}{\partial \dot\gamma}\right|_{(s,\gamma(s),\dot\gamma(s))}\right) \tag{2}$$

Proof. let ϕ be an element of C and $\alpha \in \mathbb{R}$. As γ is the minimizer of E (implying $E(\gamma) \le E(\theta) \; \forall \theta \in C$), clearly $\frac{d}{d\alpha}E[\gamma + \alpha\phi]|_{\alpha=0} = 0$.

$$0 = \frac{d}{d\alpha} E[\gamma + \alpha\phi]|_{\alpha=0}$$

$$= \frac{d}{d\alpha} \int_0^1 L(s, \gamma(s) + \alpha\phi(s), \dot\gamma(s) + \alpha\dot\phi(s))ds \Big|_{\alpha=0}$$

$$= \int_0^1 \frac{d}{d\alpha} L(s, \gamma(s) + \alpha\phi(s), \dot\gamma(s) + \alpha\dot\phi(s)) \Big|_{\alpha=0} ds$$

$$= \int_0^1 \frac{\partial L}{\partial \gamma}\Big|_{(s,\gamma(s),\dot\gamma(s))} \phi(s) + \frac{\partial L}{\partial \dot\gamma}\Big|_{(s,\gamma(s),\dot\gamma(s))} \dot\phi(s)ds$$

$$= \int_0^1 \frac{\partial L}{\partial \gamma}\Big|_{(s,\gamma(s),\dot\gamma(s))} \phi(s)ds + \int_0^1 \frac{\partial L}{\partial \dot\gamma}\Big|_{(s,\gamma(s),\dot\gamma(s))} \dot\phi(s)ds$$

$$= \int_0^1 \frac{\partial L}{\partial \gamma}\Big|_{(s,\gamma(s),\dot\gamma(s))} \phi(s)ds + \left[\frac{\partial L}{\partial \dot\gamma}\Big|_{(s,\gamma(s),\dot\gamma(s))} \phi(s)\right]_0^1 - \int_0^1 \frac{d}{ds}\left(\frac{\partial L}{\partial \dot\gamma}\Big|_{(s,\gamma(s),\dot\gamma(s))}\right)\phi(s)ds$$

$$= \int_0^1 \frac{\partial L}{\partial \gamma}\Big|_{(s,\gamma(s),\dot\gamma(s))} \phi(s)ds - \int_0^1 \frac{d}{ds}\left(\frac{\partial L}{\partial \dot\gamma}\Big|_{(s,\gamma(s),\dot\gamma(s))}\right)\phi(s)ds$$

$$= \int_0^1 \left(\frac{\partial L}{\partial \gamma}\Big|_{(s,\gamma(s),\dot\gamma(s))} - \frac{d}{ds}\left(\frac{\partial L}{\partial \dot\gamma}\Big|_{(s,\gamma(s),\dot\gamma(s))}\right)\right)\phi(s)ds$$

$$\Rightarrow 0 = \frac{\partial L}{\partial \gamma}\Big|_{(s,\gamma(s),\dot\gamma(s))} - \frac{d}{ds}\left(\frac{\partial L}{\partial \dot\gamma}\Big|_{(s,\gamma(s),\dot\gamma(s))}\right)$$

The right hand side of equation **??** is continuous and the theorem thus holds for any choice of ϕ.

$\square$

The techniques and tools employed in the Euler-Lagrange necessary condition proof provide a viable analogue to the the classic $\mathbb{R}^n$ gradient. Though theorem **??** is a powerful result, it is only a necessary condition on the minimum and does not provide a method to find γ^*.

3.2.2 Gradient Decent

Gradient Decent is one tool that has the ability to find critical points of the functional. To begin the $\mathbb{R}^n$ gradient decent schema will be illustrated then extended to the C-functional gradient decent system.

Theorem 2. *Let $f : \mathbb{R}^n \to \mathbb{R}$ such that f is differentiable. Let ν be a mapping $\mathbb{R}^+ \to \mathbb{R}^n$ defined by:*

$$\frac{d\nu(x)}{dt}(t) = -\vec\nabla f(\nu(t)) \tag{3}$$

Then $f(\nu(t))$ is a monotonically decreasing function $\forall t \in \mathbb{R}^+$.

Proof.

$$\frac{d}{dt}f(\nu(t)) = \langle \vec{\nabla} f(\nu(t)), \frac{d}{dt}\nu(t) \rangle$$
$$= \langle \vec{\nabla} f(\nu(t)), -\vec{\nabla} f(\nu(t)) \rangle$$
$$= -\langle \vec{\nabla} f(\nu(t)), \vec{\nabla} f(\nu(t)) \rangle$$
$$= -\|\vec{\nabla} f(\nu(t))\|^2$$
$$\leq 0$$

thus $f(\nu(t))$ is monotonically decreasing $\forall t \in \mathbb{R}^+$; with equality only at critical points of f. $\qquad\square$

Clearly by following this ordinary differential equation **??**, it is guaranteed that the system approaches a minimum or critical point.

To extend these results to functionals, first define the set of evolving simple closed curves.

$$C_e = \{\gamma : [0,1] \times \mathbb{R}^+ \to \mathbb{R} | \gamma \in \mathbf{C}^2, \gamma(0,t) = \gamma(1,t), \dot{\gamma}(0,t) = \dot{\gamma}(1,t), t \in \mathbb{R}^+\} \tag{4}$$

Theorem 3. *Let E be functional mapping C to $\mathbb{R}$ and let γ be an element of C_e such that:*

$$\left.\frac{d\gamma}{dt}\right|_{(s,t)} = -\left(\left.\frac{\partial L}{\partial \gamma}\right|_{(s,\gamma(s,t),\dot{\gamma}(s,t))} - \frac{d}{ds}\left(\left.\frac{\partial L}{\partial \dot{\gamma}}\right|_{(s,\gamma(s,t),\dot{\gamma}(s,t))}\right)\right) \tag{5}$$

Then $E(\gamma(t))$ is a monotonically decreasing function $\forall t \in \mathbb{R}^+$.

Proof.

$$\frac{d}{dt}E[\gamma(\cdot,t)] = \frac{d}{dt}\int_0^1 L(s,\gamma(s,t),\dot\gamma(s,t))ds$$

$$= \int_0^1 \frac{d}{dt}L(s,\gamma(s,t),\dot\gamma(s,t))ds$$

$$= \int_0^1 \frac{\partial L}{\partial \gamma}\frac{\partial \gamma}{\partial t} + \frac{\partial L}{\partial \dot\gamma}\frac{\partial \dot\gamma}{\partial t}ds$$

$$= \int_0^1 \frac{\partial L}{\partial \gamma}\frac{\partial \gamma}{\partial t} + \frac{\partial L}{\partial \dot\gamma}\frac{\partial^2 \gamma}{\partial s\partial t}ds$$

$$= \int_0^1 \frac{\partial L}{\partial \gamma}\frac{\partial \gamma}{\partial t}ds + \left[\frac{\partial L}{\partial \dot\gamma}\frac{\partial \gamma}{\partial t}\right]_0^1 - \int_0^1 \frac{d}{ds}\left(\frac{\partial L}{\partial \dot\gamma}\frac{\partial \gamma}{\partial t}\right)ds$$

$$= \int_0^1 \left(\frac{\partial L}{\partial \gamma} - \frac{d}{ds}\left(\frac{\partial L}{\partial \dot\gamma}\right)\right)\frac{\partial \gamma}{\partial t}ds$$

$$= \int_0^1 \left(\frac{\partial L}{\partial \gamma} - \frac{d}{ds}\left(\frac{\partial L}{\partial \dot\gamma}\right)\right)^2 ds$$

$$\leq 0$$

Then $E(\gamma(t))$ is a monotonically decreasing function $\forall t \in \mathbb{R}^+$. $\square$

These tools are now sufficient to begin designing and implementing an engineering solution to the problem.

3.3   Functionals

Give that 2^Ω is a difficult set to work with and regions tend to consist of contiguous sections of ω, it is reasonable to restrict regions to be those outlined by simple closed curves γ. This restriction opens up the powerful toolbox of calculus of variation. Within this framework gradient decent can be employed to numerically move towards the minima of the functionals E. Under this structure multiple functionals E were designed for specific assumptions on the region characteristics.

3.3.1   The Length Functional

Under the assumption that that the region has "smooth" boundaries penalizing length of the curve is beneficial.

$$E_L(\gamma) = \int_0^1 \|\dot\gamma(s)\|ds$$

3.3.2   The Nth Moment Functional

By assuming the distribution of gray scale within the image is conserved between regions the moment functional was devised. this functional ensures the Nth

moment is the same between regions and thus as the number of momnets used the distribution is conserved between the two.

$$E_{nM}(\gamma) = \left(\frac{\int_{R_\gamma} I_{l+1}(x)^n dx}{\int_{R_\gamma} dx} - a \right)^2 + \left(\frac{\int_{R_\gamma^c} I_{l+1}(x)^n dx}{\int_{R_\gamma^c} dx} - b \right)^2$$

3.3.3   The Center of Mass Functional

Many regions maintain a similar position in the image but may deform in space. By assuming relativity constant center of mass this trait is built int the functional.

$$E_{CM}(\gamma) = \left(\frac{\int_{R_\gamma} x dx}{\int_{R_\gamma} dx} - a \right)^2$$

3.3.4   The Area Functional

the last static property that was assumed was that area of the region is the same between images.

$$E_A(\gamma) = \left(\int_{R_\gamma} dx - a \right)^2$$

3.4   Level Sets

Though simple closed curves are very intuitive to use, in practice their discretization and their instability for numerical gradient decent make them impractical. Level sets of of a function $u : \Omega \to \mathbb{R}$ can be used to define the region. In this technique the region $R_1 = x : x \in \Omega | u(x) \geq 0$. Through the use of Green's Theorem it is possible to convert from a simple closed curve perspective to a level set perspective.

Proposition 1. *Let $\gamma \in C_e$ evolve according to the ordinary differential equation described in theorem* **??**:

$$\left. \frac{d\gamma}{dt} \right|_{(s,t)} = - \left(\left. \frac{\partial L}{\partial \gamma} \right|_{(s,\gamma(s,t),\dot{\gamma}(s,t))} - \frac{d}{ds} \left(\left. \frac{\partial L}{\partial \dot{\gamma}} \right|_{(s,\gamma(s,t),\dot{\gamma}(s,t))} \right) \right)$$

Let $\vec{N}_\gamma$ be the unit normal to γ, Then there exists a function $F : \Omega \to \mathbb{R}$ such that:

$$\left. \frac{d\gamma}{dt} \right|_{(s,t)} = F(\gamma(s,t)) \cdot \vec{N}_\gamma \tag{6}$$

4   Design

The functional pieces described in Section **??** had to be combined to form complete functionals. These functionals were created based on the invariant properties of the image sets. Of course, medical images and webcam-captured hand gesture images emphasize different invariant properties, thus functionals with different weighting on the functional pieces had to be designed for each application domain.

4.1   Medical Images

On the medical imaging side of the design, our project focused on high-contrast medical images such as those created by CT scans of bone. On these types of images, we were able to exploit high colour invariance of both the inside and outside of the outlined structure. To capture the colour properties of the images, we needed to heavily weigh inside and outside mean colour. Additionally, we also included the variance property in our functional, although this was weighted to a much weaker extent. The mean colour functional component captured the fact that the average colour on the inside and outside stayed relatively constant between images. The colour variance functional component captured a higher order term for the distribution of the pixel colours. On top of the mean and variance functionals, the length functional was added to stabilize the gradient descent.

4.2   Webcam-Generated Hand Gesture Images

Faisal

5   Implementation

Zach

6   Results

6.1   Evaluation Criteria

The standard method in medical imaging for evaluating the accuracy of segmentation algorithms is calculating the similarity of the resulting region compared to that produced by manual segmentation [1]. Since these are taken to be the most accurate possible segmentations, results produced manually are known as the ground truth.

6.2   Final Performance on Medical Images

The results of the algorithm using the final optimized parameters are shown in Figures **??**, **??**, and **??**.
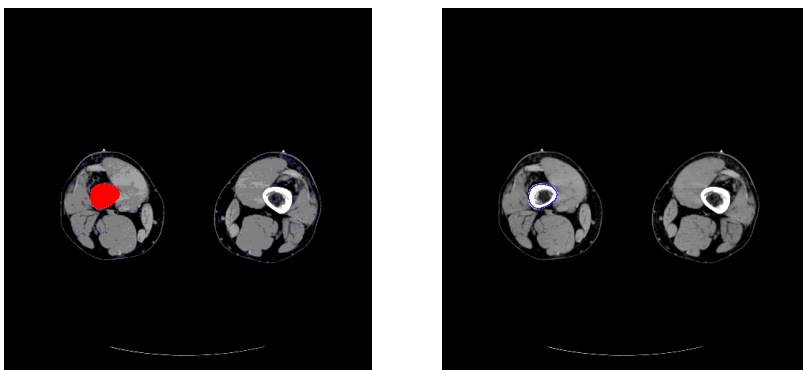
Figure 1: Hip image 25 – ground truth (left) compared to algorithm output with final parameter set (right).
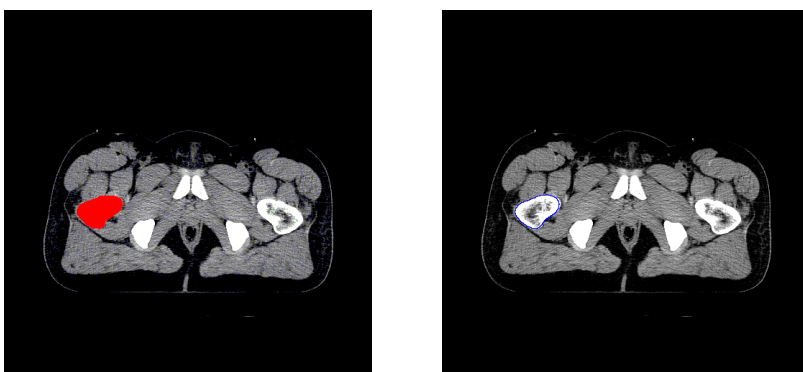

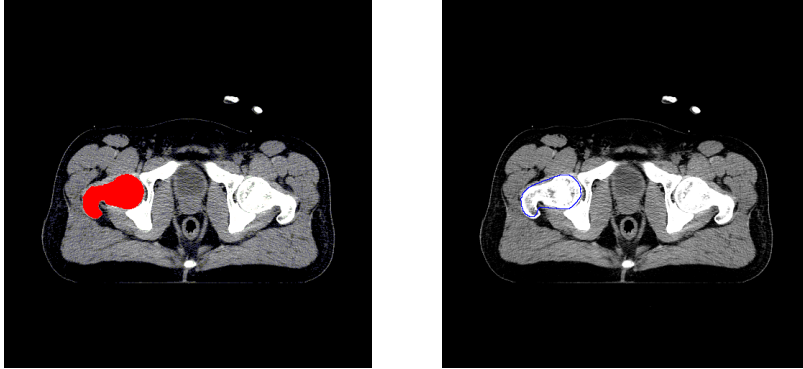**Figure 2:** Hip image 350 – ground truth (left) compared to algorithm output with final parameter set (right).

Figure 3: Hip image 375 – ground truth (left) compared to algorithm output with final parameter set (right).

On image 25 (Figure **??**), the algorithm incorrectly identified 4.13% of the image pixels by volume of the ground truth. This corresponds to an incorrect identification of 0.017% of the image pixels over the entire image. As the algorithm proceeded to subsequent images, the error rate did not increase significantly until the shape of the hip bone changed drastically around the 375th image. This is evidenced by the similar error rate obtained in image 350 (Figure **??**) compared to image 25. In image 350, the percentage of erroneous pixels was 3.82% of the volume of the ground truth, and 0.041% of volume of the entire image. However, by image 375 (Figure **??**), these error rates increased to 13.12% and 0.17%, respectively.

Overall, the algorithm error rate with these final parameters was averaged over the 25th, 75th, 350th, and 375th images. Due to the amount of manual work required to determine ground truth, the evaluation of the algorithm could only be measured over a subset of the images. The images near the beginning and the end of the sequence were assumed to be most revealing of algorithm performance due to gradual shape changes in this dataset. Over this subset, the average error rate as a fraction of ground truth was 6.00%. The average error rate as a fraction of the entire image was 0.056%. While we were aiming to have the average error rate over ground truth lower than 5%, so the algorithm's final performance did not quite reach our expectations in this regard. However, we were also targeting for our algorithm to have less than a 0.1% error rate over the entire image region, and it outperformed in this metric. See Section **??** for areas where the algorithm could have been improved to lower the error rate.

6.3 Final Performance on Hand Gesture Images

7 Discussion (3bottom)

Mandy

8   Conclusion

9   Future Work

Faisal