

## Project Proposal

Division of Applied Mathematics, Department of Mathematics

Faculty of Science, Silpakorn University

Date : 27 กันยายน 2561

Advisor : ผู้ช่วยศาสตราจารย์ ดร. นพดล ชุมชอบ

Student : นายภัคพล พงษ์ทวี รหัส 07580028

Project Title : ลบบทบรรยายแบบเข้ารหัสวิดีโอแบบอนิเมะ  
(Hard-coded subtitle remover for anime)

## 1 Introduction

บทบรรยายใต้ภาพ (subtitle) คือข้อความที่ได้จากการถอดเสียงจาก ภาพยนตร์ รายการทีวี หรือสื่อต่างๆ ที่มักแสดงอยู่ด้านล่างของหน้าจอ ซึ่งใช้เพื่อแสดงคำแปลในภาษาอื่น หรือเพื่อช่วยเหลือผู้ประสบปัญหาทางการได้ยิน โดยบทบรรยายใต้ภาพ แบ่งตามการสร้างได้เป็น 2 ประเภทได้แก่ แบบ Soft และแบบ Hard โดยแบบ Soft จะมีลักษณะเป็นข้อความที่ถูกระบุตำแหน่งเวลาไว้สามารถเปิดและปิดได้ แต่ก็มีปัญหาว่าเครื่องเล่นหรือซอฟต์แวร์ที่จำเป็นต้องรองรับรูปแบบบทบรรยายนั้นจึงจะแสดงขึ้นมาได้ และในบางครั้งก็มีปัญหาด้านการเข้ารหัส ทำให้ไม่สามารถแสดงข้อความได้อย่างถูกต้องและแบบ Hard คือ ข้อความในบทบรรยายจะถูกฝังรวมเป็นเนื้อเดียวกับวิดีโอ ซึ่งวิธีนี้จะทำให้ไม่จำเป็นต้องใช้เครื่องเล่นหรือซอฟต์แวร์พิเศษเพื่อรองรับ ทำให้ไม่มีปัญหาในการแสดงผล แต่ว่าบทบรรยายแบบนี้จะไม่สามารถทำการเปิดและปิดได้

จากการที่บทบรรยายใต้ภาพแบบ Hard ไม่สามารถเปิดและปิดได้ทำให้บทบรรยายดังกล่าวก่อความรำคาญให้กับผู้รับชม ตัวอย่างเช่น ภาพยนตร์จีนกำลังภายในที่ฉายทางโทรทัศน์ไทยมักจะมียกบรรยายภาษาจีนติดมาด้วย เนื่องจากประเทศจีนมีการใช้ภาษาจีนในสำเนียงที่หลากหลาย จึงมีบทบรรยายให้สามารถรับชมได้โดยเข้าใจความหมายที่เหมือนกัน โดยคำบรรยายที่ใช้นั้น เป็นบทบรรยายแบบ Hard ซึ่งเมื่อภาพยนตร์ดังกล่าวได้รับการให้เสียงไทยแล้ว ภาพยนตร์จึงมีบทบรรยายภาษาจีนติดมาด้วย

ด้วยปัญหาที่กล่าวมาข้างต้นโดยผู้ศึกษาได้มีความสนใจที่จะลบบทบรรยายแบบ Hard ออกจากวิดีโอ โดยเฉพาะวิดีโอแบบอนิเมะ เนื่องจากวิดีโอประเภทนี้มักมีบทบรรยายติดมาด้วย และบางครั้ง มีบทบรรยายเป็นแบบ Hard ทำให้ไม่สามารถเปิดและปิดได้ ซึ่งการลบบทบรรยายนี้ ซึ่งการลบบทบรรยายออกจากวิดีโออนิเมะ เปรียบได้กับการซ่อมภาพศิลปะที่เสียหาย ในอนาคตสามารถพัฒนาต่อยอดได้เป็นแอปพลิเคชันที่จะสามารถนำโทรศัพท์มือถือส่งไปยังงานศิลปะที่เสียหายและทำการซ่อมแซมภาพนั้นและแสดงขึ้นบนหน้าจอแบบเรียลไทม์ได้

โดยวิดีโอั้นประกอบด้วยภาพจำนวนหลายภาพต่อหนึ่งหน่วยเวลา เราจะเรียกภาพหนึ่งภาพในวิดีโอว่า เฟรม (frame) ซึ่งภาพในแต่ละเฟรมเรียกว่าภาพดิจิทัล (Digital Image) โดยภาพดิจิทัลนั้นจะสามารถนิยามได้เป็นฟังก์ชัน  $f(x, y)$  โดยที่  $x$  และ  $y$  เป็นพิกัดของภาพ และแอมพลิจูดของ  $f$  ที่พิกัด  $(x, y)$  ใดๆ ในภาพคือค่าความเข้มแสง (intensity) ซึ่งแอมพลิจูดนี้มีค่าจำกัด

ซึ่งก่อนจะลบบทบรรยายนั้น จะต้องหาบทบรรยายในภาพให้ได้เสียก่อน โดยบทบรรยายของอนิเมะนั้น มักจะขึ้นบริเวณด้านล่างของหน้าจอ และนอกจากนี้ บทบรรยายอนิเมะมักจะใช้ขอบของตัวอักษรเป็นสีตัวอีกด้วย ด้วยสมบัตินี้เองทำให้เราสามารถหาบริเวณบนเฟรมที่เป็นบทบรรยายได้โดยจะมีวิธีหาพื้นที่ซึ่งเป็นบทบรรยายดังนี้



(a) ภาพเฟรมอนิเมะที่มีบทบรรยาย



(b) ภาพหลังทำการตัดส่วนกลางและ thresholding

ตัดเฟรมมาเฉพาะส่วนกลางของเฟรมที่น่าจะมีบทบรรยายปรากฏอยู่ จากนั้นทำการ thresholding เพื่อหาบริเวณที่เป็นสีดำ เนื่องจากบทบรรยายจะถูกล้อมรอบด้วยสีดำเสมอ



(a) ภาพหลังทำการสลับสี



(b) ภาพหลังทำการเปลี่ยนพื้นที่สีขาว

ทำการสลับสีระหว่างสีดำกับสีขาวของภาพที่ทำการ thresholding หลังจากนั้นทำการเปลี่ยนพื้นที่สีขาวซึ่งติดกับขอบของเฟรมทั้งหมดให้เป็นสีดำ เพราะว่า บทบรรยายไม่อยู่ติดกับหน้าจอ เราจะถือว่าสิ่งที่อยู่ติดกับหน้าจอไม่ใช่บทบรรยาย



(a) ภาพหลังการ erode และ opening



(b) ภาพหลังการ dilate

จากนั้นนำวัตถุที่มีขนาดเล็กเกินไป หรือใหญ่เกินไปออกจากภาพด้วยวิธีการ erode และ opening จะได้ว่าส่วนที่เหลือเป็นสีขาวในภาพคือบทบรรยาย แต่ว่าขอบของบทบรรยายก็ต้องถูกลบออกไปด้วย จึงทำการ dilate เพื่อขยายขอบของบทบรรยายให้เท่ากับบทบรรยายที่อยู่ในเฟรมวิดีโอ และสิ่งที่เหลืออยู่คือ inpaint domain ที่จะนำไปใช้ในการซ่อมแซมภาพต่อไป

การซ่อมแซมภาพ คือกระบวนการที่จะเติมเต็มข้อมูลที่หายไปในพื้นที่ภาพที่กำหนด โดยมีจุดประสงค์เพื่อซ่อมแซมภาพที่เสียหาย โดยพื้นที่ภาพส่วนนั้นไม่สามารถพบได้จากการสังเกต โดยการกู้คืน สี, โครงสร้าง และพื้นผิว ที่เกิดการเสียหายเป็นวงกว้าง พิกเซลที่จะนำมาใช้ซ่อมแซมจะถูกคำนวณขึ้นมาใหม่จากข้อมูลที่พิกเซลที่อยู่โดยรอบที่ยังไม่เสียหาย ซึ่งการจะนำบทบรรยายออกจากเฟรมวิดีโอ นั้น จะพิจารณาว่าบทบรรยายนั้นเป็นส่วนที่เสียหาย แล้วจากนั้นจึงใช้การซ่อมแซมภาพเพื่อนำบทบรรยายนั้นออก

การซ่อมแซมภาพมีวิธีการทางคณิตศาสตร์ที่ใช้แตกต่างกันไปจำนวนมาก แต่เนื่องจากภาพอนิเมะที่เราต้องการซ่อมแซมนั้นเป็นรูปภาพที่ราบเรียบเป็นช่วง (piecewise smooth image) จึงเหมาะที่จะใช้วิธีการซ่อมภาพด้วยการแปรผัน (Variation)

โดยโมเดลทางคณิตศาสตร์ที่จะใช้งานในการซ่อมแซมรูปภาพนี้คือ การแปรผันรวม (Total Variation) ซึ่งมีที่มาจากปัญหา Rudin-Osher-Fatemi (ROF) โดยเป็นการแก้ปัญหาการแปรผันมีขอบเขต (bounded variation หรือ BV) ทั้งหมดโดยที่ภาพ  $u$  อยู่ใน  $BV(\Omega)$  เมื่อสามารถหาปริพันธ์ได้และจะมี Radon measure  $Du$  ซึ่ง

$$\int_{\Omega} u(x) \operatorname{div} \vec{g}(x) dx = \int_{\Omega} \langle \vec{g}, Du(x) \rangle \quad \forall \vec{g} \in C_c^1(\Omega, \mathbb{R}^2)^2$$

และจาก  $Du$  เป็น distributional gradient ของ  $u$  เมื่อ  $u$  ราบเรียบแล้ว  $Du(x) = \nabla u(x) dx$  โดย total variation seminorm ของ  $u$  คือ

$$\|u\|_{TV(\Omega)} := \int_{\Omega} |Du| := \sup \left\{ \int_{\Omega} u \operatorname{div} \vec{g} dx : \vec{g} \in C_c^1(\Omega, \mathbb{R}^2)^2, \sqrt{g_1^2 + g_2^2} \leq 1 \right\}$$

จาก  $u$  ราบเรียบแล้ว การแปรผันรวมสมมูลกับอินทิกรัลของขนาดเกรเดียนท์

$$\|u\|_{TV(\Omega)} = \int_{\Omega} |\nabla u| dx$$

จึงได้ว่าจะหาฟังก์ชันแปรผันมีขอบเขต  $u$  หาได้จาก minimization problem

$$\arg \min_{u \in BV(\Omega)} ||u||_{TV(\Omega)} + \frac{\lambda}{2} \int_{\Omega \setminus D} (f(x) - u(x))^2 dx$$

เมื่อ  $\lambda$  มีค่าบวก ปัญหา minimalization นี้จะเหมือนกับปัญหาการลบสิ่งรบกวนของ Rudin, Osher และ Fatemi เพียงแต่ปริพันธ์ลำดับอยู่บน  $\Omega - D$  แทนที่จะเป็น  $\Omega$  ถ้าผลลัพธ์ที่แน่นอนตรงอยู่ใน  $BV$  และมีค่าอยู่ในช่วง  $[0, 1]$  แล้วจะมี minimizer  $u$  แต่มักจะไม่มีเพียงหนึ่งเดียว

การซ่อมแซมรูปภาพอาจมองเป็นลักษณะการลบสิ่งรบกวนที่มี spatially-varying regularization strength เป็น  $\lambda(x)$  ทำให้ได้ว่า

$$\arg \min_u ||u||_{TV(\Omega)} + \frac{1}{2} \int_{\Omega} \lambda(x)(f(x) - u(x))^2 dx$$

โดยที่  $\lambda(x)$  จะมีค่าเป็น 0 เมื่ออยู่ใน  $D$  และ  $\lambda(x) > 0$  เมื่ออยู่นอก  $D$  ทำให้เมื่อ  $x \in D$  ที่  $\lambda(x) = 0$  ค่า  $f(x)$  จะไม่ถูกใช้ ทำให้  $u(x)$  ได้รับผลจาก  $||u||_{tv}$  เท่านั้น ส่วนที่ด้านนอก  $D$  จะเป็น TV-regularize denoising พฤติกรรมลดสิ่งรบกวนนี้อาจเป็นที่น่าพอใจเมื่ออยากที่จะระบุโดเมนที่ต้องซ่อมแซมได้อย่างถูกต้อง และเมื่อใช้ ขนาดใหญ่จะทำให้การลดสิ่งรบกวนมีผลน้อยมากจนทำให้พื้นที่นอก  $D$  แทบไม่เปลี่ยนแปลง

จากโมเดลทางคณิตศาสตร์ที่ได้กล่าวมาข้างต้น จะสามารถใช้วิธีการทางเชิงตัวเลขสำหรับการซ่อมแซมรูปภาพโดยใช้ความแปรปรวนทั้งหมด

จากความแปรปรวนทั้งหมดสามารถประมาณได้โดย  $|\nabla u_{i,j}|$  บนทุกพิกเซลนั้นคือ

$$||u||_{TV(\Omega)} \approx \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} |\nabla u_{i,j}|$$

เมื่อ  $\nabla u_{i,j}$  คือ discrete gradient วิธี split bergman คือการแยกส่วนการดำเนินการ (splitting) และการทำซ้ำ bergman (bergman iteration) ซึ่งวิธี split bergman จะนำมาใช้เพื่อแก้ minimization problem

$$\begin{cases} \arg \min_{d,u} \sum_{i,j} |d_{i,j}| + \frac{1}{2} \sum_{i,j} \lambda_{i,j} (f_{i,j} - u_{i,j})^2 \\ \text{subject to } d = \nabla u \end{cases}$$

โดยตัวแปรช่วย  $d$  คือเวกเตอร์ที่บีบบังคับ  $\nabla u$  และใช้วิธีการทำซ้ำ bergman เพื่อแก้ปัญหาที่เหมาะสมแบบมีข้อจำกัด ซึ่งในแต่ละการทำซ้ำ bergman จะเป็นการแก้

$$\arg \min_{d,u} \sum_{i,j} |d_{i,j}| + \frac{\lambda}{2} \sum_{i,j} \lambda_{i,j} (f_{i,j} - u_{i,j})^2 + \frac{\gamma}{2} \sum_{i,j} |d_{i,j} - \nabla u_{i,j} - b_{i,j}|^2$$

เมื่อ  $b$  เป็นตัวแปรของวิธีการทำซ้ำ bergman และ  $\gamma$  เป็นค่าคงที่บวกใดๆ โดยการ minimization บน  $d$  และ  $u$  จะแก้โดย alternative direction method โดยแต่ละขั้นของการหาค่าต่ำสุด ตัวแปร  $d$  และ  $u$  จะให้ตัวแปรอื่นคงค่าไว้

d subproblem เมื่อเราคงค่า  $u$  ไว้ จะได้ว่า d subproblem คือ

$$\arg \min_d \sum_{i,j} |d_{i,j}| + \frac{\gamma}{2} \sum_{i,j} |d_{i,j} - \nabla u_{i,j} - b_{i,j}|^2$$

โดยปัญหานี้เมื่อทำการแก้แล้วจะได้ว่า

$$d_{i,j} = \frac{\nabla u_{i,j} + b_{i,j}}{|\nabla u_{i,j} + b_{i,j}|} \max\{|\nabla u_{i,j} + b_{i,j}| - \frac{1}{\gamma}, 0\}$$

u subproblem เมื่อเราคงค่า  $d$  ไว้ จะได้ว่า u subproblem คือ

$$\arg \min_u \frac{1}{2} \sum_{i,j} \lambda_{i,j} (f_{i,j} - u_{i,j})^2 + \frac{\gamma}{2} \sum_{i,j} |d_{i,j} - \nabla u_{i,j} - b_{i,j}|^2$$

เมื่อแก้แล้วจะได้ว่า

$$\frac{1}{\gamma} - \Delta u = \frac{1}{\gamma} \lambda f - \text{div}(d - b)$$

โดยที่  $\text{div}$  คือ discrete divergence และ  $\nabla u$  คือ discrete laplacian เราจะประมาณคำตอบนี้โดยใช้หนึ่งรอบ Gauss-seidel ต่อหนึ่งรอบการทำซ้ำของ Bergman ซึ่ง subproblem จะถูกแก้หนึ่งครั้ง ต่อหนึ่งรอบ bergman iteration แต่ทั้งนี้ การทำซ้ำ Gauss-seidel หลายครั้ง จะทำให้การแก้ subproblem มีความแม่นยำขึ้น ส่วนตัวแปรช่วย  $b$  มีค่าเริ่มต้นเป็น 0 จากนั้นทำการปรับค่าโดย

$$b^{k+1} = b^k + \nabla u - d$$

โดยที่ความเกี่ยวข้องกันของแต่ละพื้นที่จะแรงขึ้นเมื่อ  $\gamma$  ใหญ่ขึ้น ดังนั้น  $\gamma$  ไม่ควรเล็กหรือใหญ่จนเกินไป จะทำให้ทั้งสอง subproblem ลู่เข้าได้ดี จึงได้ว่าวิธีการในภาพรวมเป็นดังนี้

```

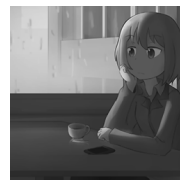
initialization  $u = 0, d = 0, b = 0$ 
while  $\|u_{cur} - u_{prev}\|_2 > Tol$  do
    Solve the  $d$  subproblem
    Solve the  $u$  subproblem
     $b = b + \nabla u - d$ 
end

```

โดยการทำซ้ำนี้จะกระทำจนกระทั่ง นอร์ม L2 ระหว่างรอบปัจจุบันต่างกับรอบก่อนหน้าไม่เกินค่า Tol ที่กำหนดไว้หรือจำนวนรอบการทำซ้ำมากจนถึงจุดสิ้นสุดที่เพียงพอที่จะให้ลู่เข้าซึ่งไม่ควรใหญ่เกินไปเพื่อไม่ให้เสียเวลาประมวลผลจนนานเกินไป โดยวิธีการซ่อมแซมภาพด้วยการแปรผันรวมได้ผลลัพธ์ดังนี้



(a) ภาพก่อนทำการซ่อมแซม



(b) ภาพหลังซ่อมแซมด้วยการแปรผันรวม

## 2 Objective

วัตถุประสงค์ของโครงการวิจัยมีดังต่อไปนี้

- (1) เพื่อลดคำบรรยายแบบ Hard ออกจากวิดีโออนิเมะ

## 3 Scope of Study

ขอบเขตของโครงการมีดังต่อไปนี้

- (3.1) วิดีโอที่ใช้ศึกษาเป็นวิดีโอประเภทอนิเมะเท่านั้น

(3.2) บทบรรยายที่ใช้ทดสอบ จะถูกล้อมรอบไว้ด้วยสีดำ ขนาดความหนาขนาดไม่น้อยกว่า 5 พิกเซล

(3.3) จะทำการทดสอบการลบคำบรรยายบน 3 ภาษาได้แก่ ไทย จีน และอังกฤษ

## 4 Methodology

วิธีการมีดังต่อไปนี้

(4.1) พัฒนาวิธีการหาบทบรรยาย

(4.2) ปรับปรุงความเร็วของ split begman

(4.3) พัฒนา Avisynth Plugin สำหรับใช้กับวิดีโอ

(4.4) เปรียบเทียบ Plug in ที่ใช้วิธีการแปรผันรวมด้วยวิธี split bergman กับวิธีอื่น

(4.5) เขียนรายงานโครงการวิจัย

## 5 Time Periods

แผนการดำเนินงานตลอดทั้งโครงการสามารถสรุปได้โดยย่อจากตารางต่อไปนี้

| แผนการดำเนินงาน                                | เดือนที่ |   |   |   |   |   |   |   |   |    |    |    |
|--|----------|---|---|---|---|---|---|---|---|----|----|----|
|  | 1        | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| พัฒนาวิธีการหาบทบรรยาย                         | x        | x |   |   |   |   |   |   |   |    |    |    |
| ปรับปรุงความเร็วของ split begman               |          |   | x | x |   |   |   |   |   |    |    |    |
| พัฒนา Avisynth Plugin สำหรับใช้กับวิดีโอ       |          |   |   |   | x | x |   |   |   |    |    |    |
| เปรียบเทียบการลบข้อวิธีการแปรผันรวมกับวิธีอื่น |          |   |   |   |   |   | x | x |   |    |    |    |
| เขียนรายงานโครงการวิจัย                        |          |   |   |   |   |   |   |   | x | x  | x  | x  |

## 6 References

พิมพ์เอกสารอ้างอิงในหัวข้อนี้