Image Registration using Primal-Dual interior point method

Ivan Bogun
Florida Institute of Technology
ibogun2010@my.fit.edu

April 30, 2014
Overview

1. Introduction

2. Feature extraction & matching
   - Feature definition
   - Feature matching
   - Learning problem setup
   - Robust least-squares

3. Interior point Primal-Dual algorithm

4. Results
   - Synthetic data
   - Application: image stitching
Introduction
Goal: given two images find a transformation from one to another.
Applications

- Augmented reality
- 3D tourism
- Computational photography: panoramas, image stitching etc.

Objective

- perform global image registration (under affine model)
- use local feature for initial matching
- calculate transformation using optimization techniques
Feature extraction & matching
Local features

**Definition**

*Local feature* in an image is a salient discriminative region in the image.
Local features

Definition

*Local feature* in an image is a salient discriminative region in the image.

Feature descriptor

*Feature descriptor* is a vector calculated in the feature region capable of qualitative comparison.
Local features

Definition

*Local feature* in an image is a salient discriminative region in the image.

Feature descriptor

*Feature descriptor* is a vector calculated in the feature region capable of qualitative comparison.

Examples of local features: SIFT, SURF, RIFT and more.
Definition

Scale-invariant feature transform is a method for local features extraction which is invariant to scale and rotation introduced by Lowe (2004)²

---

SIFT

Definition

*Scale-invariant feature transform* is a method for local features extraction which is invariant to scale and rotation introduced by Lowe (2004)\(^2\)

Algorithm

1. Extract extremas from difference of the Gaussians across different scales.
2. Eliminate low-contrast keypoints
3. Eliminate edge responses
4. Calculate descriptor on a region around keypoint

\(^2\)Lowe, “Distinctive image features from scale-invariant keypoints”.
SIFT example

SIFT features on the image
Feature matching

Let $D_1, D_2$ be two sets of SIFT feature extracted from two different images. Let $d_{i}^{j}$ be feature $i$ from the image $j$, set

$$d^* = \arg \min_{d \in D_2} \|d_{i}^{2} - d\|^2$$  \hspace{1cm} (1)

$d_{i}^{1}$ is matched against $d^*$ only if

$$\|d_{i}^{1} - d^*\|^2 \gamma \geq \min_{d \in D_2 \setminus d^*} \|d_{i}^{1} - d\|^2$$  \hspace{1cm} (2)

with $\gamma > 1$
Example: feature matching
Assume the transformation between images is in the form:

\[ [x, y]^T \rightarrow [c_{11}x + c_{12}y + c_{13}, c_{21}x + c_{22}y + c_{23}]^T \]  \hspace{1cm} (3)

Let \( X \) be a set of matched features where \( (x_1^1, x_2^1) \in X \) is a pair of matched feature from both images.
Assume the transformation between images is in the form:

\[
[x, y]^T \rightarrow [c_{11}x + c_{12}y + c_{13}, c_{21}x + c_{22}y + c_{23}]^T
\] (3)

Let \( X \) be a set of matched features where \((x_1^1, x_1^2) \in X\) is a pair of matched feature from both images.

Given \( X \) how can we find \( c \) ?
Naive way

Least-squares

Find solution in a least-squares sense:

$$\min_{c} \sum_{(x_i^1, x_i^2) \in X} \|x_i^1 c - x_i^2\|^2$$ (4)
Naive way

Least-squares

Find solution in a least-squares sense:

$$\min_c \sum_{(x_1^i, x_2^i) \in X} \|x_1^i c - x_2^i\|^2$$

(4)

Problems: any two feature matched incorrectly will serve as an outlier in the fitting and as a result will significantly influence final transformation.
Classical solution

**Idea**

Instead of the whole set of matches use only subset.

---

Idea

Instead of the whole set of matches use only subset.

Ransac

RANSAC\(^3\) (Random sampling consensus) is an iterative technique which consists of the following

1. Get subset of matched features
2. Solve 14 on that subset
3. Repeat until fit is good enough.

---

\(^3\)Fischler and Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography”.
Proposed method

Idea

Use loss function which is immune to outliers.

Robust loss functions

- Huber loss: $H_M(x) = \begin{cases} \|x\|^2 & : \|x\| \leq M \\ M(2\|x\|-M) & : \|x\| > M \end{cases}$

- Absolute value: $H_{abs}(x) = |x|$

- Turkey biweight: $H_c(x) = \begin{cases} x(1-x^2c^2)^2 & : |x| < c \\ 0 & : |x| > c \end{cases}$
Proposed method

Idea

Use loss function which is immune to outliers.

Robust loss functions

Huber loss:

\[ H_M(x) = \begin{cases} 
\|x\|^2 & : \|x\| \leq M \\
M(2\|x\| - M) & : \|x\| > M 
\end{cases} \]

Absolute value:

\[ H_{abs}(x) = |x| \] (5)

Turkey biweight

\[ H_c(x) = \begin{cases} 
x(1 - \frac{x^2}{c^2})^2 & : |x| < c \\
0 & : |x| > c 
\end{cases} \]
Loss functions plots
Huber loss properties

**Theorem**

With $M = 1.345$ least-squares with Huber is asymptotically 95% as efficient as least-squares solution and more efficient in other cases.
Huber loss properties

**Theorem**

With $M = 1.345$ least-squares with Huber is asymptotically 95% as efficient as least-squares solution and more efficient in other cases.

**Robustness to outliers**

When residual is high Huber loss function grows *linearly* as opposed to quadratically in least-squares solution.
Robust least-squares

**Definition**

*Robust least-squares* is a problem defined as:

$$\min_x \sum_{i=1}^{m} \phi_M(a_i^T x - b)$$  \hspace{1cm} (6)

where $\phi_M(\cdot)$ is a robust loss function.
Robust least-squares

Definition

Robust least-squares is a problem defined as:

\[
\min_x \sum_{i=1}^{m} \phi_M(a_i^T x - b)
\]  \hspace{1cm} (6)

where \( \phi_M(\cdot) \) is a robust loss function.

Problem equivalence

Robust least-squares with Huber loss is equivalent to:

\[
\min_{x,w} \sum_{i=1}^{m} \frac{\|a_i^T x - b\|^2}{(1 + w_i)} + M^2 \sum_{i=1}^{m} w_i
\]  \hspace{1cm} (7)

s.t. \( w_i \geq 0 \quad \forall i = 1, \ldots, m \)  \hspace{1cm} (8)
Interior point
Primal-Dual algorithm

4

4This section is based on Stephen P Boyd and Lieven Vandenberghe. Convex optimization. Cambridge university press, 2004
Definition

Problem is called **convex** if it is in the form:

\[
\begin{align*}
\min \ & f_0(x) \\
\text{s.t.} \ & f_i(x) \leq 0 \quad \forall i = 1, \ldots, m \\
\ & Ax = b
\end{align*}
\]  

(9) \quad (10) \quad (11)

where \( \forall i \ f_i : \mathbb{R}^n \to \mathbb{R} \) is convex and \( A \in \mathbb{R}^{p \times n} \) with \( \text{rank}(A) = p < n \).
Definition

Problem is called **convex** if it is in the form:

\[
\begin{align*}
\text{min} & \quad f_0(x) \\
\text{s.t.} & \quad f_i(x) \leq 0 \quad \forall i = 1, \ldots, m \\
& \quad Ax = b
\end{align*}
\]

where \( \forall i \ f_i : \mathbb{R}^n \to \mathbb{R} \) is convex and \( A \in \mathbb{R}^{p \times n} \) with \( \text{rank}(A) = p < n \).

Lagrangian

Define a function \( L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R} \), called **Lagrangian**, as

\[
L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \nu^T (Ax - b)
\]  

\( \lambda, \nu \) are called **dual variables**
Duality

Dual function

Define *Lagrangian dual function* \( g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R} \) as a minimum value of \( L(\cdot, \lambda, \nu) \) over \( x \):

\[
g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) \tag{13}
\]
Duality

**Dual function**

Define *Lagrangian dual function* $g : \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$ as a minimum value of $L(\cdot, \lambda, \nu)$ over $x$:

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) \quad (13)$$

Note that $g(\lambda, \nu)$ is concave regardless of $L$. 
Duality

**Dual function**

Define *Lagrangian dual function* $g : \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$ as a minimum value of $L(\cdot, \lambda, \nu)$ over $x$:

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) \quad (13)$$

Note that $g(\lambda, \nu)$ is concave regardless of $L$.

**Dual problem**

*Lagrange dual* problem is defined as

$$\max \ g(\lambda, \nu) \quad (14)$$

$$\text{s.t. } \lambda_i \geq 0 \quad \forall i = 1, \ldots, m \quad (15)$$
KKT conditions

Duality gap

*Duality gap* is the difference between primal and dual opt. values.
KKT conditions

Duality gap

*Duality gap* is the difference between primal and dual opt. values.

KKT conditions

Let \((x^*, \lambda^*, \nu^*)\) be points with zero duality gap, then the following conditions have to hold:

\[
\begin{align*}
  f_i(x^*) &\leq 0 & \forall i = 1, \ldots, m \\
  Ax - b & = 0 \\
  \lambda_i^* &\geq 0 & \forall i = 1, \ldots, m \\
  \lambda_i^* f_i(x^*) & = 0 & \forall i = 1, \ldots, m \\
  \nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^{p} \nu A_i & = 0
\end{align*}
\]
Relaxed problem

Consider relaxed version of the primal problem:

\[
\min \ f_0(x) + \sum_{i=1}^{m} -\frac{1}{t} \log(-f_i(x))
\]

s.t. \ Ax = b

(16)

(17)
Relaxed problem

Consider relaxed version of the primal problem:

\[
\begin{align*}
\min & \quad f_0(x) + \sum_{i=1}^{m} -\frac{1}{t} \log(-f_i(x)) \\ 
\text{s.t.} & \quad Ax = b
\end{align*}
\]  

(16)  

KKT conditions

Corresponding KKT conditions are:

\[
\begin{align*}
Ax - b &= 0 \\
\lambda_i^* f_i(x^*) &= \frac{1}{t} \quad \forall i = 1, \ldots, m \\
\nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^{p} \nu A_i &= 0
\end{align*}
\]
Modified KKT conditions

Compact form

KKT conditions can be compactly written as \( r_t(x, \lambda, \nu) = 0 \) where

\[
\begin{bmatrix}
\nabla f_0(x) + Df(x)^T \lambda + A^T \nu \\
-\text{diag}(\lambda)f(x) - \left(\frac{1}{t}\right)1 \\
Ax - b
\end{bmatrix}
= \begin{bmatrix}
r_{\text{dual}} \\
r_{\text{cent}} \\
r_{\text{pri}}
\end{bmatrix}
\]

Here we used the following notation:

\[
f(x) = \begin{bmatrix}
f_1(x) \\
\vdots \\
f_m(x)
\end{bmatrix} \quad Df(x) = \begin{bmatrix}
\nabla f_1(x)^T \\
\vdots \\
\nabla f_m(x)^T
\end{bmatrix}
\]
**Newton step**

Set $y = (x, \lambda, \nu)$, $\Delta y = (\Delta x, \Delta \lambda, \Delta \nu)$ calculate newton step from the approximation:

$$ r_t(y + \Delta y) \approx r_t(y) + D r_t(y) \Delta y = 0 \quad (18) $$

which, if expanded, has the form:

$$
\begin{bmatrix}
\nabla^2 f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla^2 f_i(x) & Df(x)^T & A^T \\
-\text{diag}(\lambda) Df(x) & -\text{diag}(f(x)) & 0 \\
A & 0 & 0
\end{bmatrix} \Delta y = -
\begin{bmatrix}
\hat{\eta}(x, \lambda) \\
0 \\
0
\end{bmatrix}
$$

Duality gap

Surrogate duality gap is given by

$$\hat{\eta}(x, \lambda) = -f(x)^T \lambda \quad (19)$$
Newton step

Set $y = (x, \lambda, \nu)$, $\Delta y = (\Delta x, \Delta \lambda, \Delta \nu)$ calculate newton step from the approximation:

$$r_t(y + \Delta y) \approx r_t(y) + Dr_t(y)\Delta y = 0 \quad (18)$$

which, if expanded, has the form:

$$
\begin{bmatrix}
\nabla^2 f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla^2 f_i(x) & Df(x)^T & A^T \\
-\text{diag}(\lambda)Df(x) & -\text{diag}(f(x)) & 0 \\
A & 0 & 0
\end{bmatrix}
\begin{bmatrix}
\Delta y \\
r_{\text{dual}} \\
r_{\text{cent}} \\
r_{\text{pri}}
\end{bmatrix}
= 0
$$

Duality gap

*Surrogate duality gap* is given by

$$\hat{\eta}(x, \lambda) = -f(x)^T \lambda \quad (19)$$
Algorithm

**Algorithm 11.2** Primal-dual interior-point method.

given $x$ that satisfies $f_1(x) < 0, \ldots, f_m(x) < 0$, $\lambda > 0$, $\mu > 1$, $\epsilon_{\text{feas}} > 0$, $\epsilon > 0$.

repeat

1. Determine $t$. Set $t := \mu m / \hat{\eta}$.
2. Compute primal-dual search direction $\Delta y_{pd}$.
3. Line search and update.
   Determine step length $s > 0$ and set $y := y + s \Delta y_{pd}$.

until $\|r_{\text{pri}}\|_2 \leq \epsilon_{\text{feas}}$, $\|r_{\text{dual}}\|_2 \leq \epsilon_{\text{feas}}$, and $\hat{\eta} \leq \epsilon$.

---

**Figure**: Primal-Dual interior point algorithm as given in

---

Results: synthetic data
Transformation

Original

Transformed
Feature matching

Feature matches between images
Reconstruction
Can we do better?
Reconstruction + sharpening

Original

Reconstructed and sharpened
Results: real data
Image stitching

Problem setup

Given a set of images of the same scene from different views reconstruct or "stitch" images together to represent combined view.
Image stitching

Problem setup

Given a set of images of the same scene from different views reconstruct or "stitch" images together to represent combined view.

Figure: How can we stitch these two images together?
Feature matching

Feature extraction and matching

Find correspondences between two images
Image Registration using Primal-Dual interior point method

## Results

Application: image stitching

---

Interior point Primal-Dual algorithm with alpha= 0.05, beta= 0.45.

<table>
<thead>
<tr>
<th>iter</th>
<th>dual gap</th>
<th>r_cent</th>
<th>r_dual</th>
<th>objective</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>60.49092705</td>
<td>9.27533272</td>
<td>475827.918031</td>
<td>204723.54837804</td>
</tr>
<tr>
<td>2</td>
<td>46.70868110</td>
<td>7.46590846</td>
<td>475662.05723931</td>
<td>147912.38735533</td>
</tr>
<tr>
<td>3</td>
<td>39.72212290</td>
<td>6.71354506</td>
<td>475473.45542507</td>
<td>127116.98248039</td>
</tr>
<tr>
<td>4</td>
<td>37.23398139</td>
<td>6.47330260</td>
<td>475302.65306721</td>
<td>119681.16128237</td>
</tr>
<tr>
<td>5</td>
<td>35.94529866</td>
<td>6.33208593</td>
<td>475136.81633145</td>
<td>115807.45439618</td>
</tr>
<tr>
<td>6</td>
<td>35.16571689</td>
<td>6.24528203</td>
<td>474973.17045421</td>
<td>113443.51991650</td>
</tr>
<tr>
<td>7</td>
<td>34.27697827</td>
<td>6.13473575</td>
<td>474751.63933781</td>
<td>111319.65494739</td>
</tr>
<tr>
<td>8</td>
<td>33.27692781</td>
<td>6.02471684</td>
<td>474460.48959382</td>
<td>109583.94164635</td>
</tr>
<tr>
<td>9</td>
<td>32.19833192</td>
<td>5.89888078</td>
<td>474055.46902172</td>
<td>108063.89932026</td>
</tr>
<tr>
<td>10</td>
<td>30.09281949</td>
<td>5.66326842</td>
<td>473287.42717818</td>
<td>106252.25945835</td>
</tr>
</tbody>
</table>

... 

| 101  | 0.00017343 | 0.00003597 | 94.98498993 | 562.68029677 |
| 102  | 0.00010881 | 0.00002203 | 72.14802702 | 550.74006338 |
| 103  | 0.00001503 | 0.00000229 | 47.51558755 | 539.78197274 |
| 104  | 0.00001229 | 0.00000255 | 38.17590650 | 539.46951871 |
| 105  | 0.00001183 | 0.00000248 | 36.63126699 | 539.42341193 |
| 106  | 0.00001162 | 0.00000244 | 35.96304578 | 539.40396942 |
| 107  | 0.00001154 | 0.00000242 | 35.66759301 | 539.39547101 |
| 108  | 0.00001135 | 0.00000238 | 35.01692864 | 539.37696590 |
| 109  | 0.00001042 | 0.00000217 | 31.87955537 | 539.29183885 |
| 110  | 0.00000139 | 0.00000113 | 2.42853146 | 538.85996497 |
| 111  | 0.00000083 | 0.00000014 | 1.34913410 | 538.85825022 |
| 112  | 0.00000009 | 0.00000000 | 0.01734665 | 538.85748058 |
| 113  | 0.00000001 | 0.00000000 | 0.00017412 | 538.85748041 |
| 114  | 0.00000000 | 0.00000000 | 0.00000174 | 538.85748041 |
| 115  | 0.00000000 | 0.00000000 | 0.00000002 | 538.85748041 |

Optimal value found 538.85748
Reconstruction

Stitched image
The End
Results

Application: image stitching


