# Efficient Coding of Image Sequences of Planar Scenes 

## Udi Shrot

Supervisors: Prof. David Malah,

Dr. Meir Bar-Zohar

## Outline

- Image Sequence coding - Problem statement.
- Image Registration.
- Various camera models and the derived spatial transforms.
- Kalman Filtering as a spatial transform estimator.
- Proposed estimator for the perspective planar transform.
- Application of the proposed estimator to improve camera motion parameter estimation.


## Image Sequence Coding

- The memory and bandwidth requirements of Image Sequence representation led to various coding methods.
- The basic idea - exploiting the information redundancy between adjacent frames - only the innovations among frames are coded.
- The General Methods such as the standards H.261,H. 263 and Mpeg's use Block Matching methods to create the difference frame.



## Image Sequence Coding (cont'd)

- There are image sequences which can be described by a single set of global motion parameters.
- For example, an image sequence taken by a moving camera.
- For such cases efficient coding is possible.
- The sequence is represented by a reference Image and the global motion parameters only.
- This work proposes a robust method for estimating the motion parameters of planar scenes.


## Image Registration

Image Registration, in the general case, means matching two images. The images can be taken from two different points of
view, at different times with different sensors under different illumination conditions.

## Image Registration - Applications

- Medical Applications - identifying abnormal differences between images taken at different times.
- Data Compression - Registration enables transmission only the differences between consecutive Images with a reduction of the transmitted bit-rate.
- Image understanding - Combining Different sensors can be very helpful for image understanding.
- Shape and depth reconstruction such as stereo.


## This Work's Registration scope

- Given two images $I_{1}(X, Y)$ and $I_{2}(X, Y)$ Find the best global spatial transform $\left(X^{\prime}, Y^{\prime}\right)=T(X, Y)$ so that

$$
I_{2}(X, Y) \approx I_{1}(T(X, Y))
$$



$$
T(X, Y)=?
$$



## Spatial Transform Estimation <br> Approaches

## I. Optical Flow.

- Estimation of Image Velocity field (for every pixel).
- These methods are usually based upon the assumptions:

1. Inter-frame motion is small.
2. Intensity function is smooth.

- These assumptions lead to the optical flow constraint equation

$$
\nabla_{t} I(\vec{x}, t)+\nabla_{\vec{x}} I(\vec{x}, t) \cdot \frac{\partial \vec{x}}{\partial t}=0
$$

- Extra smoothness constraints are usually needed
- Block matching is a variation of optical flow estimation.


## Spatial Transform Estimation Approaches (cont'd)

## II. Feature Point Matching.

- Extraction and matching of distinct tokens in the two images.
- These tokens are usually points of large intensity variations
- Feature points are more robust to varying image intensities.
- Only sparse motion field can be derived.
- No smoothness constraints are used.
- This method usually perform better than optical flow methods for global transformed images.


## Feature Point Extraction <br> Example



## Comparison between Estimation Approaches

- Performance comparison between block matching and global motion estimation method using feature points extraction.
- The graph presents the rmse between each frame and the registered previous frame.

Image sequence



## Feature Points Extraction

## Possible feature points extraction methods:

- Corners.
- Edges.
- Contours.

Method I
The edge detector's output and the chosen best points
Horizontal gradient filter

Vertical gradient filter
$45^{\circ}$ directional gradient filter
${ }^{135^{\circ}}$ directional gradient filter


## Feature Points Extraction(contd)

 [Shi \& Tomasi 94]
## Method II

I. Calculate the directional image derivatives ${ }_{I_{x}}, I_{y}$
II. For every pixel calculate the eigen-values of the matrix

$$
\lambda_{1,2}=\operatorname{eig}\left(\left(\begin{array}{ll}
\sum_{x}^{1} I_{x}^{2} & \sum_{A} I_{x} I_{y} \\
\sum_{A} I_{x} I_{y} & \sum_{A} I_{v}^{2}
\end{array}\right)\right.
$$

for a small neighborhood $A$ around it
III. Choose the pixel as a feature point if
A. $\min \left(\lambda_{1}, \lambda_{2}\right)>t o l_{1}$
B. $1 \leq \max \left(\lambda_{1}, \lambda_{2}\right) / \min \left(\lambda_{1}, \lambda_{2}\right)<t o l_{2}$

## Comparison of Feature Points Extraction Methods

- Method I - corner detection.
- Method II - eigen-value method

The graph presents the rmse between each frame and the registered reference frame.


## Camera and Motion Models

 I. Orthographic Camera Model- The depth changes in the scene are negligible relative to the camera distance from the scene $|d z| \ll|z|$

- The General Motion of a 3D object in an Orthographic Model results (on the image plane) in an affine transformation.

$$
\begin{aligned}
& X^{\prime}=a X+b Y+c \\
& Y^{\prime}=d X+e Y+f
\end{aligned}
$$

## Affine Transformation

## Definition:

$$
\left(\begin{array}{ll}
a & b \\
d & e
\end{array}\right)\binom{X}{Y}+\binom{c}{f}=\left(\begin{array}{cc}
\cos \theta_{1} & -\sin \theta_{1} \\
\sin \theta_{1} & \cos \theta_{1}
\end{array}\right)\left(\begin{array}{ll}
s_{x} & 0 \\
0 & s_{y}
\end{array}\right)\left(\begin{array}{cc}
\cos \theta_{2} & -\sin \theta_{2} \\
\sin \theta_{2} & \cos \theta_{2}
\end{array}\right)\binom{X}{Y}+\binom{c}{f}
$$

Particular cases:
original:

expansion

## shear

rotation


## Camera and Motion Models (Cont'd)

## II. Perspective Pin-hole Camera Model

- The farther the object from the camera, the smaller its size on the image plane.

- General motion of a 3D object according to this model results, on the image plane, in the perspective transformation:

$$
\begin{aligned}
X^{\prime} & =\frac{(a X+b Y+c) z+d x}{(p X+q Y+r) z+d z} \\
Y^{\prime} & =\frac{(d X+e Y+f) z+d y}{(p X+q Y+r) z+d z}
\end{aligned}
$$

## Camera and Motion Models (Cont'd)

## Perspective-Planar Transformation

- The general perspective transformation depends upon the pixel's distance from the camera plane.
- The perspective-planar transformation

$$
\begin{aligned}
X^{\prime} & =\frac{a X+b Y+c}{p X+q Y+1} \\
Y^{\prime} & =\frac{d X+e Y+f}{p X+q Y+1}
\end{aligned}
$$

Describes two specific motions:

- Camera confined to rotation only.
- General moving camera taking images of a single plane scene.


## Spatial Transforms - Examples

- Two images related by an affine transformation

Original image Affine transformed image


- Two images related by perspective-planar transformation Original image Perspective-planar transformed image



## Image Sequence Representation by a Reference Image and Motion Parameters

- Extraction of Motion Parameters between each frame and the reference frame:
Method I
calculate motion parameters between any two adjacent frames and combine these differential motion corrections.



## Image Sequence Representation by a Reference Image and Motion Parameters (cont'd)

## Method II

1. Calculate motion transform $\Delta T$ between frame $\operatorname{Im}_{N}$ and the predicted previous frame $\hat{\operatorname{Im}}_{N-1}$ (created by warping the reference image by $T$ )
2. Combine this differential motion estimation with $T$ so that $T \leftarrow T \oplus \Delta T$

3. Increment $N$ and go to the first step.

## Performance Comparison of Global Transformation Estimators

I. Incremental method
II. Warped-ref method

The graph presents the rmse between each frame and the registered reference frame - affine transformation estimation


## Performance Comparison of Estimation Methods

The comparison is performed by averaging the motion compensated images (calculated by the two methods)

Incremental method
Averaged motion compensated image by incremental transform estimation


Warped-ref method
Averaged motion compensated image by current-ref transform estimation

## Least Squares Methods for Parameters Estimation

## Assumptions:

- The measurements are contaminated with noise.
- Number of measurements exceeds the number of unknown parameters.
Least Squares:
The accurate equation is

$$
\overline{\bar{A}}_{0} \vec{x}=\vec{b}_{0}
$$

The measurements are

$$
\vec{b}=\vec{b}_{0}+\vec{n}
$$

if $\overline{\bar{A}}_{0}$ is known exactly than the LS solution is

$$
\hat{x}_{L S}=\left(\overline{\bar{A}}_{0}^{T} \overline{\bar{A}_{0}}\right)^{-1} \overline{\bar{A}}_{0}^{T} \vec{b}
$$

## Parameters Estimation (cont'd)

## Total Least Squares:

The accurate equation is

$$
\overline{\bar{A}}_{0} \vec{x}=\vec{b}_{0}
$$

The measurements are

$$
\vec{b}=\vec{b}_{0}+\vec{n}_{1} \quad \overline{\bar{A}}=\overline{\bar{A}}_{0}+\vec{n}_{2}
$$

The TLS solution is

$$
\begin{gathered}
{[\overline{\bar{A}} \vec{b}]=U \cdot \operatorname{diag}\left(\lambda_{1}, \ldots, \lambda_{N}\right) \cdot V^{T}, \quad \lambda_{1}>\lambda_{2}>. .>\lambda_{N}(\text { SVD decomposition })} \\
\hat{x}_{T L S}=-\frac{V_{(1: N-1, N)}}{V_{(N, N)}}
\end{gathered}
$$

## Estimation of Motion Model Parameters

Affine Transformation Model:

$$
\begin{aligned}
& X^{\prime}=a X+b Y+c \\
& Y^{\prime}=d X+e Y+f
\end{aligned}
$$

- Finding set of matched feature points in the two images

$$
\left(X_{i}, Y_{i}\right) \rightarrow\left(X_{i}^{\prime}, Y_{i}^{\prime}\right) \quad i=1, . ., N
$$

- The matched points $\left(X_{i}{ }^{\prime}, Y_{i}{ }^{\prime}\right)$ are contaminated with noise so a LS solution method is used:

$$
\left.\left.b=A x \Leftrightarrow\left(\begin{array}{l}
X_{1} 1_{1} \\
Y_{1} \\
\vdots \\
X_{1}^{\prime} \\
Y_{N}
\end{array}\right)=\left(\begin{array}{cccccc}
X_{N}
\end{array}\right)=\left(\begin{array}{cccccc}
a & Y_{1} & 1 & 0 & 0 & 0 \\
\vdots & 0 & 0 & X_{1} & Y_{1} & 1 \\
\vdots & & & 1 & & \\
X_{N_{N}} & Y_{N} & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & X_{N} & Y_{N} & 1
\end{array}\right) \right\rvert\, \begin{array}{c}
d x \\
d x \\
d \\
d \\
d y
\end{array}\right) \Rightarrow \hat{x}=\left(A^{T} A\right)^{-1} A^{T} b
$$

## Linear Kalman Filter

## Problem Statement

- Estimate the state vector $x_{k}$ based upon the measurements $Z_{k}=\left\{z_{0}, z_{1}, \ldots, z_{k}\right\}$ using the following assumptions:
The state equation:

$$
x_{k+1}=F_{k} \cdot x_{k}+w_{k}+u_{k}
$$

The measurement equation:

$$
z_{k+1}=C_{k} \cdot x_{k+1}+\varepsilon_{k}
$$

where,
$w_{k} \sim N\left(0, Q_{k}\right), \varepsilon_{k} \sim N\left(0, R_{k}\right), u_{k}$ is a known input.
$w_{k}, \varepsilon_{k}, x_{k}$ are all independent of each other.
$x_{k+1 \mid k+1}$ denoted the best linear estimator for $x_{k+1}$ based upon
$Z_{k+1}=\left\{z_{0}, z_{1}, \ldots, z_{k}, z_{k+1}\right\}$

## Linear Kalman Filter (contd)

- The estimator is:
prediction
innovation
Error covariance matrix of prediction

estimation
Error covariance matrix of estimation


## Kalman Filter smoothing of Affine Transform Parameters Estimation

- State vector: $x_{k}=\left(a_{k} b_{k} c_{k} d_{k} e_{k} f_{k}\right)^{T}$
- State equation: $x_{k+1}=I_{6 x 6} x_{k}+w_{k} w_{k} \sim N\left(0, Q_{k}\right)$
- Measurement equation:

$$
z_{k}=\left[\begin{array}{l}
X_{1}^{\prime} \\
Y_{1}^{\prime} \\
\vdots \\
X_{N}{ }^{\prime} \\
Y_{N}{ }^{\prime}
\end{array}\right]_{k}=\left[\begin{array}{cccccc}
X_{1} & Y_{1} & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & X_{1} & Y_{1} & 1 \\
\vdots & & & & & \\
X_{N} & Y_{N} & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & X_{N} & Y_{N} & 1
\end{array}\right]_{k} \boldsymbol{x}_{\boldsymbol{k}}+\boldsymbol{\varepsilon}_{\boldsymbol{k}}
$$

$x_{0}$ is calculated by LS solution of matched points.

## Performance of the Kalman Filter Smoothing Method

This graph displays the average distance between the estimated tracking points and their theoretical location for the synthetic rotated salesman sequence.


## Kalman Filter Smoothing of Feature Points

The affine transform estimated in the previous frame is used to improve the matching of feature points.
-State vector: $\left[\begin{array}{l}X \\ Y\end{array}\right]_{k}$
-State equation:

$$
\left[\begin{array}{l}
X \\
Y
\end{array}\right]_{k+1}=\left[\begin{array}{ll}
a_{k} & b_{k} \\
d_{k} & e_{k}
\end{array}\right]\left[\begin{array}{l}
X \\
Y
\end{array}\right]_{k}+\left[\begin{array}{l}
x_{k} \\
f_{k}
\end{array}\right]+w_{k}
$$

-Measurement equation:

$$
\left[\begin{array}{c}
X^{\prime} \\
Y^{\prime}
\end{array}\right]_{k}=\left[\begin{array}{ll}
1 & 0 \\
0 & 1
\end{array}\right]\left[\begin{array}{l}
X \\
Y
\end{array}\right]_{k+1}+\varepsilon_{k}
$$

The matched tracking points are used to estimate the affine transformation using Least Squares.

## Performance of the Kalman Filter Smoothing Method

This graph displays the average distance between the estimated tracking points and their theoretical location for the synthetic rotated 'Salesman’ sequence.


## Estimation of the Perspective-Planar <br> Transformation Parameters

$$
\begin{aligned}
X^{\prime} & =\frac{a X+b Y+c}{p X+q Y+1} \\
Y^{\prime} & =\frac{d X+e Y+f}{p X+q Y+1}
\end{aligned}
$$

- Finding matched feature points in the two images

$$
\left(X_{i}, Y_{i}\right) \rightarrow\left(X_{i}^{\prime}, Y_{i}^{\prime}\right) \quad i=1, . ., N
$$

- Algebraic manipulations results in:

$$
\left(\begin{array}{l}
X^{\prime}{ }_{1} \\
Y^{\prime}{ }_{1} \\
\vdots \\
X^{\prime}{ }_{N} \\
Y^{\prime}{ }_{N}
\end{array}\right)=\left(\begin{array}{ccccccc}
X_{1} & Y_{1} & 1 & 0 & 0 & 0 & -X_{1} X_{1}{ }^{\prime}-Y_{1} X_{1}{ }^{\prime} \\
0 & 0 & 0 & X_{1} & Y_{1} & 1 & -X_{1} Y_{1}{ }^{\prime}-Y_{1} Y_{1}{ }^{\prime} \\
\vdots & & & & & & \\
X_{N} & Y_{N} & 1 & 0 & 0 & 0 & -X_{N} X_{N}{ }^{\prime}-Y_{N} X_{N}{ }^{\prime} \\
0 & 0 & 0 & X_{N} & Y_{N} & 1 & -X_{N} Y_{N}{ }^{\prime}-Y_{N} Y_{N}{ }^{\prime}
\end{array}\right)\left(\begin{array}{l}
b \\
c \\
d \\
e \\
f \\
p \\
q
\end{array}\right)
$$

- These equation set suffers from numerical instabilities


## Second Order Motion Model

- An approximation for the perspective-planar transformation is the second order motion model:

$$
\begin{aligned}
& X^{\prime}=a X+b Y+c+h X^{2}+g X Y \\
& Y^{\prime}=d X+e Y+f+h X Y+g Y^{2}
\end{aligned}
$$

- Empirical results proved that the second order approximated model estimation is more robust than the exact perspective-planar model estimation
- The approximated model is accurate for image pairs which are related by sufficient small motion.


## Second Order Motion Model - Simulation Results

- 40 points were transformed by several planar perspective transforms with constant denominator.
- Uniform noise was added to these points and several estimation methods were tested.

Average distance between true and estimated points
$\mathrm{p}=0.0001 \mathrm{q}=-0.0002, \mathrm{r}=1$


$$
\mathrm{p}=0.0005 \mathrm{q}=-0.0005, \mathrm{r}=1
$$



The red line is the second order model. The green line is the TLS method

## Estimation of Planar Scene Motion Parameters (cont'd)

- Adjacent frames are related by sufficient small motion.


## Proposed Method

a. Estimation of second-order model between current frame \& the previously predicted frame (warping of the reference image by $T_{\text {pesesp }}$ ).

b. Transforming several point coordinates (X,Y) by $T_{\text {persp }}$ and than by

$$
\Delta T_{2 n d} d_{\text {order }}
$$

$$
\left(X^{\prime}, Y^{\prime}\right)=\Delta T_{\text {2nd _order }}\left(T_{\text {pesesp }}(X, Y)\right)
$$

c. Estimation of new $T_{\text {persp }}$ using LS based upon (X,Y) \& ( $\left.\mathrm{X}^{\prime}, \mathrm{Y}^{\prime}\right)$

## Estimation of Planar Scene Motion Parameters (contd)

## The Warped-ref Method

a. Estimation of perspective-planar transform between current frame \& the previously predicted frame (warping of the reference image by $T$ )
b. Combining $T$ \& the new transform by ordinary transform cascading to get new $T=T+\Delta T_{\text {persp }}$


The proposed method performs better than the Warped-ref method as shown in the following slides.

## Results of Proposed Algorithm

'Salesman' image warped by a fixed perspective transform ( 15 frames)

Original frame


Fifteenth frame


Averaging the fifteen motion compensated images

Proposed method


Warped-ref method


## Results of Proposed Algorithm (conta)

Performance comparison between global transform estimators:
I. Proposed method
II. Warped-ref method

The graph presents the rmse between each frame and the registered reference frame.


## Estimation of Camera Motion Parameters

The Camera motion model includes a 3-D rotation matrix and a 3-D translation vector:

$$
\left(\begin{array}{l}
x^{\prime} \\
y^{\prime} \\
z^{\prime}
\end{array}\right)=R\left(\begin{array}{l}
x \\
y \\
z
\end{array}\right)+T
$$

Existing algorithm [Huang \& Tsai 1984]:
SVD decomposition of perspective-planar transform:

$$
\begin{aligned}
& {\left[\begin{array}{ccc}
a & b & c \\
d & e & f \\
p & q & 1
\end{array}\right]=U \cdot \operatorname{diag}\left(\lambda_{1}, \lambda_{2}, \lambda_{3}\right) \cdot V^{T}} \\
& R=U \cdot\left[\begin{array}{ccc}
\alpha & 0 & \beta_{1,2} \\
0 & 1 & 0 \\
-s & \beta_{1,2} & 0 \\
s
\end{array}\right] \cdot V^{T} \quad \alpha, \beta_{1,2}=f\left(\lambda_{1,2,3}\right), \quad s=\operatorname{det}(U) \cdot \operatorname{det}(V) \\
& T=k\left(-\beta_{1,2} u_{1}+\left(\frac{\lambda_{3}}{\lambda_{2}}-s \alpha\right) u_{3}\right) k \text { is a scaling cons } \tan t
\end{aligned}
$$

## Estimation of Camera Motion Parameters (cont'd)

The rotation matrix can be described by rotation with angle $\alpha$ around center of rotation ( $n_{1}, n_{2}, n_{3}$ )
$R=\left(\begin{array}{lcc}n_{1}^{2}+\left(1-n_{1}^{2}\right) \cos \alpha & n_{1} n_{2}(1-\cos \alpha)-n_{3} \sin \alpha & n_{1} n_{3}(1-\cos \alpha)+n_{2} \sin \alpha \\ n_{1} n_{2}(1-\cos \alpha)+n_{3} \sin \alpha & n_{2}^{2}+\left(1-n_{2}^{2}\right) \cos \alpha & n_{2} n_{3}(1-\cos \alpha)-n_{1} \sin \alpha \\ n_{1} n_{3}(1-\cos \alpha)-n_{2} \sin \alpha & n_{2} n_{3}(1-\cos \alpha)+n_{1} \sin \alpha & n_{3}^{2}+\left(1-n_{3}^{2}\right) \cos \alpha\end{array}\right)$
Example of estimation of the rotation angle of a synthetic image sequence. The theoretical angle is 1.5 deo ner frame



## Conclusions

- Certain Image-Sequences can be described by a reference image and global motion parameters.
- For planar scenes the proposed method performs better than conventional methods.
- Accurate estimation of the global motion parameters is useful for camera motion estimation.


## Future Work

- Implementation of a coding system using this work results.
- Motion segmentation so that multiple plane scenes can be efficiently encoded.
- Using side information to improve coding capabilities.

