Vision-based control for an Autonomous Wheelchair

Humberto Sermeno & John Spletzer

Motivation

- Automobile operators with lower body disabilities have limited options with regards to unattended personal mobility.
- Traditional solution is a custom van conversion that places the operator *in-wheelchair* behind the steering wheel of the vehicle.

Safety Shortcomings

- Wheelchairs do not possess similar levels of crash protection afforded by traditional motor vehicle seat systems
- Provisions used for securing the wheelchair are often inadequate.
- Research in the US showed that
    - 35% of all wheelchair/automobile related deaths resulted from inadequate chair securement
    - Another 19% associated with vehicle lift malfunctions

Automated Transport and Retrieval System (ATRS)

- Wheelchair autonomously navigates between the driver's position and a powered lift at the rear of the vehicle
- Benefits
    - Safety
        - Operator and chair are separated during vehicle operations
        - The operator is seated in a crash tested seat system
    - Cost
        - It is expected that the cost of integrating ATRS will be about half that of a van conversion
        - Modifications are non-permanent, so they will not affect the vehicle resale value

Problem Statement

- *"Develop a means for reliable, autonomous docking (and undocking) of the ATRS wheelchair onto the vehicle's lift platform*."
- Initially, the problem was attacked using the onboard laser rage finder and odometry.
- Unfortunately, high noise levels from ambient sunlight resulted in localization errors that yielded inconsistent docking performance

Two-phase approach

- Navigation from the operator position to the vicinity of the lift platform is accomplished using laser based feedback control with sensors onboard the chair itself.
- In contrast, docking and undocking the chair on the lift platform relies upon a computer vision system on the minivan to estimate chair pose.

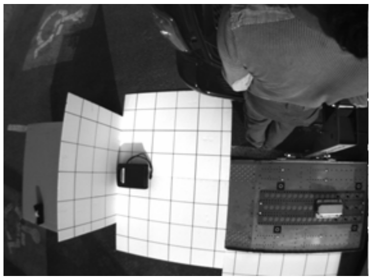
Machine Vision System

- Why Use a Camera?
  - Natural interface for operators
    - Verify the platform area is clear
    - Provides tele-operational mode for system recovery
  - Compact form factor
  - Docking Tolerances:
    - Lift/wheelchair combinations provide 2.5-5.0 cm of clearance on each side of the chair
    - The laser system can typically provide position estimates with an accuracy of 3-5 cm
    - We can do *much* better with a camera system

Vision System Characteristics

- Requirements
  - Large field of view (FOV) lens to maximize camera coverage
  - Requires high-resolution digital camera for precise localization requirements Issues
- Issues:
  - Large amounts of image distortion
  - Enormous amounts of image data (11.25 MB/s) that must be processed for real-time control
  - Must be able to operate under a wide range of illumination Levels

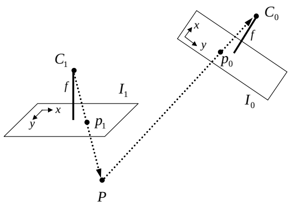
Correct for Image Distortion in Software



Simplifying the problem

- Use binary patterns (fiducials) on each armrest
- Assumptions
  - Wheelchair motion is constrained to the ground plane which is locally flat
  - The position of the lift platform with respect to the van is fixed
  - The position and orientation of the camera with respect to the lift platform is fixed
- Problem is reduced to a two-dimensional pattern matching task.

Virtual Camera



- Project p_1 to a world point P through the camera model
- Project P to camera coordinates in the true camera frame
  - $P_c = R_0^{-1}(P - C_0)$
- Prior to projecting P_C to the plane, we must account for significant image distortion

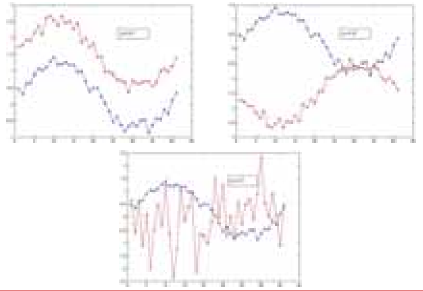
Pattern Recognition

- Remember the correlation coefficient?

$$\rho = \frac{\sum_i \left[(x_i - \mu_x)(y_i - \mu_y)\right]}{\sqrt{\sum_i (x_i - \mu_x)^2}\sqrt{\sum_i (y_i - \mu_y)^2}} \equiv \frac{C_{xy}}{\sqrt{C_{xx}}\sqrt{C_{yy}}} \equiv \frac{C_{xy}}{\sigma_x \sigma_y}$$

- The denominator normalizes the correlation coefficient to [-1,1]

Cross-Correlation Examples

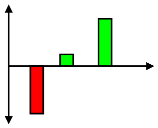


Application: 2D Pattern Recognition

- ☐ Let's assume we are given a pattern template of interest (a target, face, etc.) and must locate it within the video image
- ☐ This template can be reduced to a series of numbers corresponding to its pixel values
- ☐ We can compare these pixel values to similar size regions in the video image
- ☐ The video image position of *maximum* correlation to the template corresponds to the location of our pattern in the video image

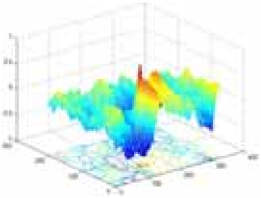
2D Pattern Recognition Example

Template



Search Region

2D Pattern Recognition Example



Actual Template

Max Correlation Template

Pattern Recognition - NID

- ☐ Normalized Intensity Distribution (NID)

$$\epsilon(T,B)=\sum_{u=1}^{m}\sum_{v=1}^{n}\left[\frac{T(u,v)-\mu_T}{\sigma_T}-\frac{B(u,v)-\mu_B}{\sigma_B}\right]^2$$

- ☐ NID has the advantage of explicitly modeling both changes in scene brightness and contrast from the reference template image
- ☐ For a binary template,

$$\epsilon(T,B)=\begin{cases}0 & g_w > g_b\\ 4N & g_w < g_b\end{cases}$$

Optimizing the NID metric

- ☐ An aside on performance…

$$\epsilon(T,B)=\sum_{u=1}^{m}\sum_{v=1}^{n}\left[\frac{T(u,v)-\mu_T}{\sigma_T}-\frac{B(u,v)-\mu_B}{\sigma_B}\right]^2$$

- ☐ After expanding:

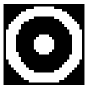
$$\epsilon(T,B)=2mn-\frac{2}{\sigma_T\sigma_B}\left[mn\mu_T\mu_B-\sum_{u=1}^{m}\sum_{v=1}^{n}T(u,v)\cdot B(u,v)\right]$$

- ☐ For a given template:

$$\epsilon_T(B)=\frac{1}{\sigma_B}\left[K\cdot\mu_B-\sum_{u=1}^{m}\sum_{v=1}^{n}T(u,v)\cdot B(u,v)\right]$$

Fiducial/Template Design

- ☐ Some nice properties
  - Invariant to rotation from an orthographic perspective
  - Binary representation exhibits a strong contrast which facilitates tracking under varying illumination conditions

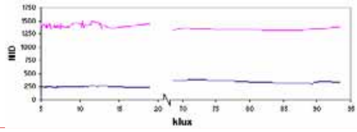


Fiducial/Template Design

- ☐ However, strong changes in illumination lead to significant changes to fiducial appearance

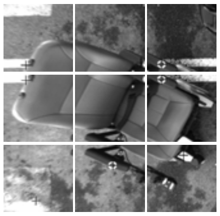
Fiducial/Template Design

- ☐ Varied grayscale value for white
- ☐ Template geometry was generated as a composite of the perceived geometry of the fiducials under light levels ranging from 5 to 95 kLux



Initial Localization

- ☐ A search window was tessellated into nine overlapping regions

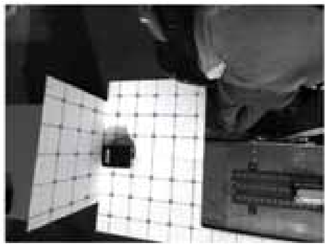


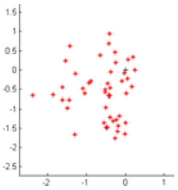
- ☐ Binary filters:
  - Distance between the fiducials in world coordinates within a predefined tolerance
  - Distance between the fiducials in pixel coordinates greater than the size of the fiducial
- ☐ Select the pair with highest combined NID values

Subsequent localization steps

- ☐ Look for the fiducial in a 144x144 pixel area centered around their last known positions
- ☐ Could be accomplished at frame rate (15 Hz)
- ☐ Allowed for a maximum linear velocity of approximately 2.5 m/s
- ☐ In practice, wheelchair velocity was limited to 0.5 m/s

How Accurate Can We Track?





Feedback control

- ☐ Control signals sent over a dedicated wireless link
- ☐ Forward velocity is fixed, while angular velocity is controlled in real-time
- ☐ A PD controller

$$\omega = -k_v \tan\theta - \frac{k_p y}{v \cos\theta}$$

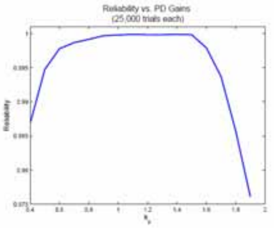
  - Rapidly converges to the desired position
  - Linear approximation, but works well in practice

Feedback control

$$\omega = \underbrace{\boxed{-k_v \tan\theta}}_{\text{Orientation}} - \underbrace{\boxed{\frac{k_p y}{v \cos\theta}}}_{\text{Position}}$$

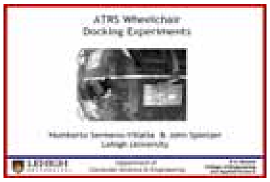
Feedback control

$$\omega = -k_v \tan\theta - \frac{\boxed{k_p} y}{v \cos\theta}$$



Experiments

- ☐ Hardware
  - Laptop with a 1.1 GHz Pentium M processor
  - Dragonfly high-resolution (1024x768) with 2.6 focal lens


ATRS Wheelchair
Docking Experiments

Humberto Sermeno-Villalta & John Spletzer
Lehigh University

Proof-of-concept Demo

Future work

- ☐ Tracking robustness must be improved
- ☐ Redundancy levels
- ☐ Assumes intrinsic and extrinsic parameters remain fixed