ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Practical Issues in Physical Sign Recognition with Mobile Devices

Christof Roduner and Michael Rohs

Institute for Pervasive Computing
Department of Computer Science
ETH Zürich

May 2006

Introduction

- Background
  - Annotation of physical objects using mobile devices (PERMID 2005)
  - Prototype based on camera phones and 2D barcode (Visual Codes)
  - Can we do it without barcodes, purely based on image matching?
  - Focus on feasibility rather than computer vision
- Outline
  - Sign annotation – use case and motivation
  - Sign recognition based on image matching
  - Comparison of matching algorithms
  - Practical problems

Digital Annotations to Physical Objects

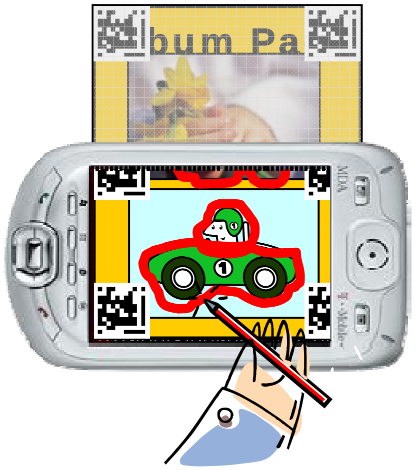
- User-generated digital media linked to physical objects
- Embed digital information into the real world
- Can be shared across space and time
- Can take multiple forms
  - Text, graphics, audio, video, hyperlinks, vCard, vCalendar

Content of Digital Annotations

- What questions do annotations answer?
  - what are similar objects?
  - what are complementary objects?
  - what similar objects are better / worse?
  - who else likes this object?
- Ratings
  - using attributes that are specific for the object class
  - using attributes of a taxonomy or ontology

Techniques for Anchoring Digital Annotations with Physical Objects

- Marker types
  - Numbers (e.g., YellowArrow)
  - Barcodes (e.g., Semapedia)
  - RFID
- Problems
  - Attaching markers is not always feasible
  - Object is not under the annotator‘s control
  - Visual markers are obtrusive
- A different approach…
  - Marker-less annotation
  - Recognize objects based on their unmodified visual appearance



Annotations by Visual Appearance

- Many regular / quadrangular shapes in urban environments
  - street signs, shop signs, indication panels
  - facades of buildings
- Use signs as annotation anchors
  - interactively supported image matching

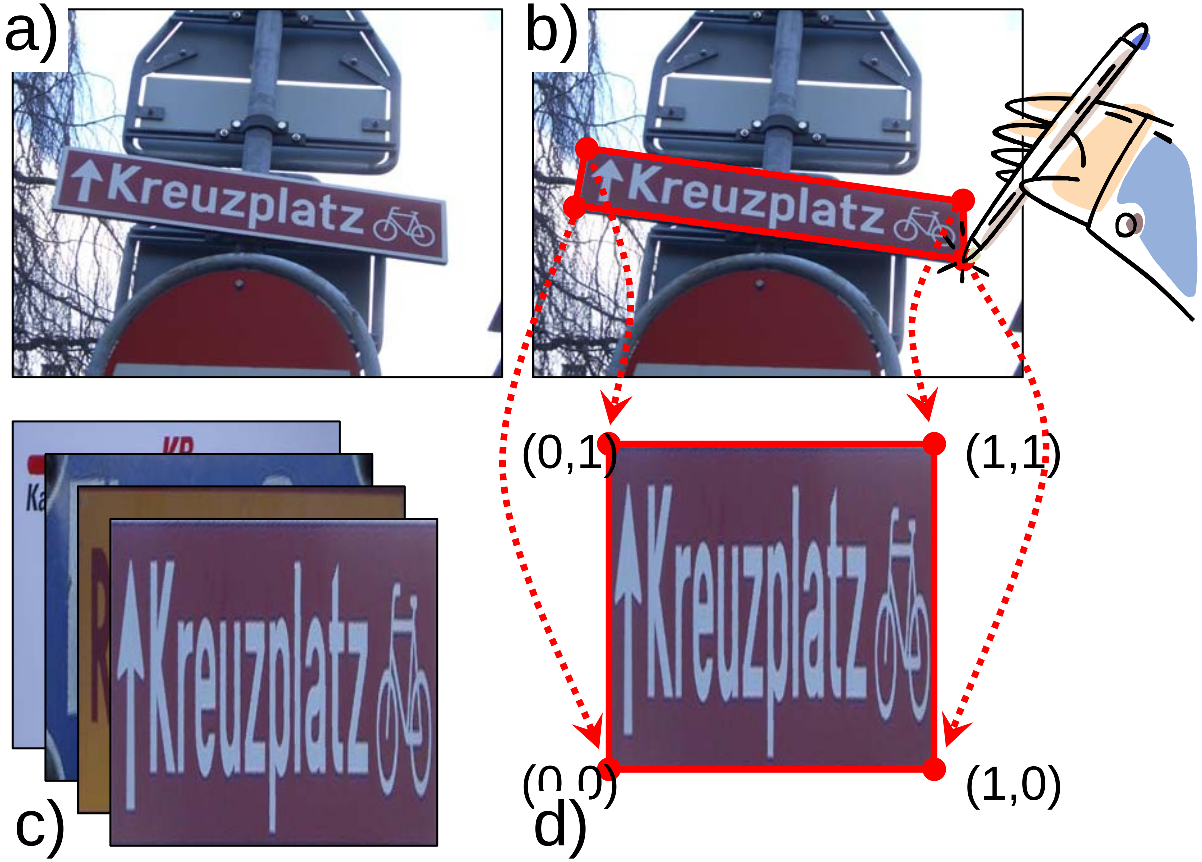
Annotating Signs using Camera Phones with Pen Input

a) captured photo

b) framing a sign with the pen
- object selection
- segmentation

c) set of templates

d) mapping framed area to unit square ("warping")



Dealing with Distortion

- Perspective distortion of sign in camera image
- Four-point correspondences
- Frame corners correspond to corners of unit square
  - Unique planar homography (projective transformation matrix)
  - Project framed part into unit square
  - Scale unit square to fixed-size request image of 128x128 pixels

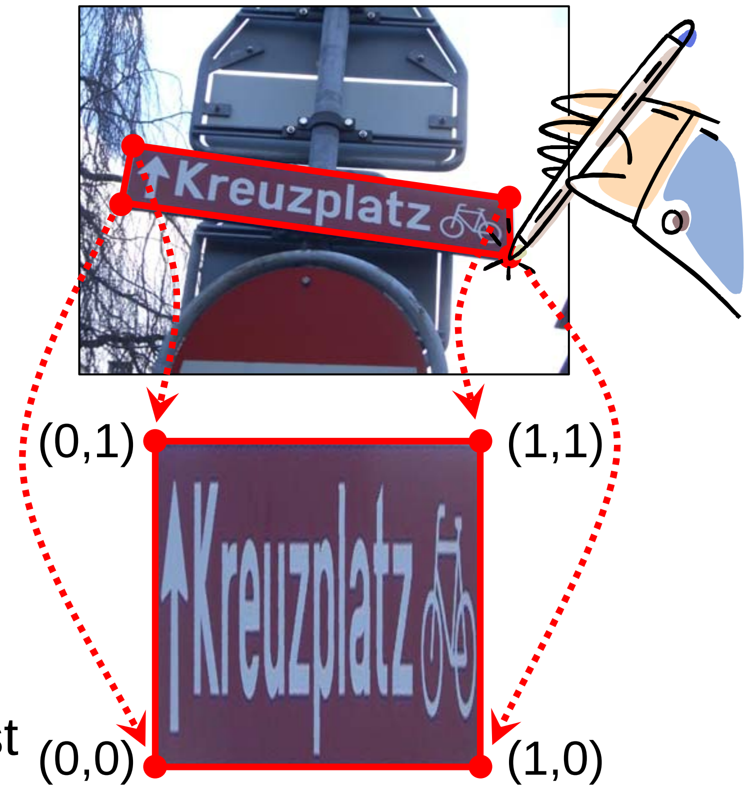


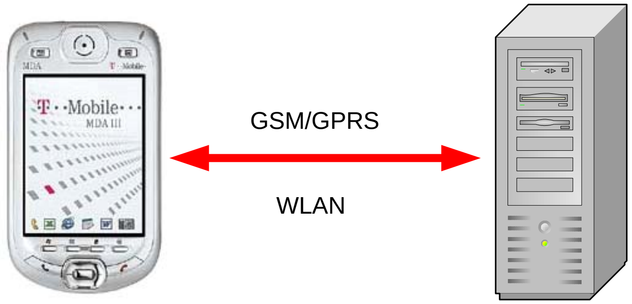
Image Matching

- Request to backend server
  - request image
- Backend server
  - executes matching algorithm
  - stores shared annotations and templates

Request image

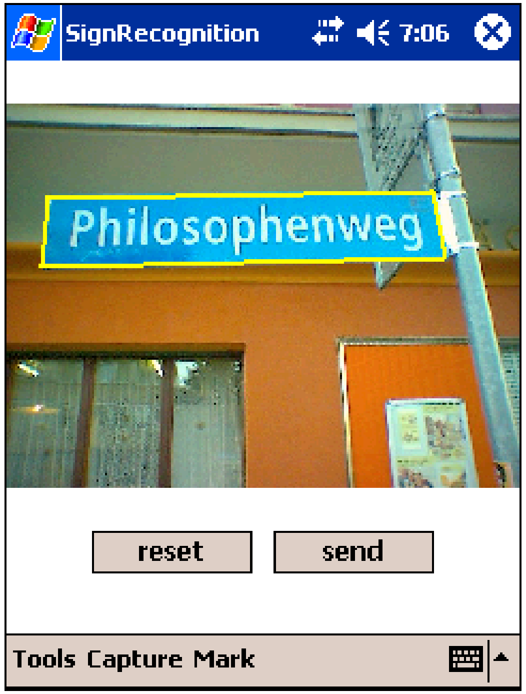
Set of templates

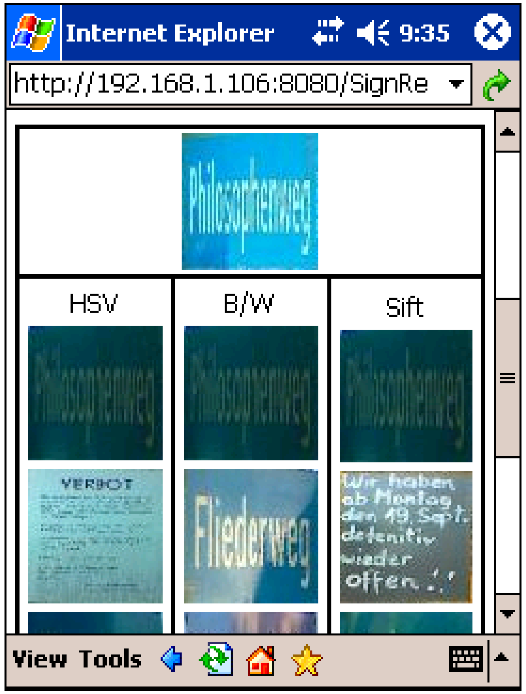
Prototype System

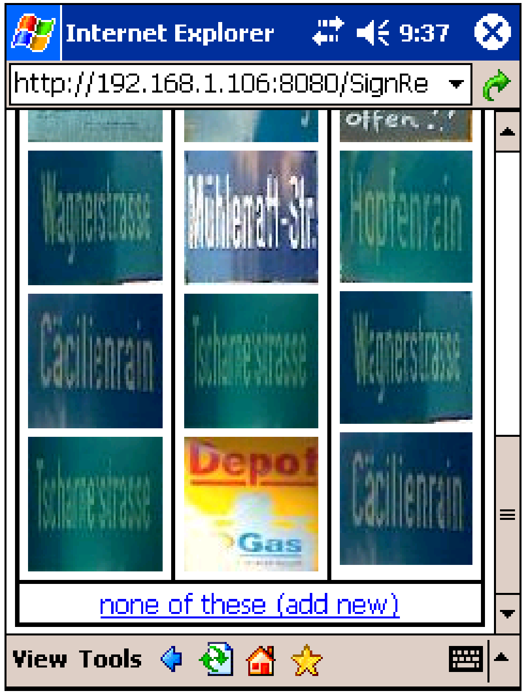


- Smartphone (T-Mobile MDA III)
  - Windows Mobile 2003
  - Pen-based input
  - Camera: 640 x 480 pixels, no optical zoom, relatively poor image quality
  - Software: Visual C++ application and web browser
- Server with Servlets and MySQL db to store templates and annotations
- Four different algorithms for image matching:
  - HSV
  - Black / White
  - SIFT (Scale Invariant Feature Transform)
  - (Wavelet)

Annotation Process







- Warped image is compared with each template
- Five best matches are presented to user as candidates
- User selects correct sign or decides to add warped image as template

Image Matching Algorithms

- HSV
  - Sum of pixel-by-pixel differences of hue value of request image and templates
- B/W
  - Correlation coefficient between monochrome request and template images
- Wavelet
  - Jacobs et al., 1995 – Java implementation based on Eikon engine
  - Signature for each image
- SIFT (Scale Invariant Feature Transform)
  - Lowe, 1999
  - Extraction of stable features (insensitive to changes in viewpoint)
  - Euclidian distance between feature vectors

Eperimental Setup

- Initialization: database with 95 template images of different objects
  - Street signs
  - Building facades
  - Company logos
  - Posters
- Evaluation: second snapshot for every object in database
  - Different perspective, lighting, and distance
  - For each matching algorithm: select 5 most similar templates from database
- Recognition successful if correct object is among the top 5 hits returned by algorithm

Results: Overview

Algorithm	HSV	Wavelet	B/W	SIFT
# snapshots	95	95	95	95
# recognized	81	69	86	89
% recognized	85.3%	72.6%	90.5%	93.7%
mean rank	1.54	1.62	1.37	1.37
% rank 1	74.1%	76.8%	86.1%	77.5%

- SIFT‘s offers the best recognition rate
- Despite their simplicity, HSV and B/W perform relatively well
- Wavelet algorithm seems less suitable
- Recognized objects are usually ranked first or second in the candidate list

Results: Street Signs

- Street signs are special
  - Omnipresent
  - Similar appearance
  - Material very susceptible to reflections and glare

Results: Street Signs

Without street signs

Algorithm	HSV	Wavelet	B/W	SIFT
# snapshots	83	83	83	83
# recognized	72	65	75	78
% recognized	86.8%	78.3%	90.4%	94.0%
% gained / lost	+1.5%	+5.7%	-0.2%	+0.3%

Street signs only

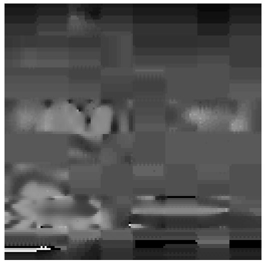
Algorithm	HSV	Wavelet	B/W	SIFT
# snapshots	12	12	12	12
# recognized	9	4	11	11
% recognized	75.0%	33.3%	91.7%	91.7%
% gained / lost	-10.3%	-39.3%	+1.1%	-2.0%

Recognition Problems

- Subtle features
  - Few distinguishable features after warping
  - Negative impact of small tanslations on pixel-based algorithms

Original

Warped



Hue

Black / White

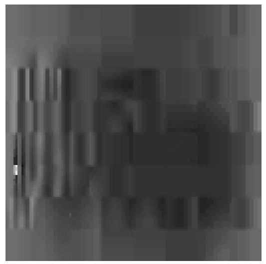
Recognition Problems

- Large width and small height
  - Warping changes aspect ratio leading to few clear features

Original

Warped

Hue



Black / White

Recognition Problems

- Perspective
  - Only problematic with pixel-based algorithms, not with SIFT

Recognition Problems

- Blurred images
  - Mainly problematic for SIFT
  - Minor issue as blur can be corrected by user

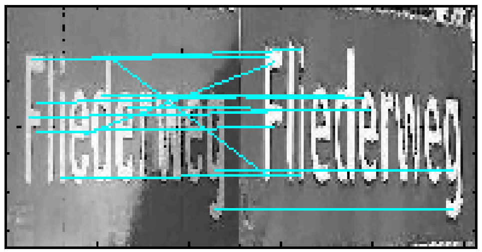
- Reflections
  - Reflecting surfaces introduce spurious features

- Non-rectangular objects
  - Not a problem for SIFT

Recognition Problems

- Lighting conditions
  - Colors changes (problematic for HSV) and reflections (problematic for B/W)

Original | Warped | Hue | Black / White



Summary

- Sign recognition works well with very simple pixel-based algorithms and best with SIFT
- Corner marking can be done accurately
- Corner marking could be omitted with SIFT
  - User still needs to select object to annotate
  - Pen input required
- Possibilities for improvement:
  - Limit search space by considering context parameters (cell id, time, weather conditions)
  - Use of more suitable data structures

Thank you!

Questions?

www.vs.inf.ethz.ch

Camera Phones with Pen Input for Generating Digital Annotations to Real-World Objects

- Interaction possibilities of camera phones with pen input
- Techniques for anchoring digital annotations with physical objects
  - Visual codes for annotations of items in printed photos
  - Suitability of camera phones as annotation devices
  - Annotation of signs using four-point correspondences