ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Camera Phones with Pen Input as Annotation Devices

Michael Rohs and Christof Roduner

Institute for Pervasive Computing
Department of Computer Science
ETH Zürich

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
  - annotations by visual appearance using image matching
- How can a mobile user interface for a generic annotation system be structured?
  - creation, access, sharing, organization of annotations

Outline

- Introduction
- Digital Annotations with Visual Codes
- Sign Annotations with Image Matching
- Open Questions

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
  - uring attributes of a taxonomy or ontology

Requirements of Annotation Systems versus Camera Phone Features

- Identification of physical objects
  - camera and image processing, barcode readers, RFID / NFC
- Graphical annotations of physical objects
  - marker-based registration
- Sharing annotations and getting up-to-date information
  - wireless connectivity
- Handle annotations in multiple media types
- Availability in everyday settings

Benefits of Pen-Based Input

- Fine-grained annotations for objects captured with the camera
  - draw frames around items
  - draw arrows to give directions
  - put predefined icons onto object images
- Interactive support for image matching algorithms
  - telling to the system what items are important
  - segment foreground from background

Problems of Mobile Input

- Simplify creation of digital annotations
  - avoid text input
- Classification in a taxonomy
  - selection within taxonomy
- Widgets for entering parameters
  - ratings
- Forms with attributes that are specific to the objects class

Outline

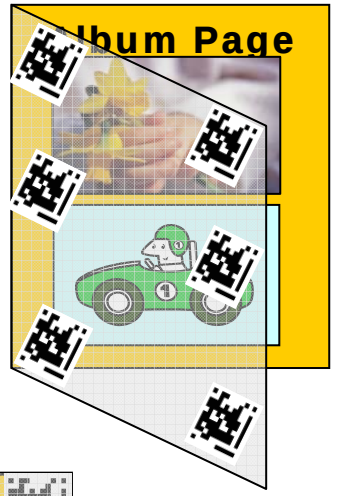
- Introduction
- Digital Annotations with Visual Codes
- Sign Annotations with Image Matching
- Open Questions

Annotating Photos in a Physical Album

- Attach media to a physical photo album
- Mobile annotations: voice, text, music, files, links
- Platform:
  - Smartphone (T-Mobile MDA III)
  - Windows Mobile 2003
  - Pen-based input
  - Camera-equipped (640 x 480 pixels)

Annotating Photos in a Physical Album

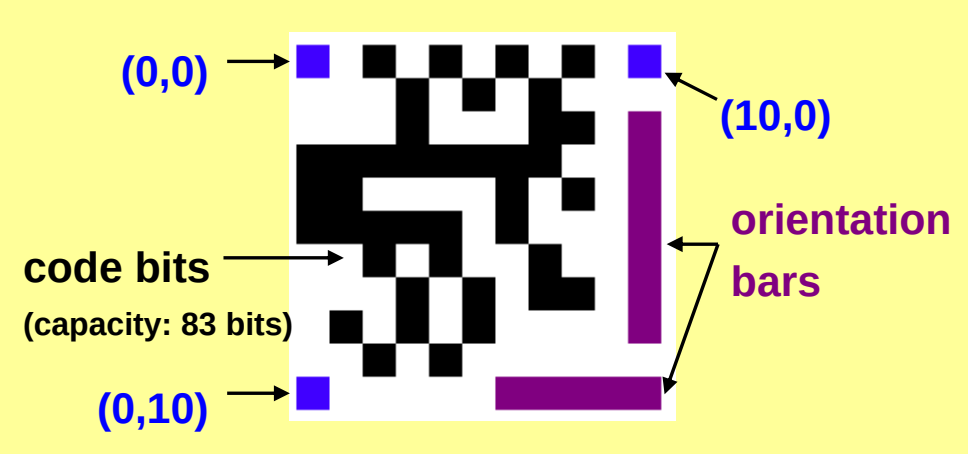
- Two-dimensional markers (Visual Codes) attached to foil covering album pages.



- Annotation process:
  - Take picture of page
  - Draw polygon around object
  - Attach media
  - Store polygon and media on server

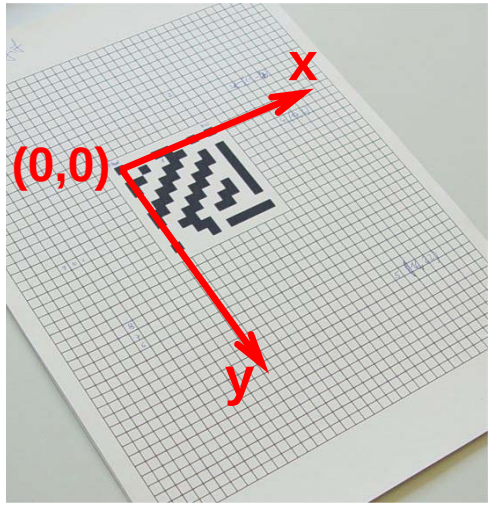
Visual Codes for Camera Phones

- For low-resolution phone cameras
  - e.g. 160x120 pixels in view-finder mode
  - requires coarsely grained code
- Arbitrary orientation because of camera mobility requires special code features
- Lightweight recognition algorithm

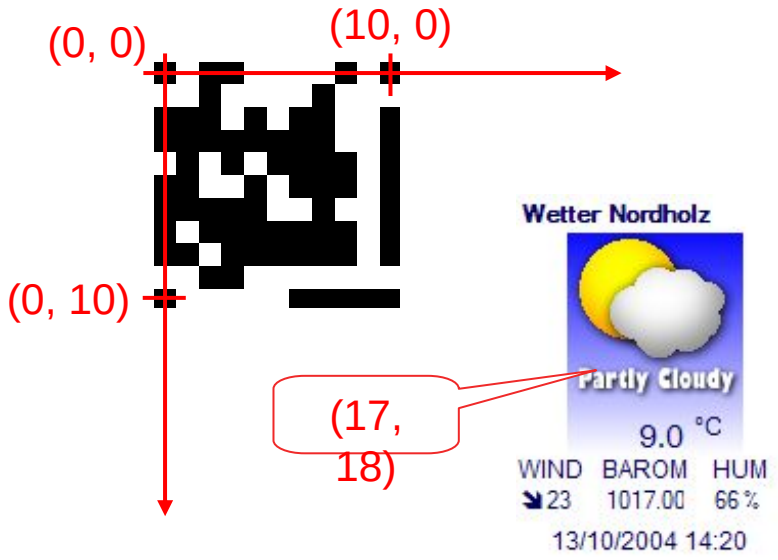


- Storage capacity
  - 76 bits / 96 bits
  - error detection (83,76,3) / (103,96,3) linear codes

Code Coordinate System



- Each visual code defines a local coordinate system
  - invariant to projective distortion
- Projective mapping (planar homography)
  - convert image pixel coordinates to code coordinates
    → create annotation polygon
  - convert code coordinates to image pixel coordinates
    → display annotation polygon



Annotation Process

Issues with Current Implementation

- Marker size: 2 x 2 cm
  - Higher resolution (1 vs. 0.3 megapixel)
- Number of markers per page (6)
  - Large distance between phone and code: code not readable
  - At least one code must be visible if camera close
- Each code has its own coordinate system
  - User must initialization code cluster
  - Allows for the transformation of coordinate systems
  - Alternative: codes pre-printing at fixed positions

Outlook

- What else to annotate...
  - X-ray images
  - Construction plans
  - Crime scenes

Outline

- Introduction
- Digital Annotations with Visual Codes
- Sign Annotations with Image Matching
- Open Questions

Annotations by Visual Appearance

- Attaching markers is sometimes not an option
  - visual markers might be too obtrusive
  - objects not under the annotator's control
- Recognize objects based on their unmodified visual appearance

Annotations by Visual Appearance

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

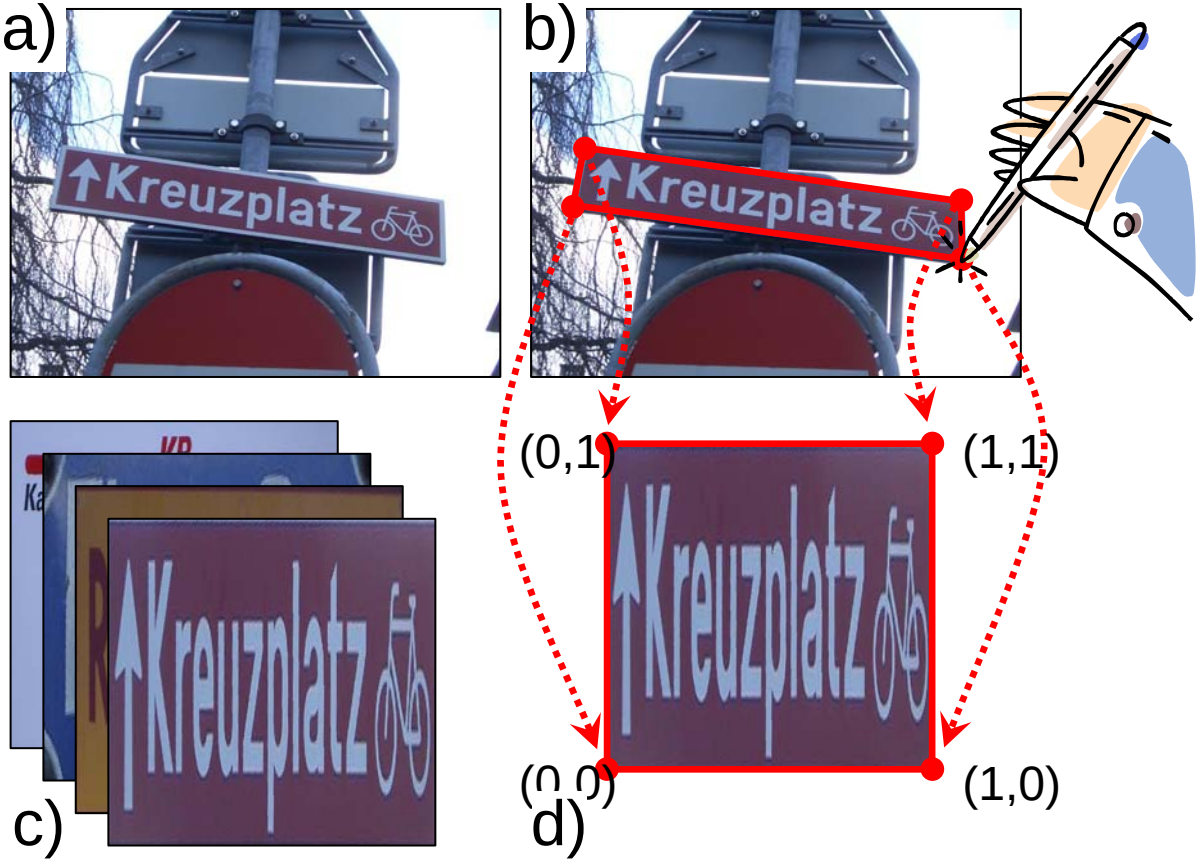
Annotating signs using camera phones with pen input

a) captured photo

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

c) set of templates

d) mapping framed area to unit square



Four-Point Correspondences

- Perspective distortion of sign in camera image
- Project framed part into unit square
  - frame corners correspond to corners of unit square
  - unique planar homography (projective transformation matrix)
  - scale unit square to fixed-size request image of 480x480 pixels

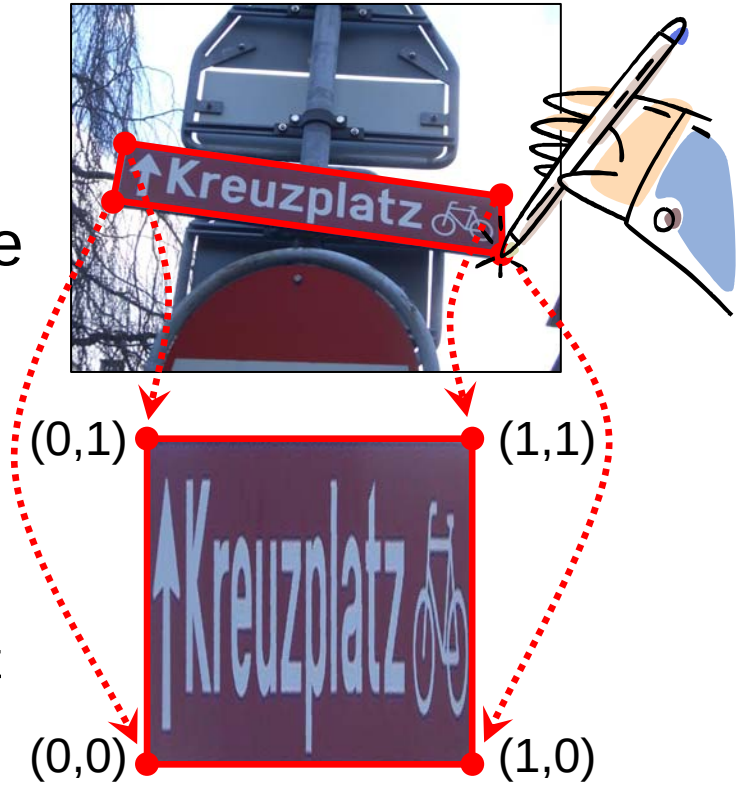
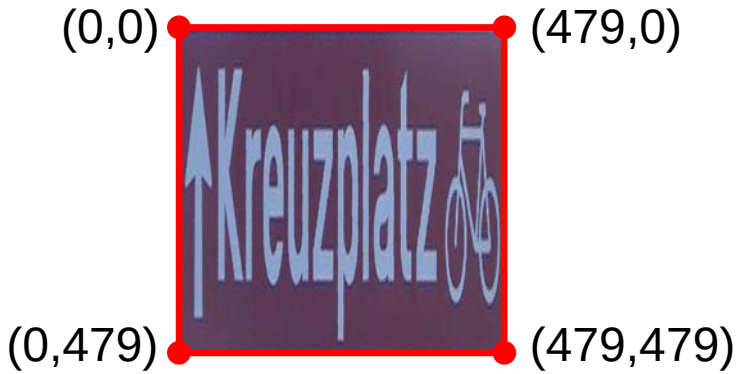


Image Matching

- Request to backend server
  - request image (480x480 pixels)
  - context parameters
    - GSM cell IDs
    - time of day (morning, noon, afternoon)
    - weather conditions (sunny, cloudy, rainy)
- Backend server
  - stores shared annotations and templates
  - executes matching algorithm



Request image

Set of templates

Image Matching

- Matching algorithm
  - sum of pixel-by-pixel differences of hue value of request image and template images
  - template images filtered by context parameters
    - cell IDs, time, weather conditions
  - filtering limits search space to a few dozen candidates
- Initial experiments show that matching works
  - problem: signs are very street similar and are not unique
  - shop signs show more variation in visual appearance

Usability Issues

- Approach is beneficial if it is less effort to take snapshot and tap corners than to enter a unique descriptor
- Upload of request image takes some time
  - less problematic in the future
- A conscious effort is required by the user
  - no automatic detection of annotations
  - no augmented reality
- Usability study necessary in some application context

Outline

- Introduction
- Digital Annotations with Visual Codes
- Sign Annotations with Image Matching
- Open Questions

Open Questions

- How accurately do users draw frames?
- In what way do imprecise frames degrade performance?
- Are there better image matching algorithms?
  - that require less interactive support
  - that don’t require interactive support
  - that are tolerant against imprecise user input
- What are compelling applications?
  - pervasive urban games? restaurant recommenders?

Summary

- Creating annotations of physical objects using camera phones with pen input
- Camera phones fulfill technical requirements
  - object identification, online connectivity, multiple media types, availability in everyday situations
- Pen-based input allows for fine-grained annotations
- Marker-based approach
  - annotate items on a printed page
- Interactively supported image matching approach
  - annotating unmodified objects

Thank you!

Questions?

www.vs.inf.ethz.ch