Event-Driven Stereo for 3D 360° Panoramic Vision

3D 360° Panoramic Vision Using High-speed Rotating Dynamic Vision Sensors

Stephan Schraml

New Sensor Technology
Digital Safety & Security Department

© AIT Austrian Institute of Technology GmbH
02/2016
Outline

I. Dynamic vision sensor
   • Bio-inspired sensor principle
   • Characteristics
   • Applications

II. 360° panoramic stereo vision system
   • Motivation
   • 360° panorama sensor system
     • Rotating stereo sensor properties
     • Geometrical analysis
   • Event-driven stereo matching method for 3D panoramic vision
     • Challenges
     • Stereo Matching
   • Results and Videos
   • Demonstration panoramic vision stereo sensor
Dynamic vision sensor (DVS) technology

- Bio-inspired optical sensor principle
  → spike based communication (address events)

- Sensitive to illumination changes
  → on-chip processing
  → data compression
  → redundancy suppression

- Feature-based and event-driven
  (instead of frame-based) image processing

- Mimic the “transient pathway” of the human/animal retina
  - “Dynamic Vision”
Sensor principle of operation

Key facts: bio-inspired dynamic vision line sensor

- **Autonomous** self-spiking and **asynchronous** operating pixel cells that
- Respond to **relative local illumination changes** (1. derivative) by generating
- Digital **events** encoding the identities of pixels that see these changes.

Sensor characteristics

- Wide dynamic range (>120 dB)
- High temporal resolution (< 1µs) (at the same time)
- Efficient encoding of local changes in scene reflectance (no images generated)
  - Redundancy suppression
  - Low data rate
- Robust to variations in illumination conditions

- Asynchronous stream of events - NO conventional images!
DVS - ATIS change detection Video

Image like representation

Events are accumulated for representation (20ms)

white..increasing of illumination (ON-event)
black..decrease of illumination (OFF-event)
Example Applications

**IVS High-speed Industrial Vision Sensor (DLS)**
- Quality control
- High-speed machine-vision applications

*) https://www.youtube.com/watch?v=evl2TdexOn8

**UCOS Stereo Vision Sensor (TMPDIFF 128)**
- e.g. People Counting

**TDS Traffic Data Sensor (TMPDIFF128)**
- Environmental monitoring

**TUCO-3D 3D 360° HDR Panoramic Imaging (ATIS-L)**
- Mobile platforms, navigation
- Surveillance
II. 360° panoramic stereo vision

Motivation

- Advantage
  - Full view, continuous observation of landmarks, ...

- Where is 3D 360° panorama actually used?
  - Google car (self driving)

- More applications
  - Mobile robots
    - Slam, mapping, landmark extraction
    - Autonomous navigation, and site modeling
    - Obstacle detection and tracking
  - Video surveillance
  - Virtual reality, Environmental simulation

- Not a new idea, but there is hardly any existing system.
Existing Technology

Laser-based technology
26.8° @5-15Hz, Velodyne HDL-64E
- Highly reliable
- 32 / 64 lines
- High price (30-70 K€)

360° (but 2D only)
Ladybug, 15 fps, Point Gray Research

Two omnidirectional cameras, aligned vertically using parabolic mirror
- Good overlapping of both views,
- Parallel epipolar lines
  Gluckman et al. (1998)
  - 600 x 60 pixel @32 disp, 7 fps
  Koyasu et al. (2001)
  - 720 x 100 @80 disp, 5 fps

Single Camera Catadioptric Stereo System
  Jungho et al. (2001)
  - Used for visual slam

Two rotating parallel stereo cameras
- Parallel epipolar lines
  Jiang and Okutomi (2006)
  - 640 x 480 pixel
Existing Technology

**Laser-based Technology**

- **+ accurate**
- **- expensive (30k-70k€)**
- **- 64 lines**

---

**Catadioptric systems**

- **+ capture panoramic scene in one image**
- **- distortions**
- **- complex geometry**
- **- lower image resolution**

- Laser-based technology
  - 26.8° @5-15Hz, Velodyne HDL-64E
  - Highly reliable
  - 32 / 64 lines
  - High price (30-70 K€)

- 360° (but 2D only)
  - Ladybug, 15 fps, Point Gray Research

**Products**

---

**Research**

**Multiperspective Systems**

- **- assemble a panoramic view from a collection of images**
- **+ distortion free**
- **+ high and adaptive spatial resolution**
- **+ simple geometrical model**

- Two rotating parallel stereo cameras
  - Parallel epipolar lines
  - Jiang and Okutomi (2006)
    - 640 x 480 pixel
  - Jungho et al. (2001)
  - Used for visual slam
  - Koyasu et al (2001)
    - 720 x 100 @80 disp, 5 fps
  - Gluckman et al. (1998)
    - 600 x 60 pixel @32 disp, 7 fps

---

24/02/16
Multi-perspective imaging

- All proposed systems rely on frame based imaging

- Limitations
  1. Limited lighting in natural scenes requires long exposure time
  2. Individual images have to be captured and concatenated
     • Eg: 0.1° horizontal resolution
       ⇒ results in 3600 images @ 1 rps or exposure time < 28 ms
       ⇒ results in 36000 images @ 10 rps or exposure time < 2.7 ms

- New sensor developments
  • Panoscan, full scan in 7s-54s, HDR in 5 min
  • Roundshot D3, full scan in 3s-60s, HDR na.
  • Spheron, full scan in 12s – 7min, HDR in 45 min

- Gap between fast stereo panoramic imaging and real-time capabilities
Panoramic stereo sensor system

- A pair of rotating line DVS
  - Events are transmitted as a stream in address-event representation
    - Address y (position), Time of occurrence t, Polarity p (OFF, ON)
  - Rotating the DVS → will detect edges in a static scene

Available stereo matching methods do not work (well)
  - Rely on prerequisites (dense images) that do not hold for events.

Other stereo matching methods using events:
  - (Adapted) classical stereo method based on area sensors (UCOS).
  - Temporal coherence is used to identify correlated events between the left and right sensor data of area sensors.
Panoramic stereo sensor system

- A pair of rotating line DVS
  - Events are transmitted as a stream in address-event representation
    - Address y (position), Time of occurrence t, Polarity p (OFF, ON)
  - Rotating the DVS -> will detect edges in a static scene

Goal: Find a Stereo Matching Method

- Challenges
  - Non-simultaneous vision
  - Sparse data
Challenges I

Non-simultaneous vision

- Dynamic scenes, motion

→ Disparity formulated as a function of time

\[ \text{disparity} = tp2 - tp1 \]

\[ \alpha \approx \text{disparity} \]

\[ r \approx \frac{\text{baseline} \times \sin(\phi)}{\sin(\alpha)} \]

- Model the effect of moving objects
- Correct the disparity from a moving sensor
Challenges II

Sparse data

- Direct matching not possible (absence of data)
- Asynchrony of events: temporal uncertainties
- Autonomy of pixel cells: number of events differ

Transformation \( e(y,t,p) \) → restore spatial context

Eventmap

- \( e_{n-5} \) - \( e_n \) time
- \( e_{n-1} \) - \( e_n \) time

\( e_{n-4} \) - \( e_{n-3} \) - \( e_{n-2} \) - \( e_n \) time

\( \text{Event-Driven Stereo Matching} \)
Event-Driven Stereo Matching I

...based on the idea that correlated sequences of events are likely to have a similar event-count and inter-event timing. Variations in the timing and the number of generated events can be quantified by the proposed cost measure.

Algorithm:
1. For each left event take the minimal distance to any right event
2. Build the sum of these distances
3. Repeat 1-2 for each right event
Event-Driven Stereo Matching II

...based on the idea that correlated sequences of events are likely to have a similar event-count and inter-event timing. Variations in the timing and the number of generated events can be quantified by the proposed cost measure.

\[
NZD^j(x, y) = \min_{i} (\text{abs}(x - i) \mid E^j(i, y) \neq 0)
\]

\[
C^{L,R}(x, y, d) = \sum_{w \in W} E^L(x+w, y) \cdot NZD^R(x+w+d, y)
\]

\[
\hat{C}^{L,R}(x, y, d) = \sum_{w \in W} NZD^L(x+w, y) \cdot E^R(x+w+d, y)
\]
Event-Driven Stereo Matching III

- **Advantages**
  - no images required
  - exploits temporal information of events
  - robust to timing uncertainties and diverging event counts
  - core algorithm can be implemented as matrix functions

- **Limitations**
  - in low light <100 lux, increase of noise
  - strong edges still ok
  - can increase sensitivity, but have to balance the amount of data

- **Motion**
  - non-simultaneous stereo vision, motion causes distortions in depth map
  - disparity form moving sensor can be corrected, if motion vector is known

- **Implementation**
  - line – wise stereo matching, local operations
  - on-line implementation possible (reduced memory)
Results I - Examples
Results II - HDR

DVS directly facing a 1KW halogen lamp

Event-driven stereo matching result

(mobile cam)
Results III - HDR

Conventional camera | no additional light | stereo result

Conventional camera | 1KW halogen lamp | stereo result

DVS | 1KW halogen lamp | stereo result
Gray-scale reconstruction

\[ I(x_0, y_j) = 0 \]
\[ I(x_i, y_j) = \beta I(x_{i-1}, y_j) + \phi(x_i, y_j) \]
Results IV - Motion

Sensor mounted on platform is moving with 0.5 m/s. Markers on the left and the right are on equal distances.

distorted disparity

motion corrected disparity
Results V - Video

Event-Driven Stereo
Thank you, for your attention!

Demo
References

TUCO-3D Specs as of Jan. 2014

**Specifications**

- **Lens focal length:** 4.5mm
- **Vertical FOV:** 48.9°
- **Horizontal FOV:** 360°
- **Depth:** 3D camera coordinates
- **Scanning speed:** 3600° /sec (max. 10 rps)
- **Image resolution:** 2300 (H) x 1024 (V)
- **On-chip compression:** > 30
- **Detector type:** CMOS (Dynamic Vision Sensor)
- **Dynamic range:** > 120dB
- **Output:** Gigabit Ethernet
- **Weight:** 1.42 kg
- **CE EMC type:** Class A
- **EMC Norms:** 2004/108/EC, EMC directive 2002/95/EC, RoHS directive

**Dimensions:**

- **Top (Dia x H):** 80 x 140 mm
- **Bottom (W x L x H):** 110 x 125 x 70 mm