A Probabilistic Approach to ToF and Stereo Data Fusion

Carlo Dal Mutto, Pietro Zanuttigh, Guido M. Cortelazzo

University of Padova
Papers

- C. Dal Mutto, P. Zanuttigh, G.M. Cortelazzo
  “A probabilistic approach to ToF and Stereo Data Fusion”
  3DPVT10, Paris, France, May 2010

- C. Dal Mutto, P. Zanuttigh, G.M. Cortelazzo “Accurate 3D Reconstruction by Stereo and ToF Data Fusion”
  GTTI Meeting, Brescia, Italy, June 2010 (BEST PAPER AWARD)
Problem

- 3D dynamic scene estimation
- Traditionally:
  Stereo vision systems (N>1 standard cameras)
- New technology:
  Matricial ToF range cameras (e.g. Mesa SR4000)
- Our approach:
  1 ToF camera + stereo pair
Previous works

• [Zhu et al. 2008]: Empirical application of Belief Propagation algorithm

• [Gudmundsson et al. 2008]: ToF as initialization for a hierarchical stereo algorithm

• [Beder et al. 2008]: Patchlets surface representation by exploiting the information from a Stereo Pair and a ToF
Outline

• Background (Stereo + ToF)
• Preliminary step: system calibration
• Problem definition and motivation
• Fusion Algorithm
• Results
• Conclusions
Conventions

- Depth = z-axis
- T = ToF camera
  - $A_T$ = Amplitude image, $D_T$ = Depth image, $I_T$ = Both ToF images
- L = Left RGB camera
  - $I_L$ = Color image
- R = Right RGB camera
  - $I_R$ = Color image
- S = Stereo Pair \{L+R\}
  - $I_S$ = Both images
Outline

• Background (Stereo + ToF)
• Preliminary step: system calibration
• Problem definition and motivation
• Fusion Algorithm
• Results
• Conclusions
Stereo (I)

- INPUT: 2 rectified RGB images \( \{I_L, I_R\} \)
Stereo (I)

- INPUT: 2 rectified RGB images $\{I_L, I_R\}$

\[ u - d, v \]

\[ [u, v] \]

\[ [u - d, v] \]

\[ d \propto \frac{1}{z} \]
Stereo (2)

- Correspondent pixels matching
  - Dissimilarity measure on a “Aggregation Support” (e.g. Absolute Difference)
  - Lowest dissimilarity ➔ correspondent points
Stereo (3)

- Output: scene disparity map (the disparity value $d$ is associated to each pixel)

- Disparity map $\rightarrow$ 3D: Triangulation principle
Stereo (4)

- Doesn’t work with textureless scenes (i.e. low SNR)
Stereo (4)

- Doesn’t work with textureless scenes (i.e. low SNR)
Stereo (5)

- Against low SRN
- Typical geometrical properties of the scene:
  - Visual consistency [Sun et al. 2005]
  - MRF smoothness assumption [Sun et al. 2003], [Kolmogorov et al. 2001]
  - Local consistency [Mattoccia 2009]
- Global methods
- Slow (post-processing)
ToF Camera (1)

- Amplitude Modulation model for a matricial ToF range camera

- 24 infrared illuminators
- Infrared carrier: 850 nm
- Modulated sinusoidal: 30 MHz
- Speed: $c = 3 \times 10^8 \text{ m/s}$
- Wavelength: 10 m
ToF Camera (1)

- Amplitude Modulation model for a matricial ToF range camera

- 24 infrared illuminators
- Infrared carrier: 850[nm]
- Modulated sinusoidal: 30[MHz]
- Speed: \( c = 3 \times 10^8[m/s] \)
- Wavelength: 10[m]
ToF Camera (1)

- Amplitude Modulation model for a matricial ToF range camera

- 24 infrared illuminators
- Infrared carrier: 850[nm]
- Modulated sinusoidal: 30[MHz]
- Speed: \(c = 3 \times 10^8[m/s]\)
- Wavelength: 10[m]
ToF Camera (2)

- Receiver matrix (176x144)

\[ r(t) \rightarrow 120[MHz] \rightarrow r(nT) \rightarrow \Delta \varphi \rightarrow \frac{1}{2} \frac{c}{2\pi f} \rightarrow d \]
ToF Camera (2)

- Receiver matrix (176x144)

\[ r(t) \rightarrow 120[MHz] \rightarrow r(nT) \]\n
\[ \Delta \varphi = \frac{1}{2} \frac{c}{2\pi f} \rightarrow d \]
ToF Camera (3)

- Output:

Depth Image: $D_T$

Amplitude Image: $A_T$
## Why ToF and Stereo

<table>
<thead>
<tr>
<th></th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ToF</strong></td>
<td>✓ Fast (20-50 fps)</td>
<td>x Noisy</td>
</tr>
<tr>
<td></td>
<td>✓ Reliable</td>
<td>x Low resolution (176x144)</td>
</tr>
<tr>
<td></td>
<td>✓ No texture scenes</td>
<td>x Illumination sensitive</td>
</tr>
<tr>
<td></td>
<td>✓ High reflectance scenes</td>
<td>x Low reflectance scenes</td>
</tr>
<tr>
<td><strong>Stereo</strong></td>
<td>✓ Less noisy</td>
<td>x Slow (&lt;10fps)</td>
</tr>
<tr>
<td></td>
<td>✓ High resolution</td>
<td>x Not always reliable</td>
</tr>
<tr>
<td></td>
<td>✓ Illumination insensitive</td>
<td>x No texture scenes</td>
</tr>
</tbody>
</table>
Outline

- Background (Stereo + ToF)
- Preliminary step: system calibration
- Problem definition and motivation
- Fusion Algorithm
- Results
- Conclusions
Calibration (1)

- **Trinocular system:**
  - 2 high resolution RGB cameras (stereo pair)
  - 1 low resolution ToF camera

- **3 single camera calibrations:**
  - Intrinsic parameters (Pinhole model)
  - Radial and tangential distortion (Heikkila model)
  - Systematic error in distance measurement for ToF

- **Multi-camera calibration**
  - Extrinsic parameters (i.e. rototranslations between the cameras)
Calibration (2)

- Stereo camera calibration for the stereo pair
  - Standard procedure
  - Calibration Toolbox for Matlab (or OpenCV)

- Stereo pair calibrated

- Estimation of the rototranslation of the ToF camera $[\hat{R}_T | \hat{t}_T]$ with respect to the stereo pair
Calibration (3)

\[
\begin{align*}
\hat{R}_T | \hat{t}_T \\
p^n_T \\
p^n_S, n = 1, \ldots, N
\end{align*}
\]

\[
\arg \min_{[R_T | t_T]} \sum_{n=1}^{N} d(p_S^n; [R_T | t_T] \times p^n_T)
\]
Calibration (4)

- Standard calibration checkerboard
- ToF:
  - 2D corners coordinates measured on the amplitude image $A_T$
  - 3D coordinates by backprojection with the depth coordinate ($z$) from $D_T$
- Stereo:
  - 2D corners coordinates measured on both the images from camera L and camera R
  - 3D coordinates by triangulation (referred to L)
Calibration (4)

- Absolute orientation problem
  Rototranslation between 2 different “clouds” of 3D points

- Closed form solution: Horn’s Algorithm
  [Horn 1987]

- Account for outliers: RANSAC
  [Fishler & Bolles 1981]
Calibration (5)

- OUTPUT:
  - Backprojection of ToF measurements: Depth map $\rightarrow$ 3D
  - Projection onto stereo images: 3D $\rightarrow$ Stereo images points

- Calibration Error: 0.6-0.7 [cm] in 3D, 0.8-0.9 [pxl] in 2D
Outline

• Background (Stereo + ToF)
• Preliminary step: system calibration
• Problem definition and motivation
• Fusion Algorithm
• Results
• Conclusions
GOAL

• Overall goal:
  best stereo features + best ToF features:
  • High accuracy
  • High resolution
  • Reliability

• This work goal:
  • High accuracy
  • Formal problem definition and resolution
Outline

• Background (Stereo + ToF)
• Preliminary step: system calibration
• Problem definition and motivation
• Fusion Algorithm
• Results
• Conclusions
Fusion Algorithm Input

- Stereo pair + 1 ToF camera

\[ I_T = \{ A_T, D_T \} \]
Fusion Algorithm Input

- Stereo pair + 1 ToF camera

\[ I_L \rightarrow I_T = \{ A_T, D_T \} \rightarrow I_R \]
Fusion Algorithm Input

- Stereo pair + 1 ToF camera

\[ I_L, I_R \]

\[ I_T = \{ A_T, D_T \} \]

\[ I_S = \{ I_L, I_R \} \]
Probability model

\[ \hat{Z} = \arg \max_Z P[Z | I_T, I_S] = \arg \max_Z \frac{P[I_T, I_S | Z] P[Z]}{P[I_T, I_S]} \]
Probability model

• \( \hat{Z} \) estimate of scene depth, wrt the reference frame of \( T \) (both images and 3D)

\[
\hat{Z} = \arg \max_Z P[Z|I_T, I_S] = \arg \max_Z \frac{P[I_T, I_S|Z]P[Z]}{P[I_T, I_S]}
\]

\[
\hat{Z} = \arg \max_Z \frac{P[I_T, I_S|Z]P[Z]}{C}
\]

\[
\hat{Z} = \arg \max_Z \frac{P[I_T, I_S|Z]P[Z]P[Z]}{P[I_T]P[I_S]}
\]

\[
\hat{Z} \approx \arg \max_Z \frac{P[I_T|Z]P[Z]}{P[I_T]} \frac{P[I_S|Z]P[Z]}{P[I_S]}
\]

\[
\hat{Z} = \arg \max_Z P[Z|I_T, I_S] \approx \arg \max_Z P[Z|I_T]P[Z|I_S]
\]
Probability model

- \( \hat{Z} \) estimate of scene depth, wrt the reference frame of \( T \) (both images and 3D)

- \[ \hat{Z} = \arg \max_Z P[Z|I_T, I_S] = \arg \max_Z \frac{P[I_T, I_S|Z]P[Z]}{P[I_T, I_S]} \] (Bayes Rule)

- \[ \hat{Z} = \arg \max_Z \frac{P[I_T, I_S|Z]P[Z]}{C} \]

- \[ \hat{Z} = \arg \max_Z \frac{P[I_T, I_S|Z]P[Z]P[Z]}{P[I_T]P[I_S]} \]

- \[ \hat{Z} \approx \arg \max_Z \frac{P[I_T|Z]P[Z]}{P[I_T]} \frac{P[I_S|Z]P[Z]}{P[I_S]} \]

- \[ \hat{Z} = \arg \max_Z P[Z|I_T, I_S] \approx \arg \max_Z P[Z|I_T]P[Z|I_S] \]
Probability model

- \( \hat{Z} \) estimate of scene depth, wrt the reference frame of T (both images and 3D)

\[
\hat{Z} = \arg \max_Z P[Z|I_T, I_S] = \arg \max_Z \frac{P[I_T, I_S|Z]P[Z]}{P[I_T, I_S]} \\
= \arg \max_Z \frac{P[I_T, I_S|Z]P[Z]P[Z]}{P[I_T]P[I_S]} \\
\approx \arg \max_Z \frac{P[I_T|Z]P[Z]P[I_S|Z]P[Z]}{P[I_T]P[I_S]} \\
\hat{Z} \approx \arg \max_Z \frac{P[I_T|Z]P[Z]}{P[I_T]} \frac{P[I_S|Z]P[Z]}{P[I_S]} \\
\hat{Z} = \arg \max_Z P[Z|I_T, I_S] \approx \arg \max_Z P[Z|I_T]P[Z|I_S]
\] (Bayes Rule)

(Den. indep. from max argument)
Probability model

- \( \hat{Z} \) estimate of scene depth, wrt the reference frame of \( T \) (both images and 3D)

\[
\hat{Z} = \arg \max_Z P[Z|I_T, I_S] = \arg \max_Z \frac{P[I_T, I_S|Z]P[Z]}{P[I_T, I_S]}
\]  
(Bayes Rule)

- \( \hat{Z} = \arg \max_Z \frac{P[I_T, I_S|Z]P[Z]}{C} \)  
(Den. indep. from max argument)

- \( \hat{Z} = \arg \max_Z \frac{P[I_T, I_S|Z]P[Z]P[Z]}{P[I_T]P[I_S]} \)  
(Uniform distribution)

\[
\hat{Z} \approx \arg \max_Z \frac{P[I_T|Z]P[Z]}{P[I_T]} \frac{P[I_S|Z]P[Z]}{P[I_S]}
\]

\[
\hat{Z} = \arg \max_Z P[Z|I_T, I_S] \approx \arg \max_Z P[Z|I_T]P[Z|I_S]
\]
Probability model

- $\hat{Z}$ estimate of scene depth, wrt the reference frame of $T$ (both images and 3D)

- $\hat{Z} = \arg \max_Z P[Z|I_T, I_S] = \arg \max_Z \frac{P[I_T, I_S|Z]P[Z]}{P[I_T, I_S]}$ (Bayes Rule)

- $\hat{Z} = \arg \max_Z \frac{P[I_T, I_S|Z]P[Z]}{C}$ (Den. indep. from max argument)

- $\hat{Z} = \arg \max_Z \frac{P[I_T, I_S|Z]P[Z]P[Z]}{P[I_T]P[I_S]}$ (Uniform distribution)

- $\hat{Z} \approx \arg \max_Z \frac{P[I_T|Z]P[Z]}{P[I_T]} \frac{P[I_S|Z]P[Z]}{P[I_S]}$ (Hypothesis)

- $\hat{Z} = \arg \max_Z P[Z|I_T, I_S] \approx \arg \max_Z P[Z|I_T]P[Z|I_S]$
Probability model

- \( \hat{Z} \) estimate of scene depth, wrt the reference frame of \( T \) (both images and 3D)

- \( \hat{Z} = \arg \max_Z P[Z|I_T, I_S] = \arg \max_Z \frac{P[I_T, I_S|Z]P[Z]}{P[I_T, I_S]} \)  
  (Bayes Rule)

- \( \hat{Z} = \arg \max_Z \frac{P[I_T, I_S|Z]P[Z]}{C} \)  
  (Den. indep. from max argument)

- \( \hat{Z} = \arg \max_Z \frac{P[I_T, I_S|Z]P[Z]P[Z]}{P[I_T]P[I_S]} \)  
  (Uniform distribution)

- \( \hat{Z} \approx \arg \max_Z \frac{P[I_T|Z]P[Z]}{P[I_T]} \cdot \frac{P[I_S|Z]P[Z]}{P[I_S]} \)  
  (Hypothesis)

- \( \hat{Z} = \arg \max_Z P[Z|I_T, I_S] \approx \arg \max_Z P[Z|I_T]P[Z|I_S] \)  
  (Final result)
Probability model

- \( \hat{Z} \) estimate of scene depth, wrt the reference frame of \( T \) (both images and 3D)

\[
\hat{Z} = \arg\max_Z P[Z|I_T, I_S] = \arg\max_Z \frac{P[I_T, I_S|Z]P[Z]}{P[I_T, I_S]} \quad \text{(Bayes Rule)}
\]

- \( \hat{Z} \) = \arg\max_Z \frac{P[I_T, I_S|Z]P[Z]}{C}

- \( \hat{Z} = \arg\max_Z \frac{P[I_T, I_S|Z]P[Z]P[Z]}{P[I_T]P[I_S]} \quad \text{(Uniform distribution)}
\]

- \( \hat{Z} \approx \arg\max_Z \frac{P[I_T|Z]P[Z]}{P[I_T]} \cdot \frac{P[I_S|Z]P[Z]}{P[I_S]} \quad \text{(Hypothesis)}
\]

- \( \hat{Z} = \arg\max_Z P[Z|I_T, I_S] \approx \arg\max_Z P[Z|I_T]P[Z|I_S] \quad \text{(Final result)}
\]

- Assumption: pixel-by-pixel independency for \( Z(p) \)

lunedì 7 febbraio 2011
Visually

For each of these samples calculate:

\[ P[Z|I_S] \quad \& \quad P[Z|I_T] \]
ToF Probability (1)

- Two types of noise: [Kahlmann et al. 2008]
  - Thermal noise
    - Camera acquisition properties
    - Gaussian distribution
    - Dominant far from depth discontinuities
  - Scattering generated noise
    - Noise due to scene geometry
    - Gaussian distribution
    - Dominant near from depth discontinuities
ToF Probability (2)

- Thermal noise: far from depth discontinuities
- Gaussian distributed noise: $P[Z(p)|I_T] \sim \mathcal{N}(d, \sigma_T^2)$
  - Mean: 0 ($D_T(p) = d$)
  - Variance: function of the amplitude image in the considered pixel ($\sigma_T^2$)
    (High amplitude value $\leftrightarrow$ low variance)
ToF Probability (3)

- Near depth discontinuities: scattering generated noise
- Gaussian distributed noise $P \left[ Z(p) \mid I_T \right] \sim \mathcal{N}(d, \sigma_S^2)$
  - Mean: 0 ($D_T(p) = d$)
  - Variance: variance of the depth image in the second order neighborhood $p$ of $\sigma_S^2$:
ToF Probability (4)

- Calculate $\sigma_T^2$
- Calculate $\sigma_S^2$
- $\sigma_W^2 = \max\{\sigma_T^2, \sigma_S^2\}$
- $P[Z(p)|I_T] \sim \mathcal{N}(d, \sigma_W^2)$
- $3 - \sigma$ rule $\rightarrow [d - 3\sigma_W, d + 3\sigma_W]$ is sampled in $m$ points
Stereo Probability (1)

- Heuristic probability
  - Each sample $p_i, i = 1, \ldots, m \in [d - 3\sigma_T, d + 3\sigma_T]$ is reprojected into the stereo images $I_L$ and $I_R$
  - TAD (Absolute Difference) cost function is calculated: $C_i, i = 1, \ldots, m$
Stereo Probability (1)

- Heuristic probability
  - Each sample $p_i, i = 1, ..., m \in [d - 3\sigma_T, d + 3\sigma_T]$ is reprojected into the stereo images $I_L$ and $I_R$
  - TAD (Absolute Difference) cost function is calculated: $C_i, i = 1, ..., m$
Stereo Probability (1)

- Heuristic probability
  - Each sample $p_i, i = 1, \ldots, m \in [d - 3\sigma_T, d + 3\sigma_T]$ is reprojected into the stereo images $I_L$ and $I_R$
  
- TAD (Absolute Difference) cost function is calculated: $C_i, i = 1, \ldots, m$
Stereo Probability (2)

\[ C_i = \min \left\{ \sum_{W} \left| W_R^R - W_R^L \right| + \left| W_L^G - W_R^G \right| + \left| W_L^B - W_R^B \right|, T_h \right\} \]
Stereo Probability (3)

- Probability:

\[ P[Z(p) = z(p_i)|I_S] \propto \exp - \frac{C_i(p)}{\sigma_i} \]

- where \( \sigma_I \) is the noise standard deviation in \( \{I_L, I_R\} \) (to be estimated, classical value 0.05)

- [Sun et al. 2003]
Combined Model

- For each pixel $p$ the ToF and the stereo probabilities are calculated and multiplied
- The maximum is selected
Outline

- Background (Stereo + ToF)
- Preliminary step: system calibration
- Problem definition and motivation
- Fusion Algorithm
- Results
- Conclusions
Synthetic Setup
Numerical Results (1)

- Fixed noise in $I_s$, increasing noise in $D_T$
Numerical Results (2)

- Fixed noise in $D_T$, increasing noise in $I_S$

![Graph showingToF, Stereo, Fusion Algorithm Performances]
Middlebury Dataset

- Classical stereo dataset

Noiseless

Noisy image before the fusion algorithm

Noisy image after the fusion algorithm
Real Scene
Outline

• Background (Stereo + ToF)
• Preliminary step: system calibration
• Problem definition and motivation
• Fusion Algorithm
• Results
• Conclusions
Conclusion

- ToF and stereo data fusion algorithm
- AM model for ToF camera
- Probabilistic framework
- Interesting applications
  - Autonomous navigation
  - Gaming (Microsoft Kinect)
  - Free view-point video
Carlo Dal Mutto

dalmutto@dei.unipd.it
www.dei.unipd.it/~dalmutto
twitter: carlodalmutto
Multimedia Technology and Telecommunication Laboratory

lttm.dei.unipd.it