Accurate 3D Reconstruction by Fusion of ToF and Stereo Data

C. Dal Mutto, P. Zanuttigh, G.M. Cortelazzo
Department of Information Engineering
University of Padova
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
Stereo (I)

- Input: 2 rectified stereo images

\[ [u, v] \]
Stereo (I)

- Input: 2 rectified stereo images
Stereo (2)

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

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

- Doesn’t work with textureless scenes
Stereo (3)

- Doesn’t work with textureless scenes
ToF Camera (I)

- 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 (I)

- 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 (I)

- 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)

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

- Receiver matrix (176x144)

\[
\Delta \varphi \times \frac{1}{2} \frac{c}{2\pi f}
\]

\[r(t) \downarrow 120[MHz] \quad r(nT) \quad \Delta \varphi \quad d\]
ToF Camera (3)

- Output:
  - Distance Image: $D_T$
  - Amplitude Image: $A_T$
Fusion Algorithm Input

- Stereo pair + 1 ToF camera

$$I_T = \{A_T, D_T\}$$
Fusion Algorithm Input

- Stereo pair + 1 ToF camera

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

- Stereo pair + 1 ToF camera

\[ I_L, I_R \]

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

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

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

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

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

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

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

martedì 22 giugno 2010
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 P[I_T, I_S|Z]P[Z] 
\]  
(Den. indep. from max argument)

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

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

- \( \hat{Z} = \arg\max_Z \frac{P[I_T, I_S \mid Z]P[Z]}{C} \)  
  (Density independent from max argument)

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

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

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

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

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

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

- Assumption: pixel-by-pixel independency $Z(p)$
ToF Probability (1)

- Gaussian distributed noise: \( P[Z(p)|I_T] \sim \mathcal{N}(d, \sigma_T^2) \)
  - Mean: 0 \( (I_T(p) = d) \)
  - Variance: function of the amplitude image in the considered pixel \( (\sigma_T^2) \)
  - 3 $\sigma$ Rule $\rightarrow$ only interval: \([d - 3\sigma_T, d + 3\sigma_T]\) $\rightarrow$ \(m\) points

![Histogram of 100 depth measurements]
Stereo Probability

- 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 $I_R$

- AD (Absolute Difference) cost function is calculated: $C_i, i = 1, \ldots, m$

- Probability: $P[Z(p) = z(p_i)|I_S] \propto \exp\left(-\frac{C_i(p)}{\sigma_i}\right)$, where $\sigma_I$ is the noise standard deviation in $\{I_L, I_R\}$
Stereo Probability

- 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 $I_R$
  - AD (Absolute Difference) cost function is calculated: $C_i, i = 1, \ldots, m$
  - 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 $\{I_L, I_R\}$
Stereo Probability

- 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 $I_R$
  - AD (Absolute Difference) cost function is calculated: $C_i, i = 1, \ldots, m$
  - Probability: $P[Z(p) = z(p_i)|I_S] \propto \exp\left(-\frac{C_i(p)}{\sigma_i}\right)$, where $\sigma_I$ is the noise standard deviation in $\{I_L, I_R\}$
Combined Model

- For each pixel \( p \) the ToF and the stereo probabilities are calculated and multiplied
- The maximum is selected
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 $I_S$
Middlebury Dataset

- Classical stereo dataset

![Noiseless image](image1)

![Noisy image before fusion algorithm](image2)

![Noisy image after fusion algorithm](image3)
Real Scene
Conclusion

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

dalmutto@dei.unipd.it

http://lttm.dei.unipd.it