Video Based 3D Imaging:
Milestones and The Horizon

June 20, 2014

Guna Seetharaman Ph.D.
Information Intelligence and Analysis Division
Information Directorate, Rome, NY
Gunasekaran.seetharaman@us.af.mil
315-330-2414

Prof. K. Palaniappan, University of Missouri at Columbia
Dr Steve Suddarth, (Ret Col USAF), Transparent Sky LLC, ABQ

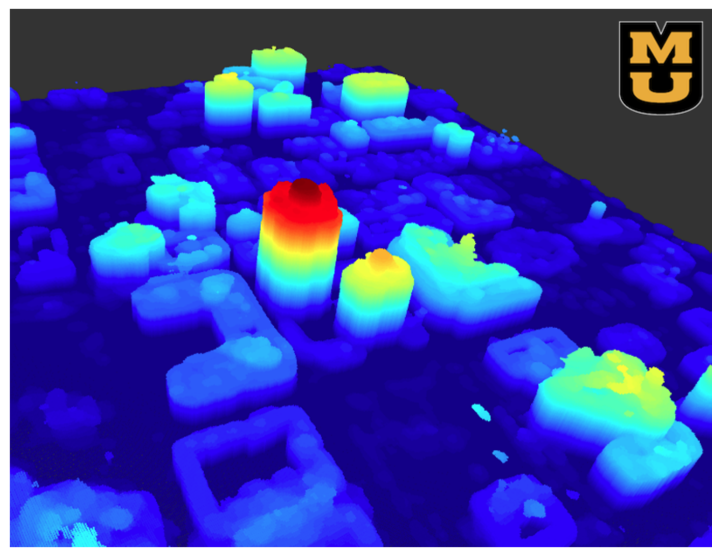
Integrity ★ Service ★ Excellence

Historical Milestones

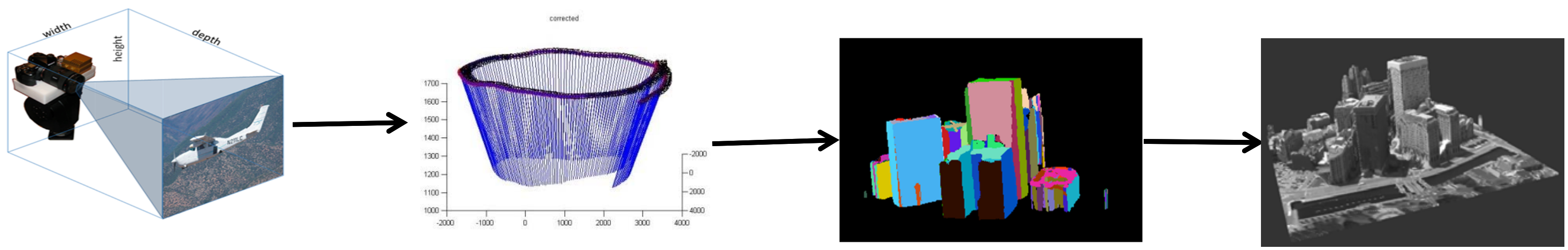
- Physically measure 3D: Surveying and Spherometer
- Stereo and triangulation
- Integral Photography
- Sonar and RADAR SAR --- Synthetic Holograms
- Point and Shoot LIDAR such NASA 60's and SICK-gener 90s
- Steven White scanner – Laser line and line induced contour 80s
- Projected patterns – Stockman et.al and 1988
- DeMorie Patterns – Montreal Group 1986
- Kanade's distributed angle of arrival sensor 1988
- Hybrid Range intensity sensing – Seetharaman 91 and Medioni 95
- Multiview Imaging – Zisserman et.al - Point clouds
- Canesta - Kinect – 2003 rapid point patterns
- Bundle Adjustment – Triggs et.al. 2005; most recent rapid growth

Current State of the Art: VB3D Reconstruction

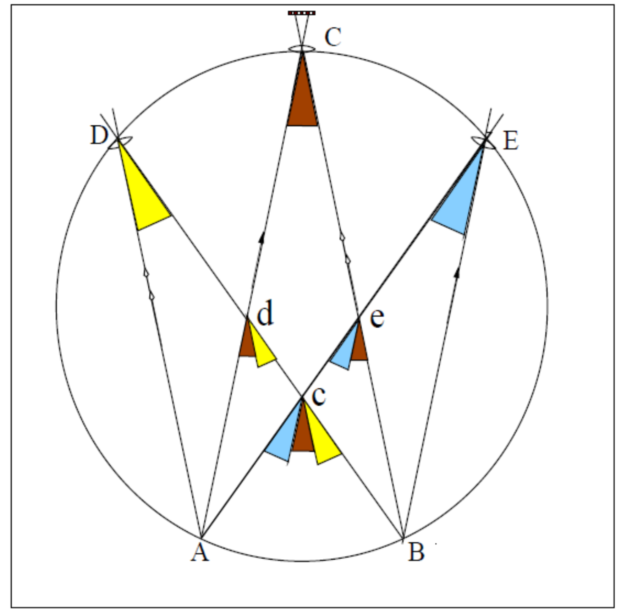
HD motion imagery of ABQ, NM.
Courtesy: Transparent Sky LLC, NM



VB3D: Video Based Point Cloud Creation



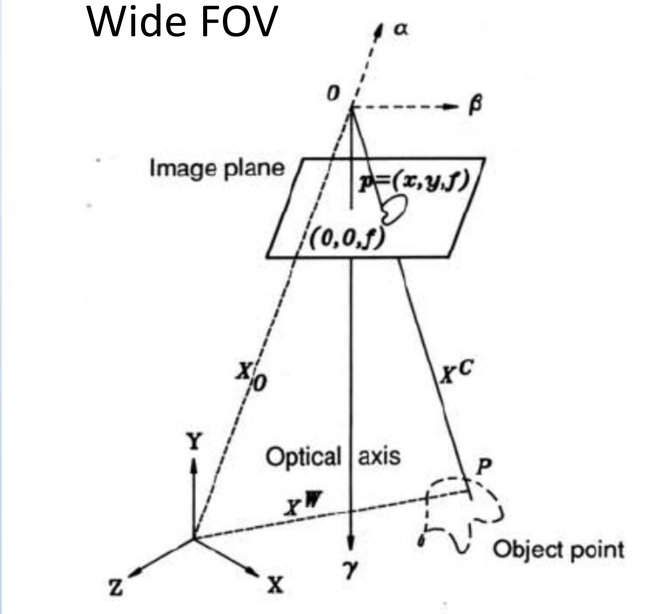
Persistent Sensors vary in SWAP and CONOPS



	Realtime Consumption	Forensic Consumption
Image Registration	Geo-registration before communication	Exhaustive registration between retinal images, after communication
Exploitation	Tactical, visual process (Humans in the Loop) and forensic	Context, location-specific history, enterprise scale forensic processes
3-D models	a-priori models via DTED	3D is extracted; but, long-term learning is feasible

Two imaging models

Large world
Wide FOV



$$x = \frac{fX^O}{f + Z_0 + Z^O}$$

$$y = \frac{fY^O}{f + Z_0 + Z^O}$$

Epipolar line equation holds.

Vanishing points are unique for each bundle of parallel lines.

Image of 3-D points on a plane all obey a single quadratic relationship across time.

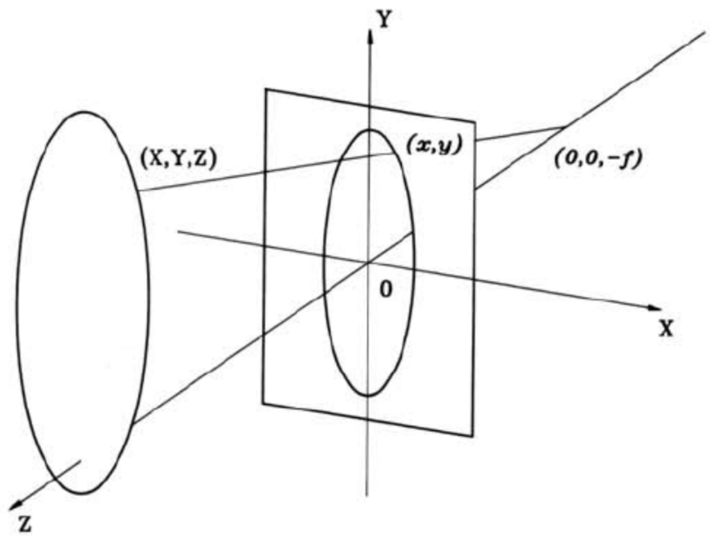
Non Euclidean imaging!

Challenge: Even a simple cube has six faces; each visible partition in the image manifests as a different transformation – parameters TBE.

True, even in the case of a single motion – of the cube or camera.

Two imaging models

Small world
Narrow FOV



$$x = \frac{fX^O}{f + Z_0 + Z^O} = \frac{fX^O}{(f + Z_0)}\left(1 + \frac{Z^O}{Z_0 + f}\right)^{-1},$$

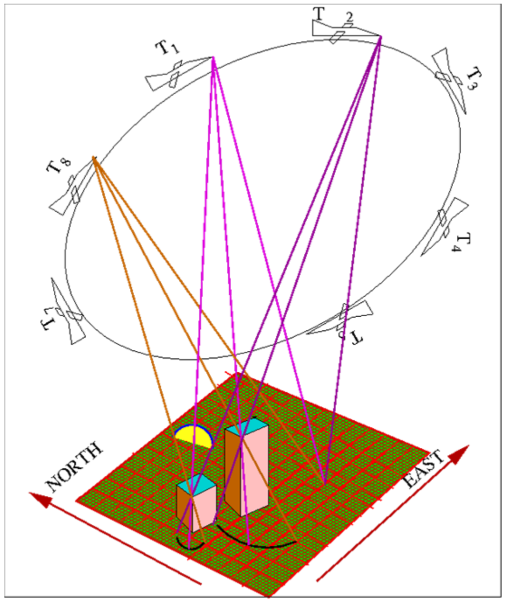
$$y = \frac{fY^O}{f + Z_0 + Z^O} = \frac{fY^O}{(f + Z_0)}\left(1 + \frac{Z^O}{Z_0 + f}\right)^{-1}.$$

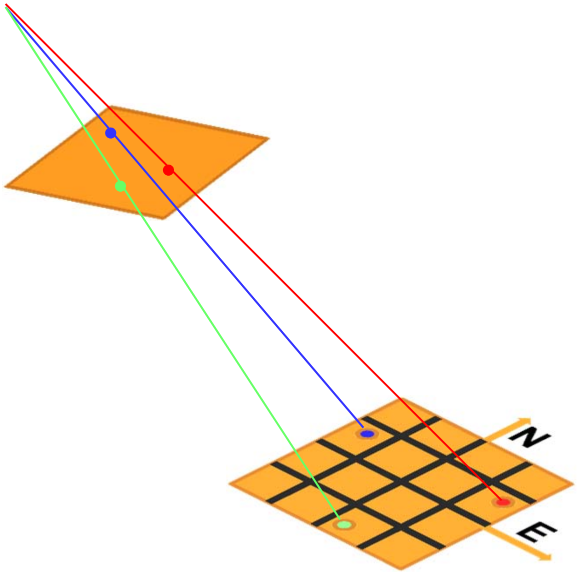
$$\begin{bmatrix} x \\ y \end{bmatrix} = \left(1 + \frac{Z_0}{f}\right)^{-1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X^O \\ Y^O \\ Z^O + Z_0 \end{bmatrix},$$

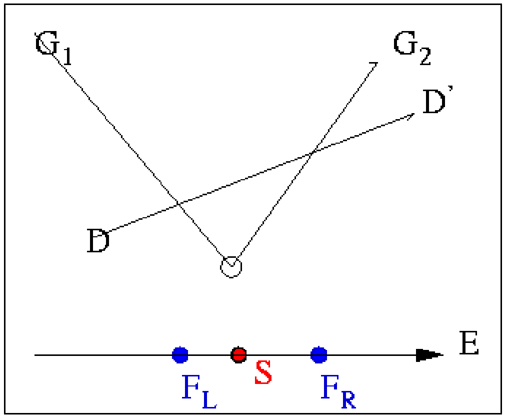
$Z_0 \to \infty$, implies far-away objects
$f \to \infty$, implies large-focal length.
Telephoto imaging

Euclidean in the small-world.
EE-s recall small signal analysis.
Piecewise linear, and bounded total distortion on the global scale.
How to locate and deal with discontinuities!?

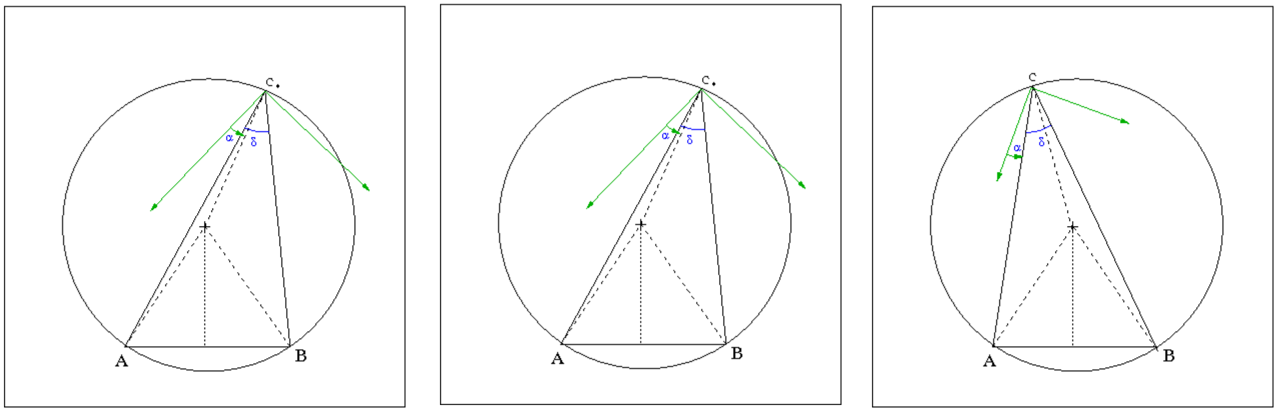
Registration: Mythical Flat Earth







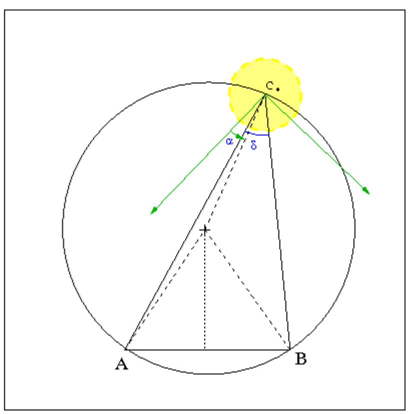
Basic Constraints on G and R from Visual Landmarks



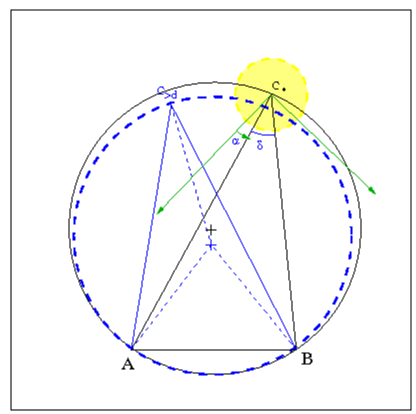
Longlook-97
IJDSN - 2007

Triplets – 198x

Given, **a** and **b** in the image. A and B are known. Camera position is constrained onto a unique circle.



Fusing GPS helps narrow the uncertainty



Imprecision in the pixel positions adds to radial uncertainty.

Features of MU BA

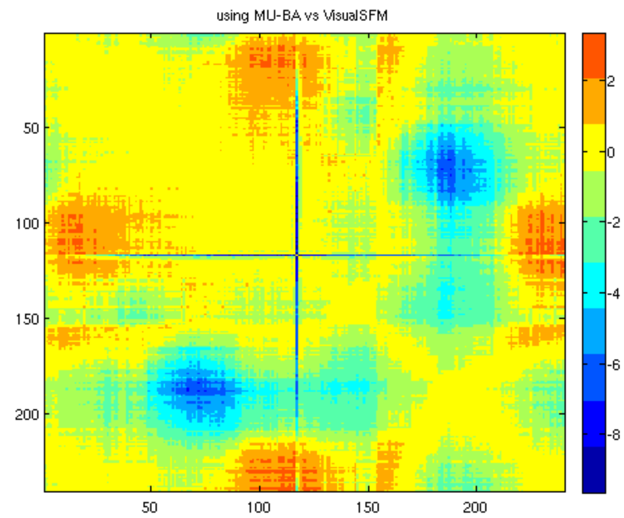
- Sequential feature tracking tailored for WAMI collection
- No RANSAC, No Kalman or other filtering
- No Funda $\frac{1}{2}\sum_{i=1}\rho_i\left(\left\|f_i\left(x_{i_1},\ldots,x_{i_k}\right)\right\|^2\right)$ stimation
- Robust weighted reprojection error loss function

- **Very fast: 6.5 minutes vs 6-8 hours with VisualSfM –combination of GPU, C++, Matlab**

BA Albuquerque WAMI Dataset

- Number of cameras: 215
- Image size: 6600 x 4400
- Number of feature observations: 668,000
- Number of 3D points: 141,559
- Size of Jacobian matrix: 1,336,000 x 425,968
- Total time: Less than 6.5 minutes (4 min Matlab)

EEE of MU-BA vs VisualSFM



Errors (m)	Mean	Std Deviation
MU-BA	1.8276	1.0495
VisualSFM	2.4866	1.5856

Percentage of MU-BA overperforming VisualSFM is 66.29

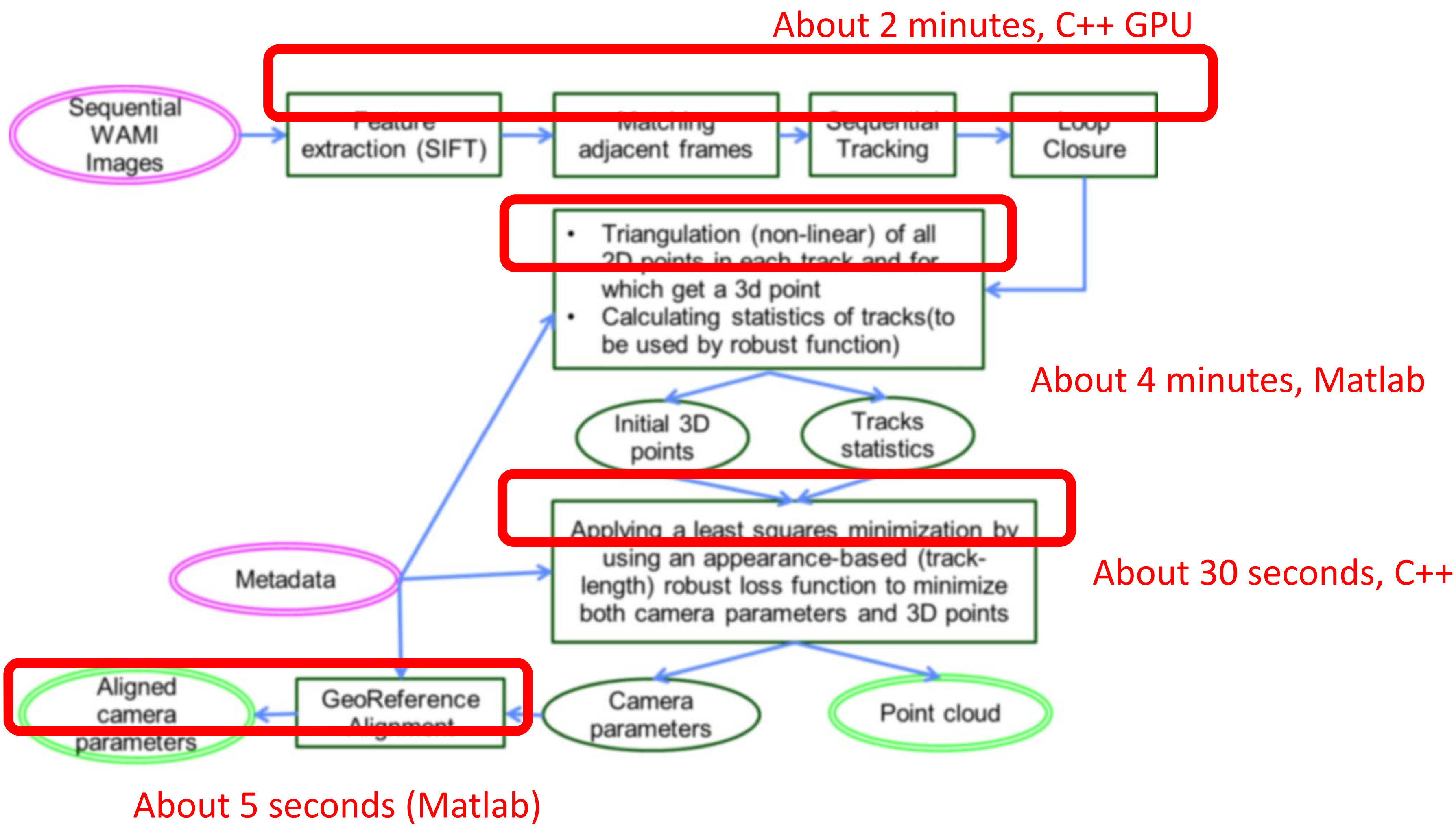
Percentage of VisualSFM overperforming MU-BA is 33.30

Percentage that both VisualSFM and MU-BA have no error 0.42

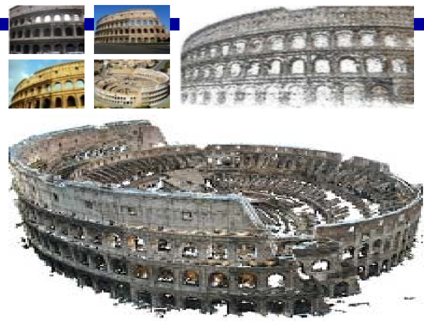
Highlights

- MU BA pipeline uses fewer points (less computation) compared to other BA such as VisualSFM, but yields better results
- For WAMI-BA, using extra matches does not necessarily result in improved accuracy
- Much faster than VisualSFM which takes about 5-6 hour on Albuquerque dataset
- MU BA does not use thresholding or RANSAC to eliminate outliers, but a robust loss function

BA on Albuquerque Dataset

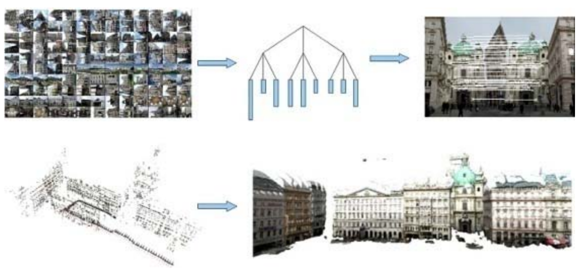


State of the Art – 3D Reconstruction Techniques

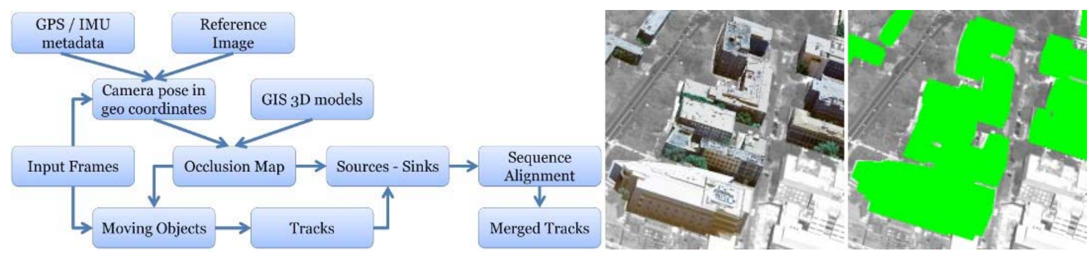


Large unorganized collection of ground based photos
Snavely (Cornell), Seitz (Univ. Washington) & Szeliski (Microsoft) ICCV, 2009

Volumetric appearance modeling , Pollard & Mundy, CVPR09
Comprehensive 3-D change detection using volumetric appearance modelling, B. Pollard, PhD Thesis, Brown Univ. 2009
A continuous probabilistic scene model for aerial imagery, D. Crispell, PhD Thesis, Brown Univ. 2010



Large structured collection of ground based photos,
Bischof (TU Graz) CVIU, 2012



Using 3D scene structure to improve tracking , Prokaj & Medioni (USC) CVPR11
Accurate image registration through 3D reconstruction, Y. Lin, PhD Thesis,, USC, 2010

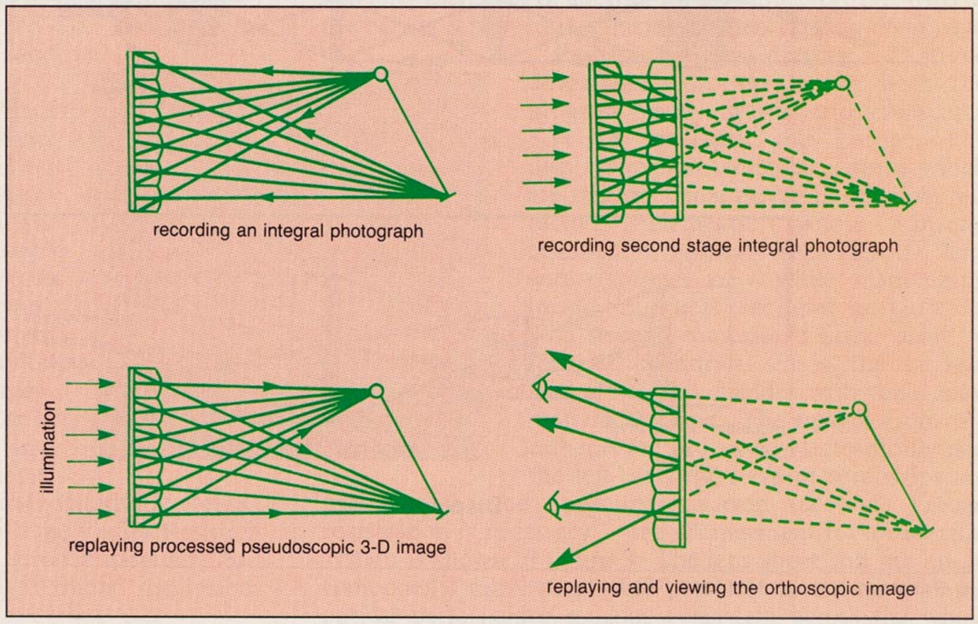
Limitations of Current Methods:

- Probabilistic voting vs feature-constrained voting
- Accuracy of point set models and planar models is not adequate
- Shadow and occlusion maps need to be integrated into tracking models
- High computational overhead cost and offline process for large urban coverage

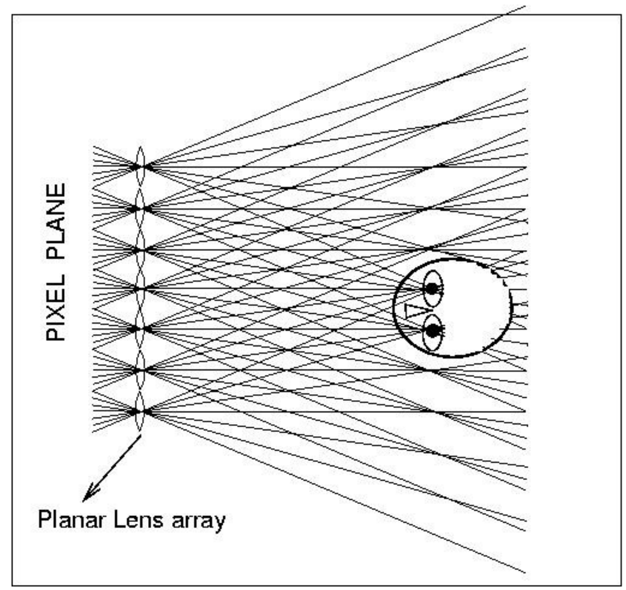
A. Zakhor, UCB

Major Milestones

3DTV - Approach 2: Integral Photography



5 Integral photography



Integral photo ‘1960s; technologic
the 90s.

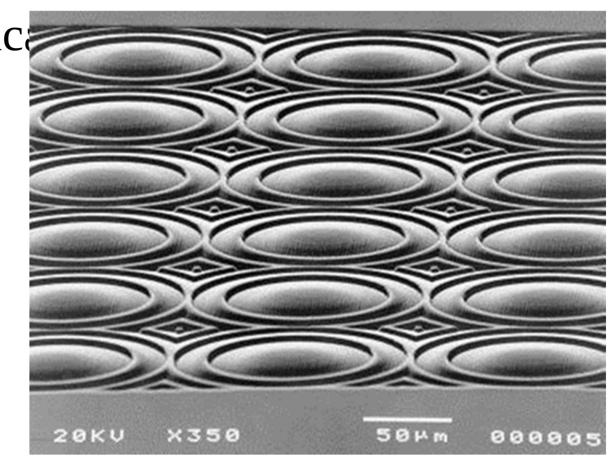


Image Pickup

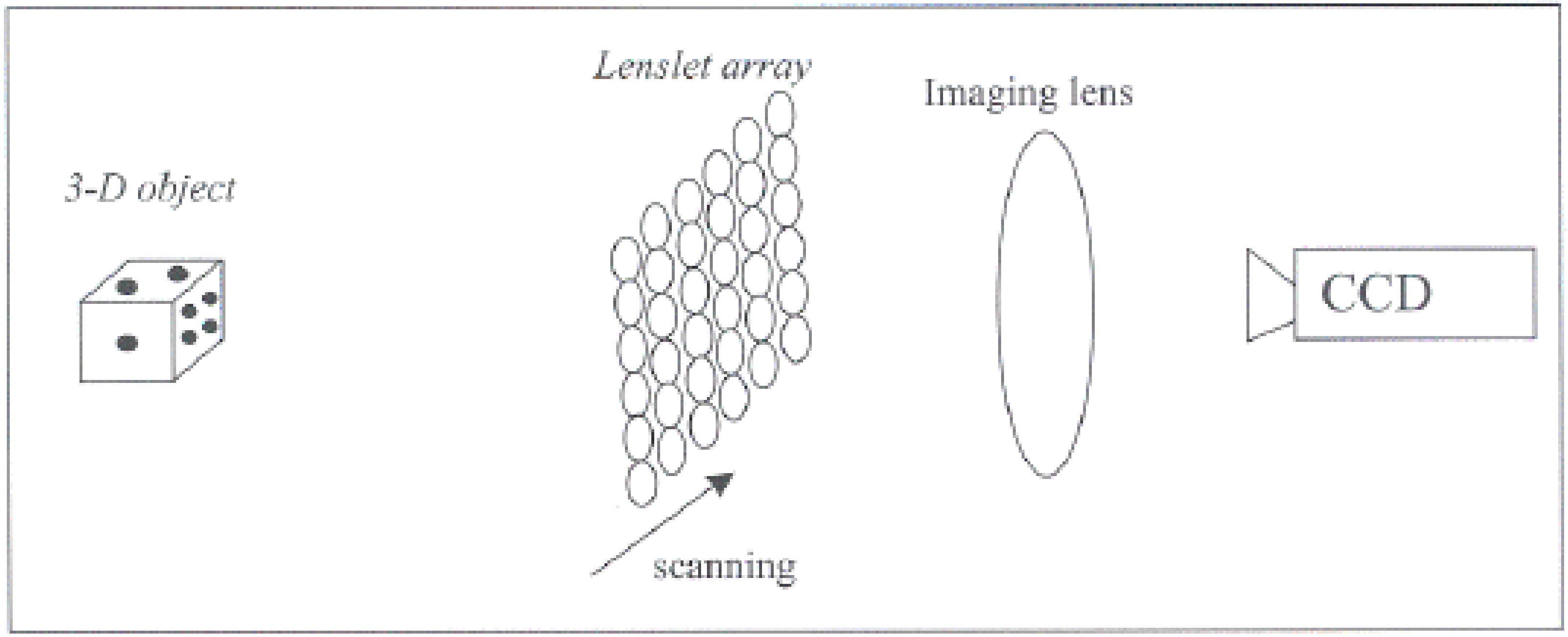
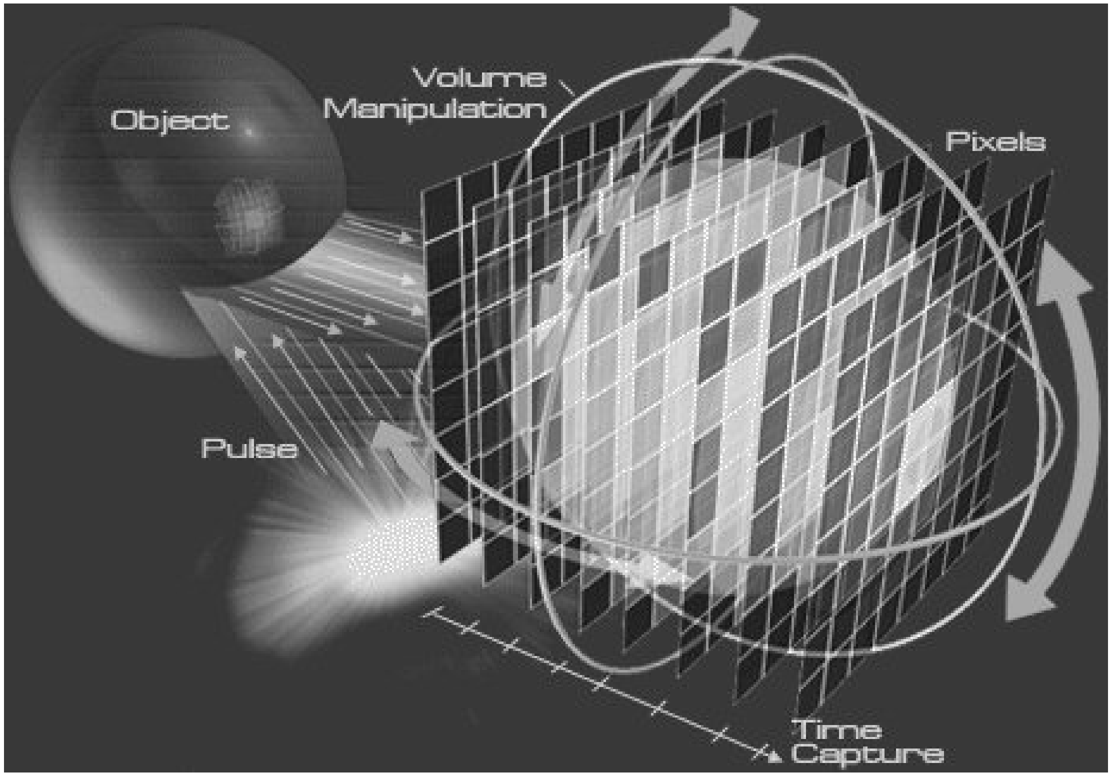
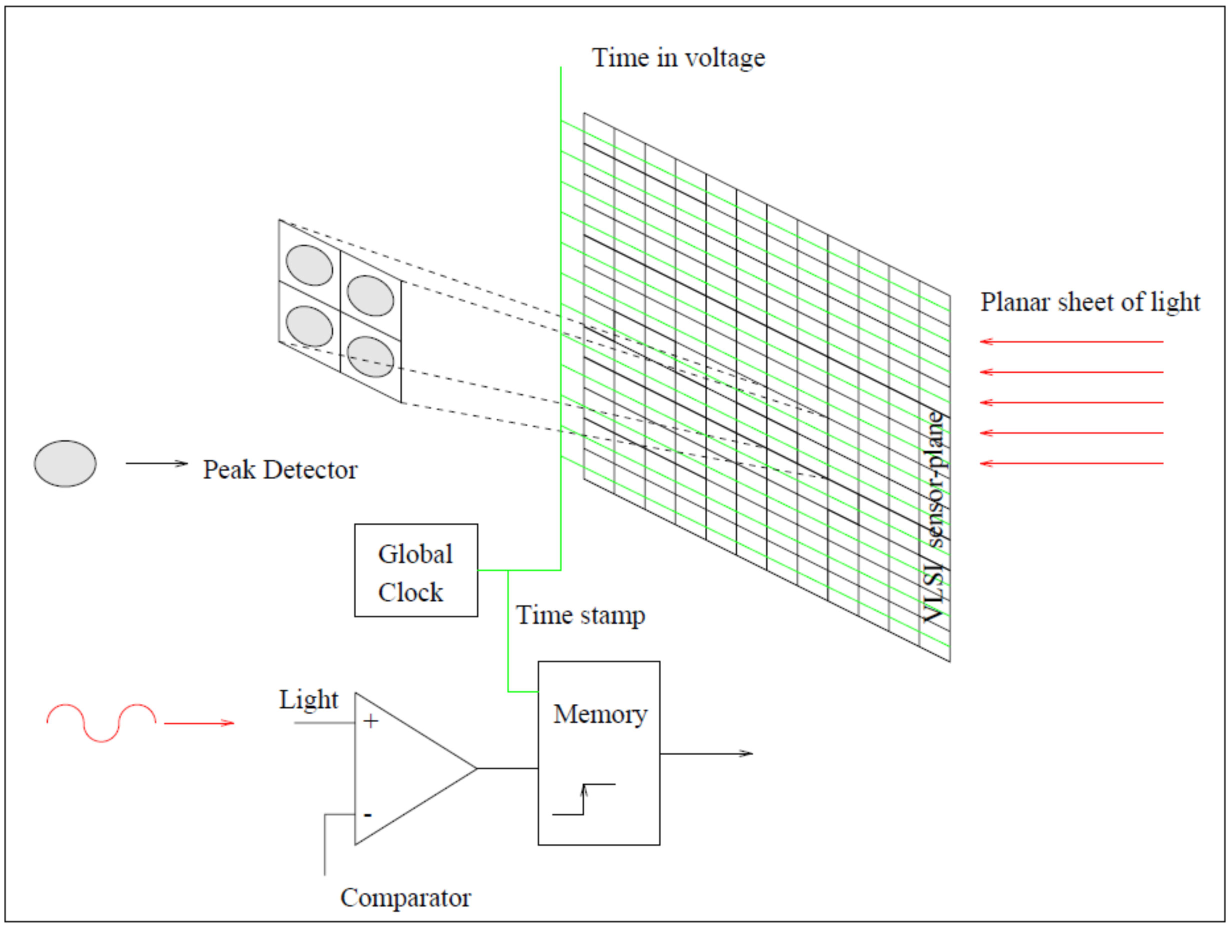
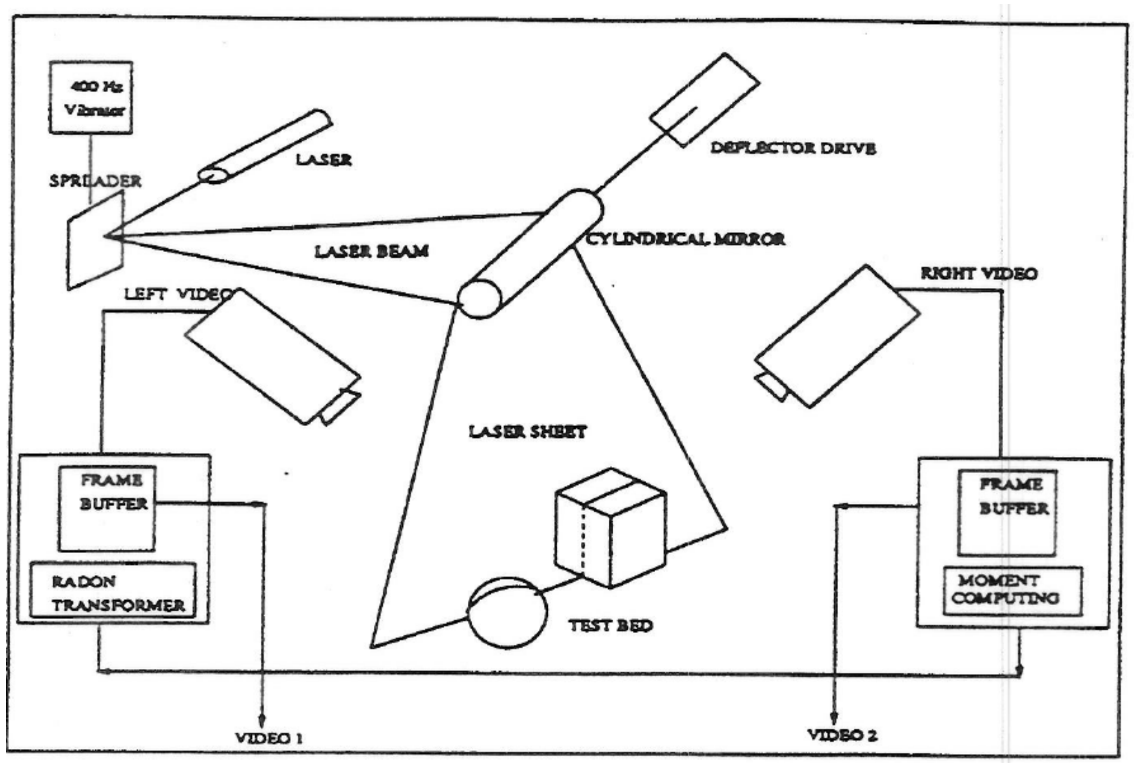


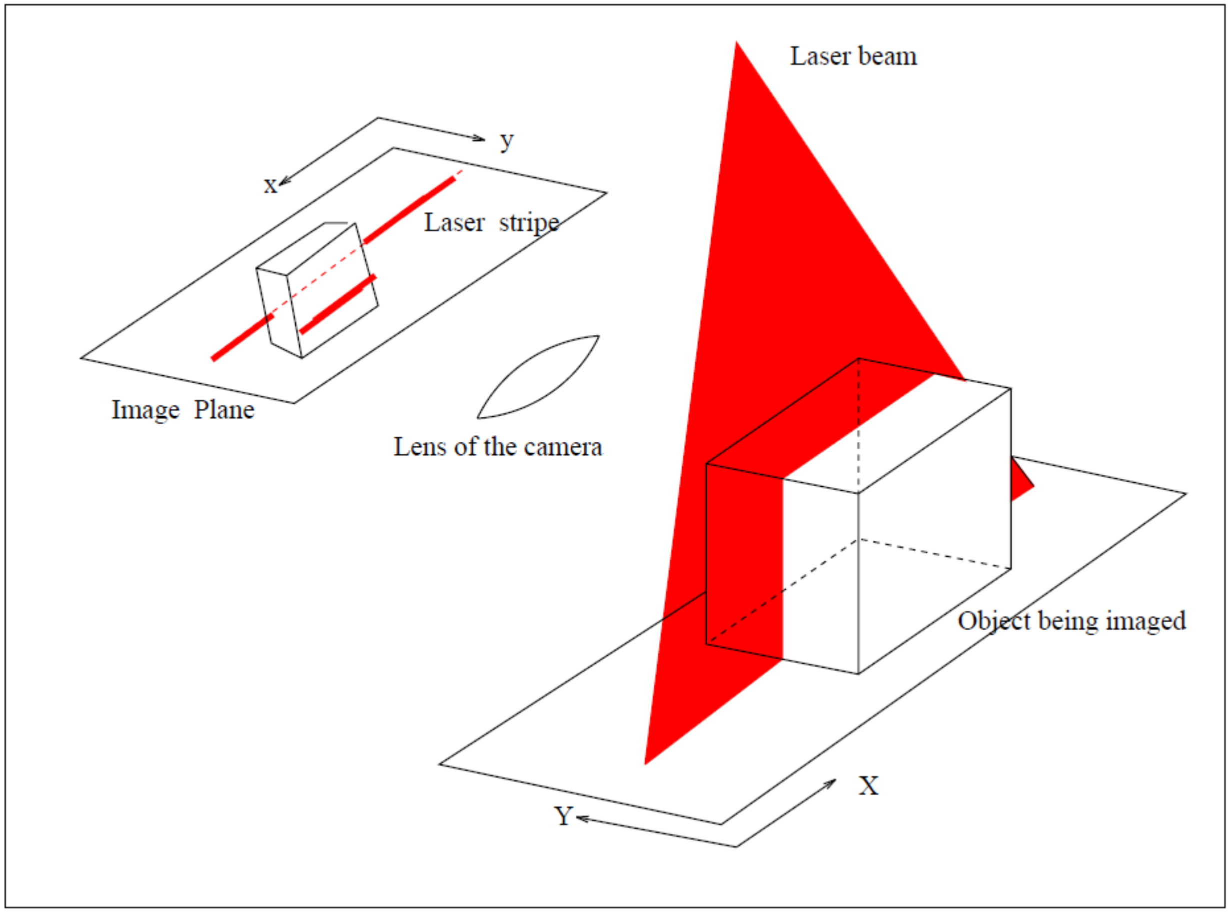
Fig. 1 Typical Integral Imaging capturing system. With CII images representing perspectives of 3-D objects are reconstructed digitally. In the proposed method the lenslet array is moved in a plane perpendicular to the optical axis.

Javidi LEOS2004

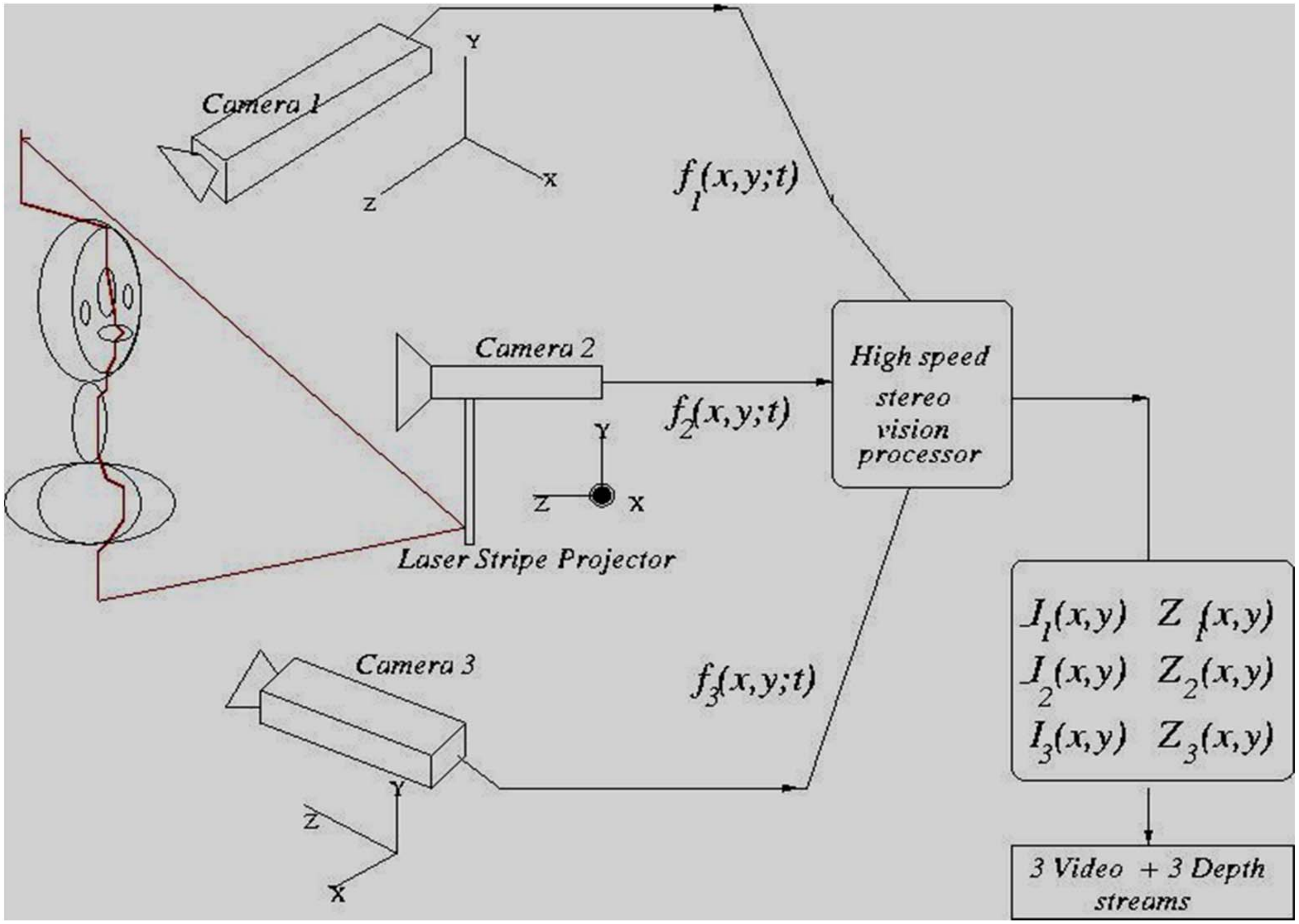

Object
Volume
Manipulation
Pixels
Pulse
Time
Capture

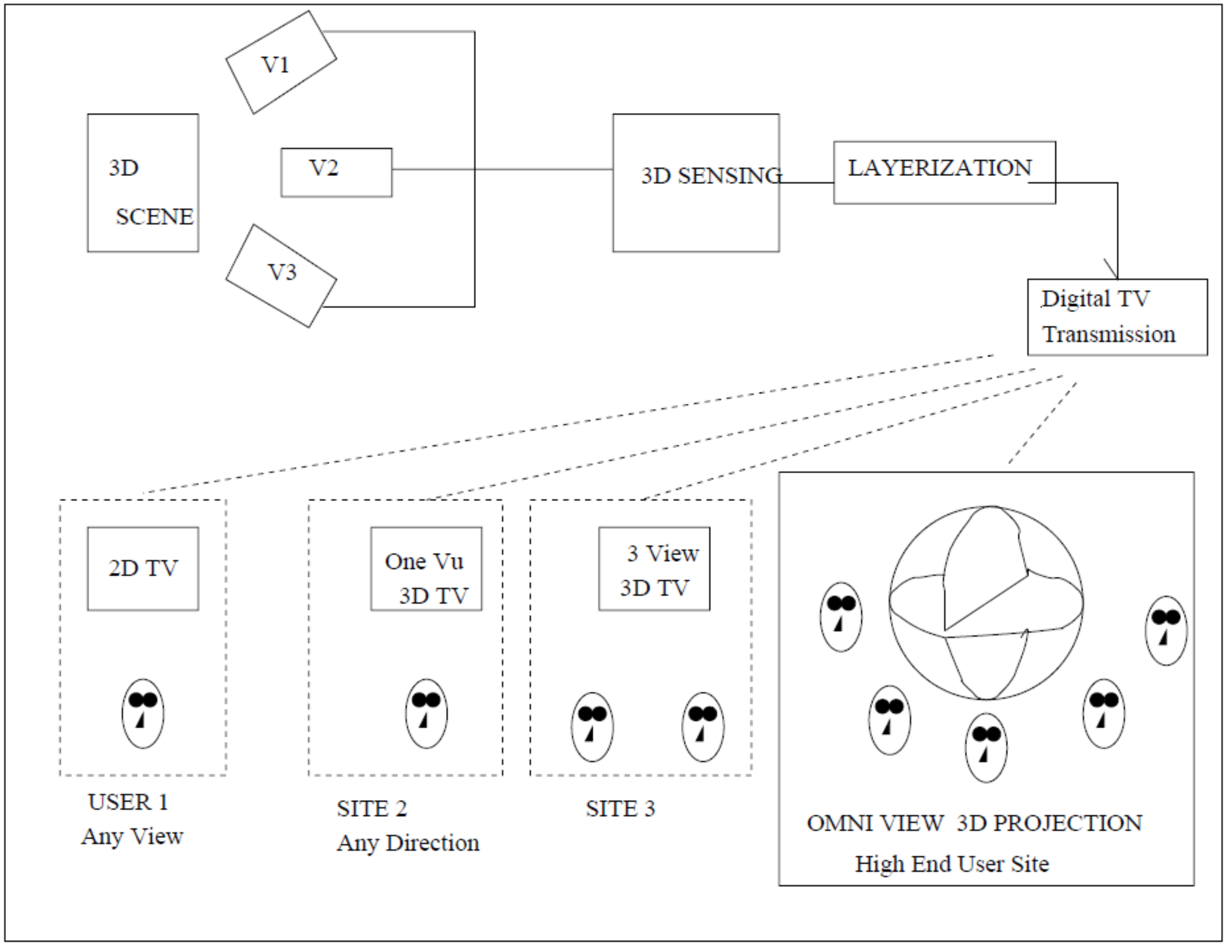

Time in voltage
Planar sheet of light
Peak Detector
Global
Clock
Time stamp
VLSI sensor-plane
Light
+
-
Memory
Comparator


400 Hz
Vibrator
SPREADER
LASER
DEFLECTOR DRIVE
LASER BEAM
CYLINDRICAL MIRROR
LEFT VIDEO
RIGHT VIDEO
LASER SHEET
FRAME
BUFFER
RADON
TRANSFORMER
FRAME
BUFFER
MOMENT
COMPUTING
TEST BED
VIDEO 1
VIDEO 2

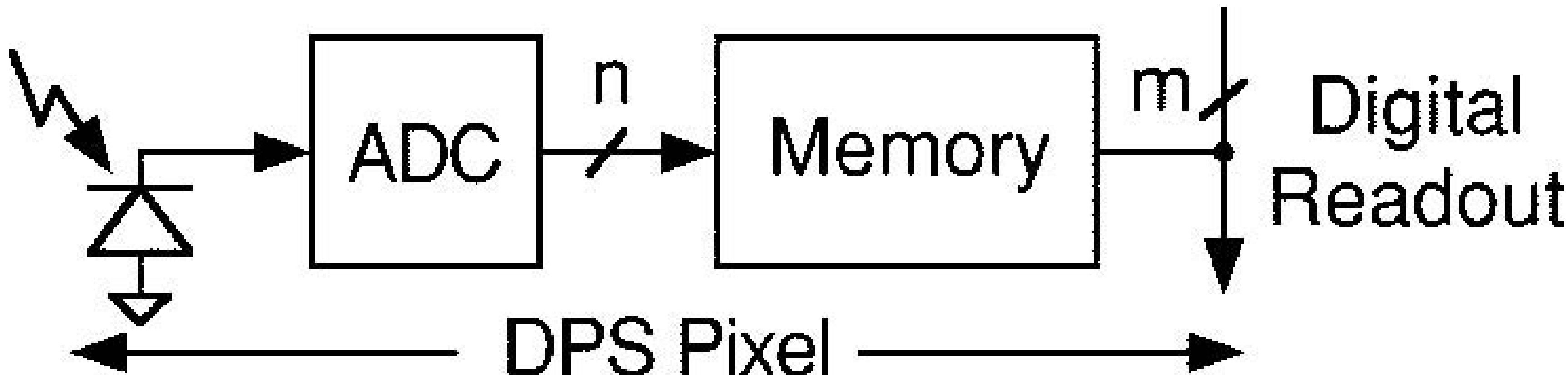

Laser beam
y
x
Laser stripe
Image Plane
Lens of the camera
Object being imaged
X
Y

Real-time multi-view I & Z sensor (Seetharaman et.al. CAMP-1991)



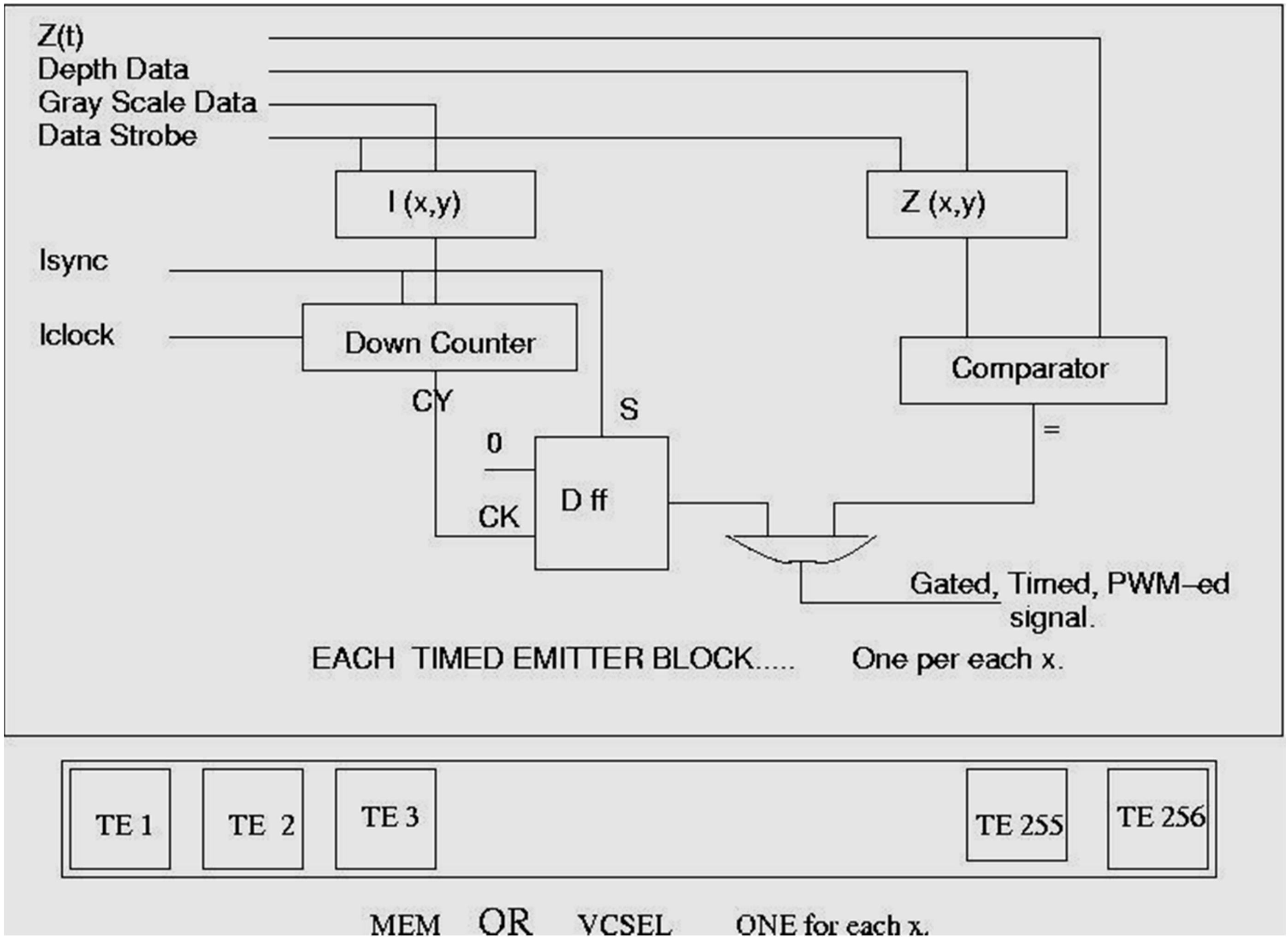

V1
3D
SCENE
V2
V3
3D SENSING
LAYERIZATION
Digital TV
Transmission
2D TV
One Vu
3D TV
3 View
3D TV
USER 1
Any View
SITE 2
Any Direction
SITE 3
OMNI VIEW 3D PROJECTION
High End User Site

10000 FPS CMOS Pixels



IEEE JSSC 2001. El Gamal. Stanford U.

Parallel indexed Timed Emitters (DCV02)



ALL PASSIVE – NON EMISSIVE SENSING

LA Results March 2014