Registration Methods for Augmented Reality

Uma Mudenagudi

Department of Computer Science Engineering
Indian Institute of Technology
Delhi

OUTLINE

- Augmented Reality Overview.
- AR applications.
- Typical AR system.
- Brief Literature review.
- Algorithms of AR system.
- Reference.

AR overview

- Milgram et al(1994) describes the relation of AR and VR.

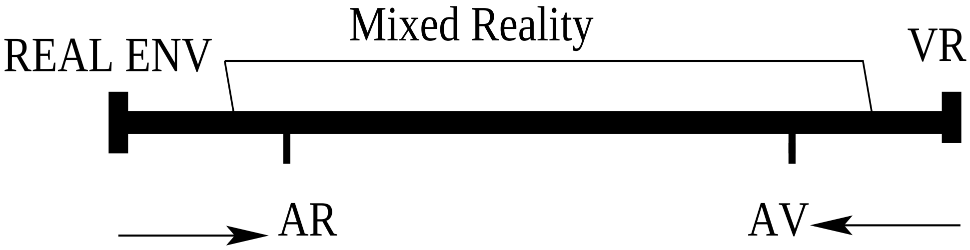

Figure 1: Relation between AR and VR

- AR lies near the real world end of line.

- Very difficult to produce real time photorealistic virtual scene.
- Different Display techniques used by AR system
- AR is not just superimposition of graphics object into real scene.

AR applications

- Medical field: Pre-operative imaging studies and avoiding usage of typical stereotactic frame.
- Entertainment:Simple form of AR system is weather report, Movie special effects and advertisement.
- Military training.
- Engineering Design.
- Robotics and Telerobotics.
- Manufacturing, maintenance and repair.
- Consumer design.

Typical AR system

- **AR SYS = Computer vision + Computer graphics + User interfaces.**

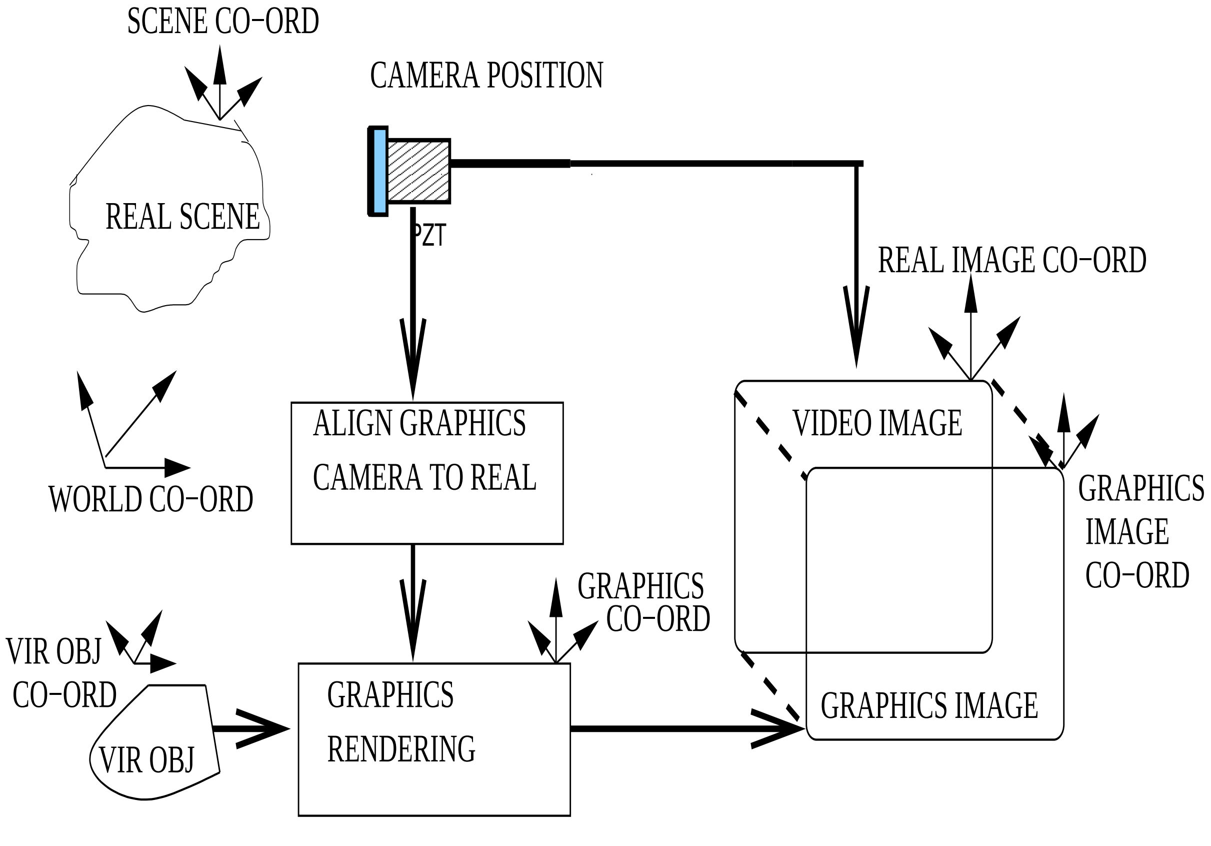

Figure 2: Typical AR System

- Camera performs the perspective projection of 3D world onto 2D image plane.

- The parameters of camera determine exactly what is projected.
- The virtual image is generated with the std computer garphics system.
- The virtual objects are modeled in an object reference frame:which requires exact imaging of real scene so that it can render the objects correctly.
- Video image and computer generated object with same view as that of the scene are merged to form augmented reality image.

Research

Research in AR center around two aspects of the problem.

- Methods to register two distinct sets of images and keep them registred in real time(tracking and registration).
- Second direction of research is in display technology for merging the two images.

Performance Issues

- Update rate for generating augmented image.
- Accuracy of registration: Noise and delay in the system.

Display Technologies

- Monitor based viewing of augmented scene:'window on the world'.
- Head mounted Displays:
  - Video see-through AR display.
  - Optical see-through AR display.

Brief Literature review

- Simon, Berger et.al have presented different vision based registration methods.
- Simon,Zisserman et.al have presented real time registration method using planar structure in the scene.
- Kanade et.al presented Z-keying method of merging images in real time.

- Kanade et.al presented probabilistic registration of 3D medical images.

Registration Methods

- AR sys loop:

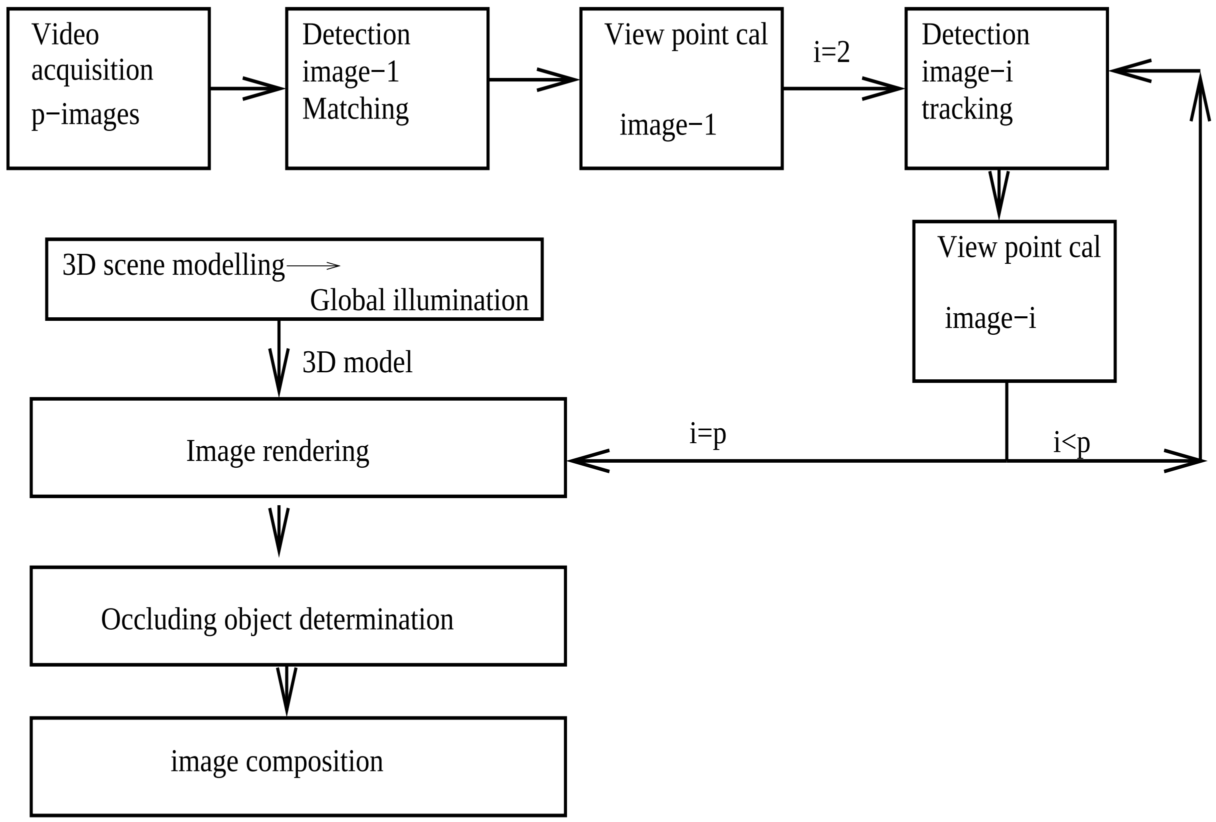

Figure 3: Typical AR loop structure

Registration method by G.Simon et.al

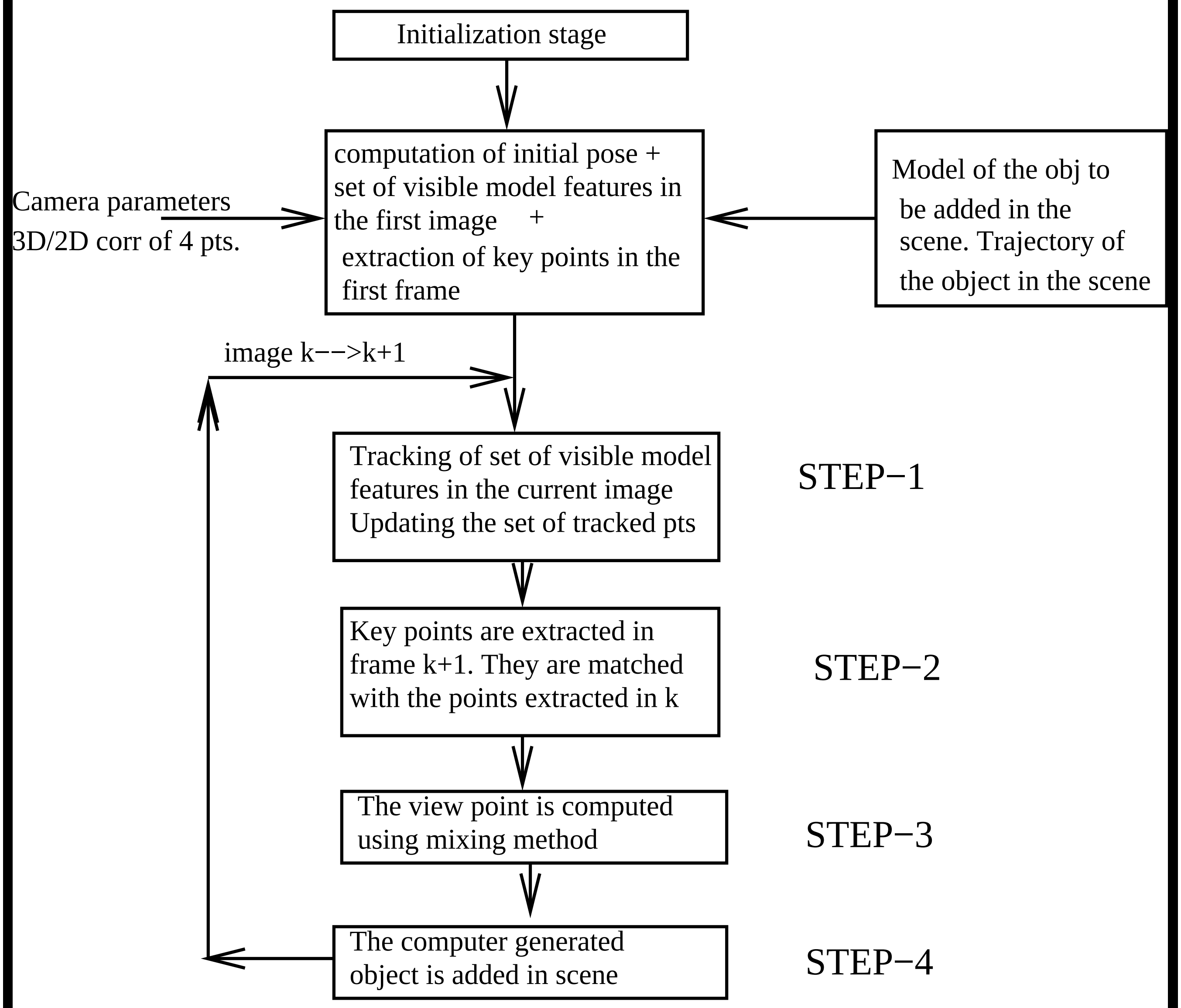

Figure 4: Registration method proposed by G.Simon et.al

Initialization

- System is initialized with the known camera parameters.
- User specified set of four 2D pts along with their 3D counterparts.
- The 2D features corresponding to the visible model are automatically determined using a two stage robust statistical method proposed by G.Simon and Berger.
- Model of the object to be added with its trajectory.
- Once initialized it follows four step loop:
  - s1:Tracking assessment and pose computation(Using RPC algo).
  - s2:Accounting for 2D/2D correspondence.
  - s3:Viewpoint calculation using mixing method.
  - s4:Addition of computer generated obj.

s1:RPC algorithm

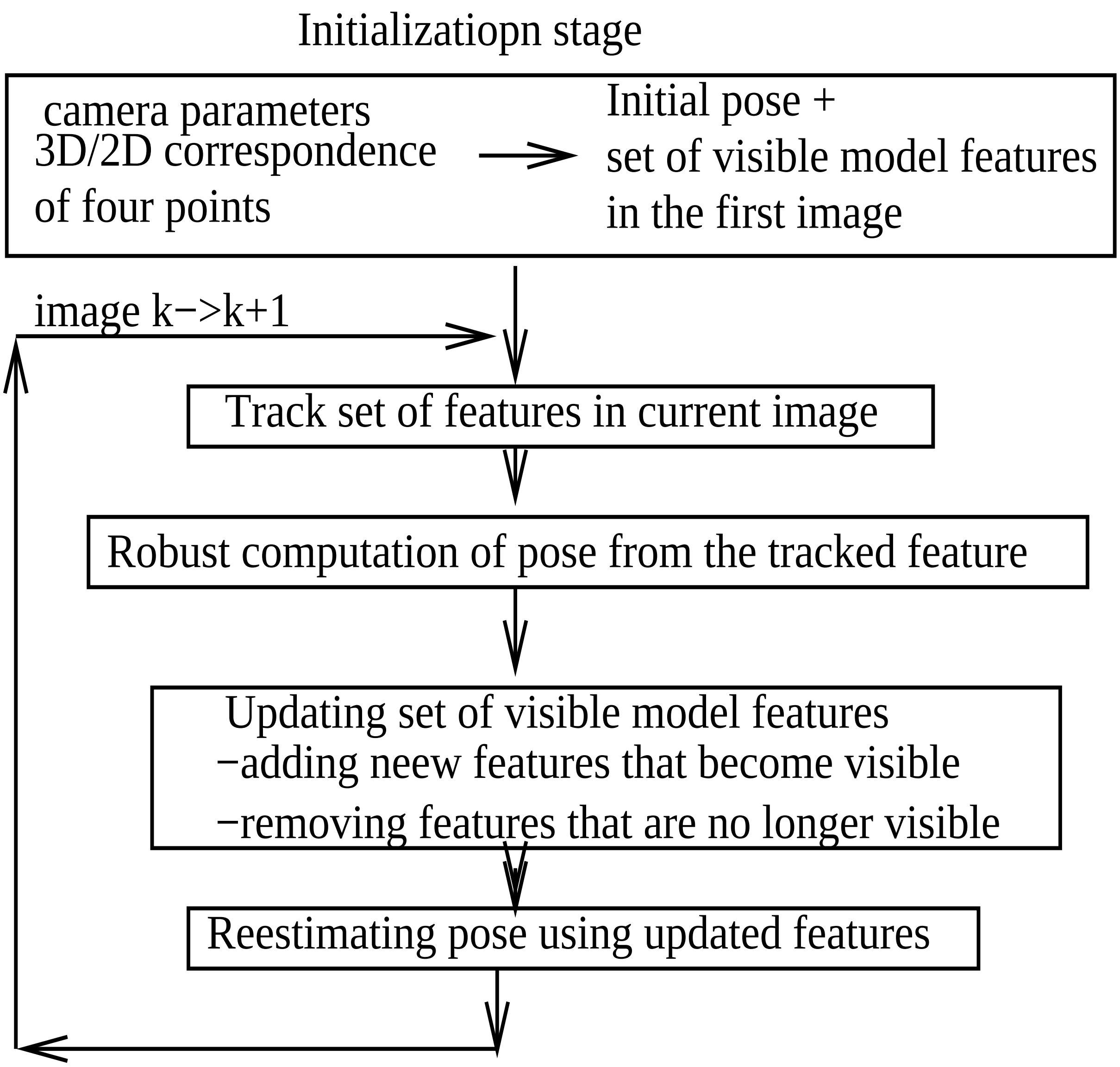

Figure 5: RPC algorithm proposed by G.Simon et.al

RPC algorithm

- Tracking:The set of features are tracked in the current image using Berger's tracking algorithm.
- Computing pose:Iterative closest point algorithm(ICP algo).
- Updating the visible model features:
- Re-estimation of the pose:
- Pose computation algorithm.
  - If camera intrinsic parameters are known, then number of unknowns are six(3 for R and 3 for t).
  - C_i be a 3D curve, described by the chain of 3D points $\{M_{i,j}\}_{1 \leq j \leq l_i}$.
  - c_i be the projection of C_i in the image plane, described by the chain of 2D points $\{m_{i,j}\}_{1 \leq j \leq l_i}$, where $m_{i,j} = Proj\{\mathbf{RM_{i,j}} + \mathbf{t}\}$.

- $c_i^{'}$ be the detected curve corresponding to C_i.
- A simple solution would be to perform one stage minimization
  $$\min \sum_{ij} \rho(d_{i,j})$$
  where $d_{i,j} = Dist(m_{i,j}^{'}, c_i)$ and ρ is a positive, symmetric function.
- To reduce the errors Simon et.al proposed robust estimate using two stage process.
- Global and Local stage
- Local stage:which computes the robust residual for each feature point.
  - Reduces the influence of erroneous sectors of the contours.
  - M-estimation technique is used by taking
    $$r_i^2 = \frac{1}{l_i^{'}} \sum_{j=1}^{l_i^{'}} \rho(d_{i,j})$$
    where ρ is Huber's function.
- Global stage:Which minimizes the robust function of feature residual.

– M-estimation technique is used by taking $\min_p \sum_{i=1}^{n} \rho(r_i)$, where ρ is Tukey's function.

- The residual is compared with the standrd deviation of all the residuals.
- If it is larger define as outlier and discard the feature outlier.
- Minimization is done using Powell's method.

s-2:Key point detection and Matching

- Tracking only model feature does not give sufficient accuracy, hence key points are detected and matched.
- Detecting key points:Two methods
  - Extracting edges searching for maximum curvature.
  - Gray level image using autocorrelation function.
- Detecting or extracting key points is carried

out using Harri-Stephen's autocorrelation method.

- Matching :
  - Matching is done with correlation technique of C Schmid et.al.
  - Each point is characterized locally by a vector of differntial invarients under a group of displacements.
  - Measure the similarity between the invarient vectors leads to matches.

s3:Mixing Method

- Let $[\mathbf{R_k}, \mathbf{t_k}]$ and $[\mathbf{R_{k+1}}, \mathbf{t_{k+1}}]$ are viewpoint in k and k+1, $\triangle R, \triangle t$ be the displacement of the camera.

- Simple way to improve is to minimize

$$\min_{R_{k+1}, t_{k+1}} \{\frac{1}{n} \sum_{i=1}^{n} \rho_1(r_i) + \frac{\lambda}{m} \sum_{i=1}^{m} \rho_2(v_i)\}\text{i},$$

s4:Resolving Occlusions:

Resolving occlusions consists of four steps.

- Computation of the initial mask of the virtual object, ie. the region in theimage to which the object corresponds, assuming that whole object is visible.
- Tracking the contours: The contour chains are extracted in the region of the image correspoding to mask. Then these chains are tracked using a curved-based tracking. Finally, matching of the contours is done using epipolar constraint.
- Labeling the contour points with *behind* or $in-front-of$:
- Computation of the occlusion mask.

Labelling contour points

- Let $m_1 \in I_1$ and $m_2 \in I_2$, Z_{real} is the 3D pt. Now, let M_{obj} be the 3D pt belonging to virtual object whose projection in I_1 is m_1. This virtual objects occludes the real one only if $Z_{obj} < Z_{real}$.
- The projection of M_{obj} in I_2 can be computed which lies on the epipolar line and can be compared. label the points.
- Identify the occluding objects:Consider new connected chains that are only made up of $in - front - of - points$ and group the chains that correspond to the same object.
- Computing the occlusion masks using proximity creterion and regularization theory.

Move Matching for general motion

by, G.simon, Zisserman et.al

- **Initialization:**
  - **Manually indicate the planar region in image-0.(corner detection and matching are restricted to this region).**
  - **Detect interest points in image-0.**
  - **Initialize camera calibration K.**
- **Steady state: Comuting H from frame i to $i+1$.**
  - **Detect N interest points in frame $i+1$, giving the set $\{x_j^{i+1}\}_{j=1}^{N}$.(Using public domain code for Harris corner detector.)**
  - **Match interest points from frame i to $i+1$. This generates set of correspondences $x_j^i \leftrightarrow x_k^{i+1}$.(Each**

pixel is compared with 10 others).

- **From the set of correspondences, robustly compute H_i^{i+1}.**
- **Compute pose from H_i^{i+1}.**

Computing P^i

- Compute the homography of world co-ordinates to one of the frame, say frame-0, ie. H_w^0. Then measure H_k^{k+1} for any pair of frames $(k, k+1)$ from the tracked points. $H_0^i = H_{i-1}^i H_{i-2}^{i-1} H_{i-3}^{i-2} .. H_0^1$, and we get $H_w^i = H_0^i H_w^0$.
- We know $H_w^i = K[r_1^i \quad r_2^i \quad t^i]$ and if the calibration is known, then $P^i = K[r_1^i \quad r_2^i \quad (r_1^i \times r_2^i) \quad t^i]$.
- With the known H_w^0, one can compute camera positions for every frame of the sequence, using only the frame-to-frame homographies H_i^{i+1}.

Z-keying Method

- TV weather report-uses chroma-keying, which can't deal with occlusions.
- Uses dense depth map as a switch.
- For each pixel, the z-key switch compares the pixel depth values of two images, and routes the colour value of the foreground image that is nearer to the camera for the merged output image.
- Real and virtual objects will occlude correctly.
- Uses real time stereo-machine,(specs).
  - Number of cameras-2-6.
  - Frame rate -max 30frames/sec.
  - Depth image size-up to 256×240.
  - Disparity search range-60 pixels.

References

- Gaurav Agarwal, Dinesh RAthi, Prem Kalra, Subhashis Banerjee, A System for Image Based Rendering of Walk-throughs,*Accepted for Computer Graphics International (CGI2002),* Bradford, UK. July, 2002.
- G. Simon, V. Lepetit, M.-O. Berger Registration Methods for Harmonious Integration of Real Worlds and Computer Generated Objects *In Proceedings of the Advanced Research Workshop on Confluence of Computer Vision and Computer Graphics,* Ljubljan (Slovenia),1999.
- G. Simon, A. Fitzgibbon, A. Zisserman Markerless Tracking using Planar Structures in the Scene *In Proceedings of the International Symposium on Augmented Reality,* Munich (Germany),2000.
- T.Kanade et.al, Video-Rate Z Keying: A New Method for Merging Images. *Tech. report*

CMU-RI-TR-95-38, Robotics Institute, Carnegie Mellon University December, 1995. (India), 1998.

- M. Uenohara and T.Kanade. Vision based object registration for real time image overlay. *Journal of computers in biology and Medicine*,1996.
- M.-O. Berger, C. Chevrier, G. Simon Compositing Computer and Video Image Sequences: Robust Algorithms for the Reconstruction of the Camera Parameters, *Eurographics. Computer Graphics Forum,* Blackwell Publishers,1996.
- M.-O. Berger Resolving Occlusion in Augmented Reality: A Contour-Based Approach Without 3D Reconstruction *In Proceedings of the Conference on Computer Vision and Pattern Recognition,*pages 91-96,1997.

THANK YOU