Video Fields: Fusing Multiple Surveillance Videos Into a Dynamic Virtual Environment

Ruofei Du, Sujal Bista, and Amitabh Varshney
www.Video-Fields.com
www.Augmentarium.com

Augmentarium | Department of Computer Science | UMIACS
University of Maryland, College Park
Video Fields: Fusing Multiple Surveillance Videos into a Dynamic Virtual Environment

Ruofei Du, Sujal Bista, Amitabh Varshney
The Augmentarium | UMIACS | University of Maryland, College Park
{ruofei, sujal, varshney} @ cs.umd.edu
www.VideoFields.com
Introduction

Surveillance Videos - Monitoring
Introduction
Surveillance Videos – Shopping Centers
Introduction

Surveillance Videos - Airports
Introduction

Surveillance Videos – Train stations
Introduction

Surveillance Videos - Campuses
Introduction
Surveillance Videos - Conventional
Introduction

Surveillance Videos – Cognitive Burden
Introduction
Surveillance Videos – Fusing & Interpreting
Related Work

Fusing Multiple Static Photographs
Related Work

Fusing Multiple Static Photographs

Abstract

We present a system for interactively browsing and exploring large unstructured collections of photographs as a scene using a novel 3D interface. Our system consists of an image-based modeling stage that automatically reconstructs the geometry and photometric properties of each photograph as well as a rendering stage that越多 the scene and image properties.

Our system employs image-based rendering techniques to seamlessly blend between photographs, while maintaining full"Photographic" realism of the image and its world configuration, along with additional information such as viewpoint.

Our system also makes it easy to construct photo tours of scenes of interest, allowing us to accurately depict details, which are automatically annotated to the scene layout. We demonstrate our system using several large photo collections as well as images gathered from Internet photo-sharing sites.

CCS Concepts: I.3.4 [Information Interfaces and Presentation]: Multimedia Information Systems—Architectural requirements; I.3.7 [Human--Computer Interaction]: User Interface—Models and Usability of graphical interfaces.

Keywords: image-based modeling, image-based rendering, photo tours, travel, virtual environments.
Related Work

Fusing Multiple Static Photographs
Related Work

Fusing Multiple Static Photographs

Social Snapshot: A System for Temporally Coupled Social Photography

Robert Paiz, Cheuk Yin Ip, Sujal Bista, and Aalabh Varsaniy • University of Maryland, College Park

Social Snapshot actively acquires and personalizes temporally dynamic data. The system enables spatial-temporal 3D photography using commodity sensors assisted by new auxiliary sensors and network functionality. It augments users, making them active rather than passive participants in data acquisition.

The next phase in the photography revolution, 3D photography, can bring users together to socialize and collaboratively take pictures in an entirely new way. However, transforming a photographic scene from 2D to 3D requires introducing multiple layers of scene understanding geometry from different viewpoints. The reconstruction of 3D geometry from multiple viewpoints is one of the classic structures-from-motion (SFM) problems in computer vision. Typically, the instruments used to acquire photographs are calibrated to produce precise measurements.

To simplify 3D photographs, our SocialSnapshot system performs active photography and socialization.

Social Snapshot’s Contributions

Social Snapshot’s contributions are naturally into two categories: technical and social.

The technical contributions are improved algorithms and techniques that enhance our system’s accuracy and scalability. For example, Social Snapshot produces a temporally and colored-depth reconstruction from a sparse oriented photo collection, rather than the sparse or dense point reconstructions produced by related approaches. In addition, it features locally optimized mesh generation and viewing. Finally, it provides camera network capabilities to support synchronized capture of temporally dynamic data.

The social contributions lead to a new way of thinking about the interplay between data acquisition and social interactions. They also let us define social photography as an active, rather than passive, medium. For example, Social Snapshot encourages collaborative photography as a social experience, letting users capture dynamic scenes by synchronizing their photographs. It leverages social norms, such as explicit media sharing and digital organization, to spur social data acquisition into mode.

For a look at some of the working features, please visit the accompanying demo page.
Related Work

Fusing Multiple Static Photographs

Social Snapshot: A System for Temporally Coupled Social Photos

Robert Patra, Cheuk Yin Leung

Social Snapshot actively acquires and reconstructs temporally dynamic data. The system enables spontaneous social photography using commodity devices, such as new auxiliary sensors and network functionalities. It supports work making them active rather than passive participants in data acquisition.
Here is a comparison between real and virtual reconstructed models.
Related Work

Fusing Multiple Dynamic Videos
Related Work

Fusing Multiple Dynamic Videos
Related Work

Fusing Multiple Dynamic Videos
Dense 3D Motion Capture from Synchronized Video Streams

Jean Pouye1,2
Willow Team
LIENS (CNRS/ENS/INRIA, UMR 8548)
Ecole Normale Supérieure, Paris, France

Yasutaka Furukawa1
Department of Computer Science
and Beckman Institute
University of Illinois at Urbana-Champaign, USA

Abstract: This paper proposes a novel approach to nonrigid, markerless motion capture from synchronized video streams acquired by calibrated cameras. The instantaneous geometry of the observed scene is represented by a polyhedron with fixed topology. The initial mesh is constructed in the first frame using the publicly available PWMS software for multi-view stereo [17]. Its deformation is captured by tracking its vertices over time, using two optimization processes on each frame, a local one using a rigid motion model in the neighborhood of each vertex, and a global one using a regularized nonrigid model for the whole mesh. Quantitative and qualitative experiments using several real datasets show that our algorithm effectively handles complex nonrigid motions and motion outliers.

1. Introduction

The most popular approach to motion capture today is to attach distinctive markers to the body and/or face of an actor, and track these markers in images acquired by multiple calibrated video cameras. The marker tracks are then matched, and triangulation is used to reconstruct the corresponding positions and velocity information. The accuracy of any motion capture system is limited by the temporal and spatial resolution of the cameras. In the case of marker-based technology, it is also limited by the number of markers available. Although relatively few (say, 50) markers are sufficient for rigid body configurations, estimates of nonrigid motions. Markerless technology using special makeup is indeed emerging in the entertainment industry [15], and several approaches to local scene flow estimation have also been proposed to handle less constrained settings [6, 13, 16, 19, 23]. Typically, these methods do not fully exploit global spatiotemporal consistencies.

1.1. Related Work

Three-dimensional active appearance models (AAMs) are used for facial motion capture [11, 14]. In this approach, parameters encoding head facial shape and appearance are fitted to one or several images sequencers. AAMs require at a priori parametric face model and arc, by design, aimed at tracking relatively coarse facial motions rather than reconstructing fine surface detail and subtle expressions. Active appearance approaches to motion capture use a projected pattern to independently estimate the shape in each frame, then use optical flow and/or surface flow techniques to recover the frame-to-frame changes between adjacent frames to recover the frame-to-frame changes between adjacent frames. Although qualitative results are impressive, these methods typically do not exploit the redundancy of the multi-viewpoint information, and may be incapable to estimate parameters of the camera motion. Several passive approaches to capture flow compensation [11, 13, 16, 19, 23]. Some
Related Work

Fusing Multiple Dynamic Videos

Performance Capture from Sparse Multi-view Video

Edithon de Aguiar, Carlos Stoff, Christian Theobalt, Navneet Alahakoon, Hans-Peter Seidel, Sabineillion

MPI Informatik, Saarbrücken, Germany
Stanford University, Stanford, USA

Figure 1: A sequence of pose-captured from eight static recordings of a single actor. Our algorithm detects real-time temporally coherent geometry of the moving performer that captures both the scene, moving performer and a few keyless multi-view.

Abstract

This paper presents a new method for capturing real-time performances from multiple single-view video. Our algorithm can instantly reconstruct spatio-temporally coherent geometry, motion and multi-view appearance of a performer without any special markers or rigid markers. Furthermore, since our algorithm is purely mesh-based and makes as few as possible prior assumptions about the type of subject being tracked, it can even capture performances of people using only a mouse, without a detailed wearing a suit. To achieve this, our system uses efficient and effective techniques to process large amounts of data and enable the system to run under tight constraints. The framework allows multiple users to work together, and each user can provide input from different viewpoints. Our method delivers spatially coherent performance in all high-level of detail, is highly scalable, and is applicable to any computer setup that would not be feasible by alternative methods that would be infeasible for this task.

1 Introduction

The recently released 3D Freeform (3DF) version of the 1996 film "The Fugitive" (1993) makes extensive use of the keyframe animation that was developed to create the final visual output. In this paper, we describe our approach to this problem, and discuss some of the challenges associated with it. Our algorithm is designed to provide a better understanding of the relationship between human and computer interfaces, and the challenges associated with them. The algorithm is designed to be scalable, and is compatible with existing computer hardware and software. It is therefore not new to the field, but it is important to note that our system is capable of capturing and rendering highly detailed multi-view videos for interactive applications. Additionally, our system allows for multi-view capture of multiple performers in real-time, providing a powerful tool for interactive applications.

Keywords: Computer Graphics, Image Processing, Virtual Reality, Human-Computer Interaction.
Probabilistic Deformable Surface Tracking From Multiple Videos

Cedric Cagniart¹, Edmond Boyer², and Slobodan Ilic¹

¹ Technische Universität München
² Grenoble Université - INRIA Rhône-Alpes
{cagniart, slobodan.ilic}@in.tum.de, edmond.boyer@inrialpes.fr

Abstract. In this paper, we address the problem of tracking the temporal evolution of arbitrary shapes observed in multi-camera setups. This is motivated by the ever growing number of applications that require consistent shape information along temporal sequences. The approach we propose considers a temporal sequence of independently reconstructed surfaces and iteratively deforms a reference mesh to fit these observations. To effectively cope with outlying and missing geometry, we introduce a novel probabilistic mesh deformation framework. Using generic local rigidity priors and accounting for the uncertainty in the data acquisition process, this framework effectively handles missing data, relatively large reconstruction artefacts and multiple objects. Extensive experiments demonstrate the effectiveness and robustness of the method on various 4D datasets.

1 Introduction

Inferring shapes and their temporal evolutions from image data is a central problem in computer vision. Applications range from the visual restoration of live events to their analysis, recognition and even synthesis. The recovery of shapes using multiple images has received considerable attention over the last decade and several approaches can build precise static 3D models from geometric and photometric information, sometimes in real time. However, when applied to temporal sequences of moving objects, they provide temporally inconsistent shape models by treating each frame independently hence ignoring the dynamic nature of the observed event.

Most methods interested in tracking deformable surfaces in multi-camera systems deform a reference template to fit observed geometric cues as well as possible at each time frame. These cues appear in the literature as photogrammetric observations or intensity measurements.
Deformable 3D Fusion: From Partial Dynamic 3D Observations to Complete 4D Models

Weipeng Xu1,2, Mathieu Salzmann2,3, Yongfeng Wang1, Yue Liu1
1Beijing Institute of Technology, China
2NICTA, Canberra, Australia
3CVLab, EPFL, Switzerland

Abstract
Capturing the 3D motion of dynamic, non-rigid objects has attracted significant attention in computer vision. Existing methods typically require either highly complete 3D reference observation, or a shape template. In this paper, we introduce a template-less 4D reconstruction method that incrementally fuses highly incomplete 3D observations of a deforming object, and generates a complete, temporally-coherent shape representation of the object. To this end, we design an online algorithm that alternatively registers new observations to the current model estimate and updates the model. We demonstrate the effectiveness of our approach on reconstructing non-rigid moving objects from highly incomplete measurements on both sequences of partial 3D point clouds and Kinect views.

1. Introduction
In this paper, we introduce an approach to estimating a temporally coherent 3D model of a non-rigid object given a dynamic sequence of highly-incomplete 3D observations of the object undergoing large deformations. Capturing the 3D motion of dynamic objects, or 4D reconstruction, has been a longstanding goal of computer vision. Ultimately, the resulting models should yield a temporally-coherent shape representation of the observed deformable object.

Multi-view reconstruction methods have been well studied to address 4D reconstruction. While current methods achieve impressive results [12, 9, 6, 21, 22, 26], they typically require well-calibrated setups.

In the case of rigid motion, several fusion techniques have been proposed to combine multiple partial 3D observations [23, 28, 41]. However, when the object undergoes ( quasi-) rigid motion, to acquire a complete 3D template of the object, which will then be deformed to match new non-rigid data.

By contrast, in this paper, we introduce a template-less 4D reconstruction method which is robust to uncalibrated cameras and can effectively handle large deformations.
Related Work

Fusing Multiple Dynamic Videos

Real-time Simultaneous Pose and Shape Estimation for Articulated Objects Using a Single Depth Camera

Mao Ye  Ruihang Yang
University of Kentucky
Lexington, Kentucky, USA, 40506

Abstract

In this paper, we present a novel real-time algorithm for simultaneous pose and shape estimation for articulated objects, such as human beings and animals. The core of the pose estimation component is to embed a novel deformation model with explicit pose constraints. The pose estimation component is based on a robust probabilistic model, where each candidate pose is associated with a score. The shape estimation component, on the other hand, is based on a shape reconstruction algorithm that computes the shape of the object by optimizing its appearance model.

1. Introduction

The estimation of pose and shape is a fundamental problem in computer vision. For articulated objects, the estimation of pose and shape is even more challenging due to the complex motion and deformation of the object. In this paper, we present a novel real-time algorithm for simultaneous pose and shape estimation for articulated objects, such as human beings and animals.

Figure 1: Our novel algorithm effectively estimates the pose of articulated objects using a single depth camera, such as human and dogs even with challenging scenes.
Related Work

Fusing Multiple Dynamic Videos
Related Work

Fusing Multiple Dynamic Videos

Abstract

This paper proposes a novel technique for fusing multiple dynamic videos. The technique is based on a model-based approach, where a deformable model is used to represent the objects in the scene. The model is then registered to the videos, allowing for the synchronization of the content across different views.

1. Introduction

The goal of this paper is to develop a method for fusing multiple dynamic videos into a single, coherent representation. The proposed method is based on a combination of model-based and appearance-based techniques, allowing for a flexible and robust approach to video fusion.

Keywords:
- Dynamic fusion
- Model-based
- Appearance-based

Related Work

Fusing Multiple Dynamic Videos

DynamicFusion: Reconstruction and Tracking of Non-rigid Scenes in Real-Time

Richard A. Newcombe
necsoftsigns.com, washington.edu

Andrew J. Davison
Imperial College London

Dieter Fox
necsoftsigns.com, washington.edu

Steven M. Seitz
washington.edu, seattle

Abstract

We present a novel approach for fusing multiple dynamic videos into a single, coherent representation. The approach is based on a combination of model-based and appearance-based techniques, allowing for a flexible and robust approach to video fusion.

Figure 1. Example reconstruction of a moving scene with DynamicFusion. Both the person and the camera are moving. The visually consistent fusion is achieved by fusing multiple views into a single, coherent representation.

3D scene reconstruction involves capturing and representing the spatial relationships between objects in a scene. The proposed approach is based on a combination of model-based and appearance-based techniques, allowing for a flexible and robust approach to video fusion.

DynamicFusion: Reconstruction and Tracking of Non-rigid Scenes in Real-Time

Richard A. Newcombe
necsoftsigns.com, washington.edu

Andrew J. Davison
Imperial College London

Dieter Fox
necsoftsigns.com, washington.edu

Steven M. Seitz
washington.edu, seattle

Abstract

We present a novel approach for fusing multiple dynamic videos into a single, coherent representation. The approach is based on a combination of model-based and appearance-based techniques, allowing for a flexible and robust approach to video fusion.

Figure 1. Example reconstruction of a moving scene with DynamicFusion. Both the person and the camera are moving. The visually consistent fusion is achieved by fusing multiple views into a single, coherent representation.

3D scene reconstruction involves capturing and representing the spatial relationships between objects in a scene. The proposed approach is based on a combination of model-based and appearance-based techniques, allowing for a flexible and robust approach to video fusion.

DynamicFusion: Reconstruction and Tracking of Non-rigid Scenes in Real-Time

Richard A. Newcombe
necsoftsigns.com, washington.edu

Andrew J. Davison
Imperial College London

Dieter Fox
necsoftsigns.com, washington.edu

Steven M. Seitz
washington.edu, seattle

Abstract

We present a novel approach for fusing multiple dynamic videos into a single, coherent representation. The approach is based on a combination of model-based and appearance-based techniques, allowing for a flexible and robust approach to video fusion.

Figure 1. Example reconstruction of a moving scene with DynamicFusion. Both the person and the camera are moving. The visually consistent fusion is achieved by fusing multiple views into a single, coherent representation.

3D scene reconstruction involves capturing and representing the spatial relationships between objects in a scene. The proposed approach is based on a combination of model-based and appearance-based techniques, allowing for a flexible and robust approach to video fusion.

Related Work

Fusing Multiple Dynamic Videos
Related Work

Fusing Multiple Dynamic Videos
Related Work

Fusing Multiple Dynamic Videos

SIGGRAPH 2016

Wednesday, 3:30 - 4:00 PM
Related Work

Fusing Multiple Dynamic Videos
Related Work

Fusing Multiple Dynamic Videos
Related Work

Fusing Multiple Dynamic Videos
Our Approach?
Video Fields
Fusing multiple RGB surveillance videos without feature matching algorithms nor calibration patterns.
Introduction

Video Field
Introduction

Surveillance Videos

They monitors a variety of activities in shopping centers, airports, train stations, and university campuses.
Introduction
In this paper we introduce, Video Fields, a novel web-based interactive system to create, calibrate, and render...
Conception, architecting & implementation

Video Fields

A mixed reality system that fuses multiple surveillance videos into an immersive virtual environment,
Integrating automatic segmentation of moving entities

Video Fields Rendering

Real-time fragment shader processing
Two algorithms to fuse multiple videos

Early & deferred pruning

These methods use voxels and meshes respectively to render moving entities in the video fields.
Achieving cross-platform compatibility by WebGL + Three.js

smartphones, tablets, desktop, high-resolution large-area wide field of view tiled display walls, as well as head-mounted displays.
System Overview
Architecture

Video Fields Flowchart

Input Video Streams

WebGL Modeling Interface
- Calibrate Cameras
- Create Building Geometries

Video-Field Server
- Background Modeling

WebGL Renderer
- Interactive Segmentation
- Tracking Moving Entities
- Transparency Control

Early Pruning
- Deferred Pruning

Rendering to Target Displays
Architecture

Video Fields Flowchart

- Surveillance video streams
- Calibration of camera world matrices
- Static 3D models and satellite image

Video Fields Mapping

- Dynamic virtual environment
- Automatic segmentation and view-dependent rendering
Video-Fields
Mockup & Calibration

... dynamic video-based virtual reality scenes in head-mounted displays, as well as high-resolution wide-field-of-view tiled display walls.
Architecture

Video Fields Flowchart

Input Video Streams

WebGL Modeling Interface
  - Calibrate Cameras
  - Create Building Geometries

Video-Field Server
  - Background Modeling

WebGL Renderer
  - Interactive Segmentation
  - Tracking Moving Entities
  - Transparency Control

Early Pruning
Deferred Pruning

Rendering to Target Displays
• Provide a background texture for each camera
• Identify moving entities in the rendering stage
• Reduce the network bandwidth requirements
\( T(u, v)_i = \{ T(u, v, j), 1 \leq j \leq i \} \)

\[ \mathcal{D}(T(u, v)_i) = \sum_{j=1}^{N} \mathcal{N}(T(u, v)_i | \mu_{ij}, \Sigma_{ij}) \cdot \omega_{ij} \quad (1) \]

\[ \mathcal{N}(T(u, v)_i | \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{1}{2}}} \cdot \frac{1}{\Sigma^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}(T(u, v)_i - \mu_{ij})^T \Sigma_{ij}^{-1} (T(u, v)_i - \mu_{ij})} \quad (2) \]

\[ \omega_{ij} \leftarrow (1 - \alpha)\omega_{i(j-1)} + \alpha M_{ij} \quad (3) \]
More adaptive with:
- different lighting conditions,
- repetitive motions of scene elements,
- moving entities in slow motion
\[ T' \leftarrow \mathcal{G}(\sigma) \otimes T, \quad B' \leftarrow \mathcal{G}(\sigma) \otimes B \]

\[ F \leftarrow \delta(||I' - B'||) \]
Background Modeling
Gaussian Mixture Models (GMM)

(a) source video texture
(b) background model by GMM
(c) segmentation without Gaussian convolution
(d) segmentation with Gaussian convolution
Our system integrates background modeling and automatic segmentation of moving entities with rendering of video fields.
Architecture

Video Fields Flowchart

Input Video Streams

WebGL Modeling Interface
- Calibrate Cameras
- Create Building Geometries

Video-Field Server
- Background Modeling

WebGL Renderer
- Interactive Segmentation
- Tracking Moving Entities
  - Transparency Control
- Early Pruning
- Deferred Pruning

Rendering to Target Displays
(a) Rendering before visibility testing and opacity modulation

(b) Rendering after visibility testing and opacity modulation
Video-Fields

See-through Buildings

It allows users to adjust camera parameters, navigate through time, walk around the scene, and see through the buildings.
Architecture

Video Fields Flowchart

Input Video Streams

- Calibrate Cameras
- Create Building Geometries

Video-Field Server
- Background Modeling

WebGL Modeling Interface

WebGL Renderer
- Interactive Segmentation
- Tracking Moving Entities
- Transparency Control

Early Pruning
- Deferred Pruning

Rendering to Target Displays
1. Vertex in the 3D models -> Pixel in the texture space
2. Pixel in the texture space -> Vertex on the ground

• The second is useful for projecting a 2D segmentation of a moving entity to the 3D world
\[ \hat{p}_{xyzw} \leftarrow C \cdot G \cdot (p_{xyz}, 1.0) \]  \hspace{1cm} (6)\\

\[ t_{uv} \leftarrow \left( \frac{\hat{p}_x + \hat{p}_w}{2\hat{p}_w}, \frac{\hat{p}_y + \hat{p}_w}{2\hat{p}_w} \right) \]  \hspace{1cm} (7)
Video Fields Mapping

Perspective correction

(a) Video Fields mapping before perspective correction

(b) Video Fields mapping after perspective correction
Video Fields Mapping

Depth Map / Hashing Function

\[ H : t_{uv} \rightarrow p_{xyz} \]
ALGORITHM 1: Early Pruning for Rendering Moving Entities

Input: foreground $F$ and the set of bounding rectangles $R$ of moving entities

Output: a 3D point cloud $P$ visualizing the moving entities

1. Initialize a set of points for the video visualization. (Run once);
2. For each pixel $t$ inside the bounding box, calculate the intersection point $t_{\perp}$ between its perpendicular line and $t_{1}t_{3}$;
3. for each pixel $t$ from the video do
   4. if $t \notin F$ then
      5. discard $t$ and continue;
   6. set the color of the pixel: $c \leftarrow \text{texture2D}(F, t)$;
   7. look up the corresponding projected points in the 3D scene:
      $p \leftarrow \mathcal{H}(t), \ p_{\perp} = \mathcal{H}(t_{\perp})$;
   8. update the $z$ coordinate of the 3D point:
      $p_{z} \leftarrow |p - p_{\perp}| \cdot \tan(\theta_{p})$;
   9. use the $x, y$ coordinates of $t_{\perp}$ to place the point vertically:
      $p_{xy} \leftarrow t_{uv}$;
10. render the point $p$;
ALGORITHM 2: Deferred Pruning for Rendering Moving Entities

**Input:** foreground $F$ and the set of bounding rectangles $R$ of moving entities

**Output:** a set of billboards rendering the moving entities

1. Initialize a set of billboards to display moving objects. (Run once);
2. for each detected bounding box $r$ in $R$ do
   3. calculate the bottom-left, bottom-middle, bottom-right and top-middle points $t_1$, $t_2$, $t_3$, $t_4$ in $r$, as illustrated in Fig. 5;
   4. look up the corresponding projected points in the 3D scene:
      
      $p_i \leftarrow \mathcal{H}(t_i), i \in \{1, 2, 3, 4\}$;

   5. calculate the width of the billboard in the 3D space:
      
      $w \leftarrow |p_3 - p_1|, h \leftarrow |p_4 - p_2| \cdot \tan(\theta_{p_4})$;

   6. Reposition a billboard to the position $p_1 + p_3 \over 2$ with width and height $w$ and $h$;

   7. In the fragment shader of the billboard, sample the color from $I$ as described in Equation. 6 and 7, but replace $G$ with the current billboard’s model matrix; discard pixels which does not belong to the foreground $F$;
Visual Comparison

Early Pruning vs. Deferred Pruning

(a) early pruning for rendering moving entities
(b) deferred pruning for rendering moving entities
View-dependent Rendering
View-dependent Rendering
View-dependent Rendering
View-dependent Rendering
## Experimental Results

### Early Pruning vs. Deferred Pruning

<table>
<thead>
<tr>
<th>Render Algorithm</th>
<th>Resolution</th>
<th>WebVR</th>
<th>Framerate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early Pruning</td>
<td>2560 × 1440</td>
<td>No</td>
<td>60.0 fps</td>
</tr>
<tr>
<td></td>
<td>2 × 960 × 1080</td>
<td>Yes</td>
<td>55.2 fps</td>
</tr>
<tr>
<td></td>
<td>6000 × 3000</td>
<td>No</td>
<td>48.6 fps</td>
</tr>
<tr>
<td>Deferred Pruning</td>
<td>2560 × 1440</td>
<td>No</td>
<td>60.0 fps</td>
</tr>
<tr>
<td></td>
<td>2 × 960 × 1080</td>
<td>Yes</td>
<td>41.5 fps</td>
</tr>
<tr>
<td></td>
<td>6000 × 3000</td>
<td>No</td>
<td>32.4 fps</td>
</tr>
</tbody>
</table>
Video Fields: Fusing Multiple Surveillance Videos Into a Dynamic Virtual Environment

Ruofei Du, Sujal Bista, and Amitabh Varshney
www.Video-Fields.com
www.Augmentarium.com

Augmentarium | Department of Computer Science | UMIACS
University of Maryland, College Park

Vocal: Sai Yuan; BGM: Ukulele by Bensound CC
## Experimental Results

### Early Pruning vs. Deferred Pruning

<table>
<thead>
<tr>
<th>Render Algorithm</th>
<th>Resolution</th>
<th>WebVR</th>
<th>Framerate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early Pruning</td>
<td>2560 × 1440</td>
<td>No</td>
<td>60.0 fps</td>
</tr>
<tr>
<td></td>
<td>2 × 960 × 1080</td>
<td>Yes</td>
<td>55.2 fps</td>
</tr>
<tr>
<td></td>
<td>6000 × 3000</td>
<td>No</td>
<td>48.6 fps</td>
</tr>
<tr>
<td>Deferred Pruning</td>
<td>2560 × 1440</td>
<td>No</td>
<td>60.0 fps</td>
</tr>
<tr>
<td></td>
<td>2 × 960 × 1080</td>
<td>Yes</td>
<td>41.5 fps</td>
</tr>
<tr>
<td></td>
<td>6000 × 3000</td>
<td>No</td>
<td>32.4 fps</td>
</tr>
</tbody>
</table>
## Experimental Results

### Early Pruning vs. Deferred Pruning

<table>
<thead>
<tr>
<th>Render Algorithm</th>
<th>Resolution</th>
<th>WebVR</th>
<th>Framerate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early Pruning</td>
<td>$2560 \times 1440$</td>
<td>No</td>
<td>60.0 fps</td>
</tr>
<tr>
<td></td>
<td>$2 \times 960 \times 1080$</td>
<td>Yes</td>
<td>55.2 fps</td>
</tr>
<tr>
<td></td>
<td>$6000 \times 3000$</td>
<td>No</td>
<td>48.6 fps</td>
</tr>
<tr>
<td>Deferred Pruning</td>
<td>$2560 \times 1440$</td>
<td>No</td>
<td>60.0 fps</td>
</tr>
<tr>
<td></td>
<td>$2 \times 960 \times 1080$</td>
<td>Yes</td>
<td>41.5 fps</td>
</tr>
<tr>
<td></td>
<td>$6000 \times 3000$</td>
<td>No</td>
<td>32.4 fps</td>
</tr>
</tbody>
</table>
Visual Comparison
Early Pruning vs. Deferred Pruning

(a) early pruning for rendering moving entities
(b) deferred pruning for rendering moving entities
In this paper we introduce, Video Fields, a novel web-based interactive system to create, calibrate, and render...
Future Work
Scale Up - Hundreds of cameras
Future Work

Bandwidth Problem
Future Work

Holoportation with RGB cameras