Sports VR Content Generation from Regular Camera Feeds

Kiana Calagari et al.

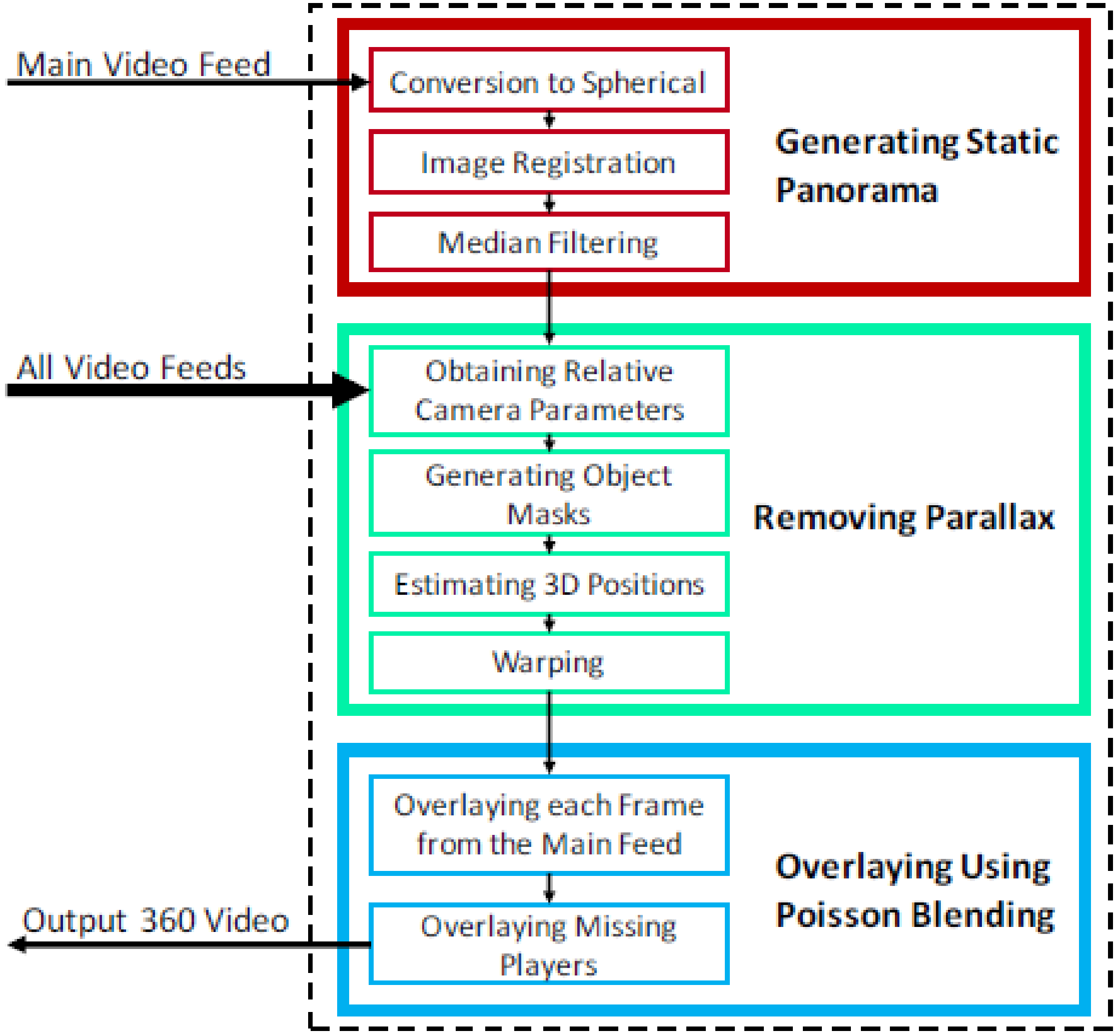
Multimedia Conference (MM), ACM, 2017

Introduction

- One problem remains for VR being adopted on a wider range: the lack of the real content
- Commonly used VR camera rigs expensive to set-up and operate
- The cameras originally deploy in the stadium can be utilized for the 360 content
- The paper proposed a solution for high-quality VR content generation

Proposed Method

1. Generate wide-view static panorama as background
2. Remove the parallax between all video feeds
3. Overlay the information onto the static background



Static Panorama

- Generate the static wide viewing angle panorama as the background
- This stage can perform only once, or periodically during a long time to update the information

Spherical Coordination Conversion

- Use spherical projection to avoid the distortion when aligning planar images
- For each pixel (x, y), we map it to the spherical coordinate (θ, φ)

$$\varphi = \arctan\left(\frac{x}{Z_{img}\cos(\alpha) + y\sin(\alpha)}\right)$$

$$\theta = \frac{\pi}{2} + \arctan\left(\frac{Z_{img}\sin(\alpha) - y\cos(\alpha)}{\sqrt{\left(Z_{img}\cos(\alpha) + y\sin(\alpha)\right)^2 + x^2}}\right)$$

Rotation matrix of the camera

Rotation matrix of the camera

Image Registration

- Transform the camera rotation to a wider angle of view by performing image registration
- Extract and match SIFT features between consecutive frames.
- Use RANSAC (random sample consensus) to select a set of inliners that are compatible with a homography between the frames.
- We then align the frames according to the homography by applying a similarity transformation

Median Filtering

- The static panorama is extracted from the aligned frames using median filtering
- the moving objects will be removed, leaving only the static background
- applying median filter only on key frames can generate sharper results than applying it to all frames

Parallax Removal

- A huge amount of parallax is generated for the cameras are usually placed meters away from each other
- Remove such parallax by warping the camera feeds to the main camera position

(a) Left camera (b) Middle/Main camera (c) Right camera

Relative Camera Parameters

- Obtain an estimation of the relative camera parameters using the VisualSFM
- Estimate the 3D position of each pixel using plane fitting and object masks
- VisualSFM performs a sparse 3D model reconstruction and provides the estimation
  - relative camera positions (C)
  - relative camera rotation matrixes (R)
  - camera focal lengths (f)
  - sample 3D points (X_w, Y_w, Z_w)

Object Masks

- An object mask indicates the pixels which are not part of the static background
- Estimate the 3D position of the objects, identify the missing objects, and copy them on the panorama
- Background subtraction technique

Pixel Estimation

1. A 3D point (X_w, Y_w, Z_w) in world coordinates is first projected to the camera coordinates (X_c, Y_c, Z_c)

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = [R|T] \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

Rotation matrix of the camera

translation vector based on the camera positions $(T = -RC)$

2. 3D point is then projected on the 2D image

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} f\, X_c/Z_c \\ f\, Y_c/Z_c \end{bmatrix}$$

the image coordinates with the origin being at the image center

Focal length

Pixel Estimation cont.

- The field in camera's coordinate system is a plane parallel to the x axis which can be presented as $bY_c + cZ_c = 1$
- Estimate the plane parameters b and c by fitting a plane to sample (X_c, Y_c, Z_c) points
- With the plane parameters, Z_c of each field pixel (x_i, y_i) can be estimated

$$Z_c = \frac{1}{by_i + c}$$

- Based on the object mask, we estimate the Z_c of each object to be the Z_c of the place where it touch the ground

Warping

1. revert the camera projection to find the world coordinates for each pixel (x_i, y_i)

$$\begin{bmatrix} X_{c_{main}} \\ Y_{c_{main}} \\ Z_{c_{main}} \end{bmatrix} = R_{main}\left(R^{-1}\left(\frac{Z_c}{f}\begin{bmatrix} x_i \\ y_i \\ f \end{bmatrix} - T \right) \right) + T_{main}$$

2. project each point back to a 2D image, the new coordinates $(x_{i_{main}}, y_{i_{main}})$ are calculated according to the main camera parameters

$$\begin{bmatrix} x_{i_{main}} \\ y_{i_{main}} \end{bmatrix} = \begin{bmatrix} f_{main}\, X_{c_{main}} / Z_{c_{main}} \\ f_{main}\, Y_{c_{main}} / Z_{c_{main}} \end{bmatrix}$$

Warping cont.

Panorama Overlaying

- Overlay parts of each camera feed on the static panorama to generate 360 video
- Use Poisson image editing to seamlessly blend the overlaid parts with the background
- For each frame, we first overlay the main feed.
- Objects missing from the main feed are then identified and copied from the complementary feeds

Overlaying Main Camera Feed

- First align the camera feed with the panorama, then blend the frame borders using Poisson blending
- To reduce the effect of possible misalignments, if a object is on the borders of the frame, it is removed and considered as a missing object

Overlaying Objects

- Missing objects are identified using object masks
- Align the warped complementary frames with the panorama for identifying and the missing objects
- For a better alignment, convert the field panorama from spherical to planar format for image registration

$$x = \tan(\varphi)\,(Z_{img}\cos(\alpha) + y\sin(\alpha))$$

$$y = Z_{img}\frac{\sin(\alpha) - \dfrac{\tan(\theta - \frac{\pi}{2})\cos(\alpha)}{\cos(\varphi)}}{\cos(\alpha) - \dfrac{\tan(\theta - \frac{\pi}{2})\sin(\alpha)}{\cos(\varphi)}}$$

Overlaying Objects cont.

(a) Main view overlaid on the static panorama

(b) Zoomed version of final panorama with all missing players overlaid

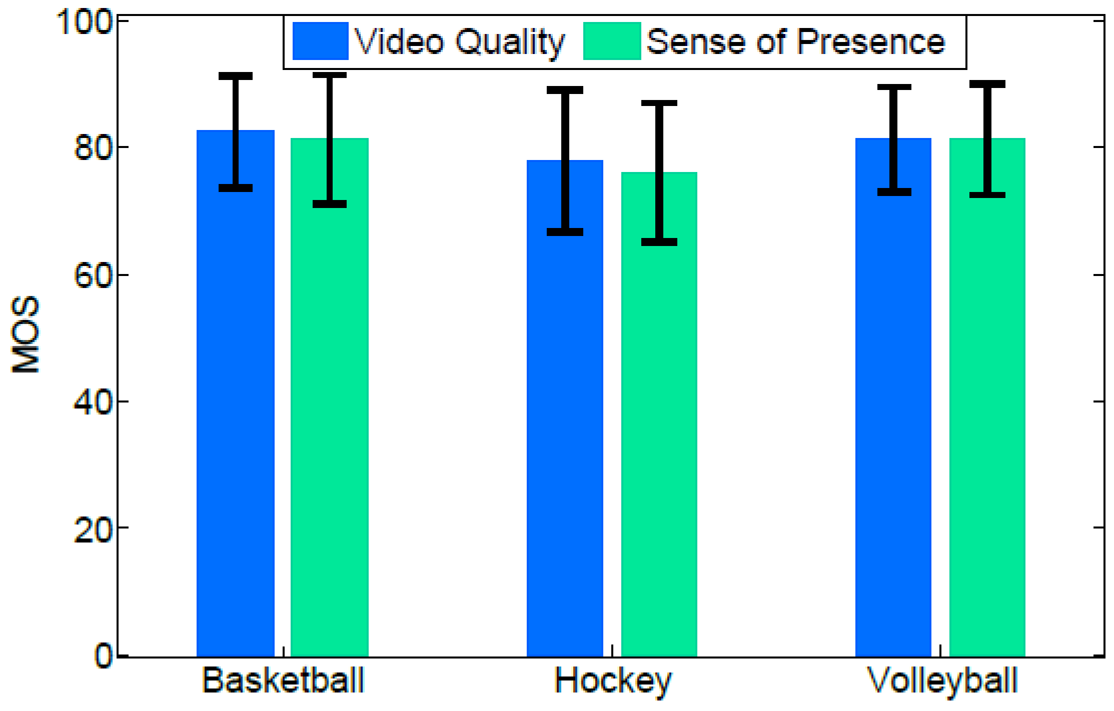
Evaluation Setup

- GoPro Omni 360 camera vs 3 individual GoPro Hero4-black cameras
- All cameras captured the scene simultaneously
- Take Omni 360 camera content as ground-truth
- The 3 individual cameras were deployed at the left, right and middle of the field, the middle one rotates with the action, and is considered as the main feed
- Capture data from basketball, ice hockey, and volleyball game
- A 30-second sequence was chosen and converted to VR content using the proposed method
- Fifteen participants took part in the experiments

Generated Scene

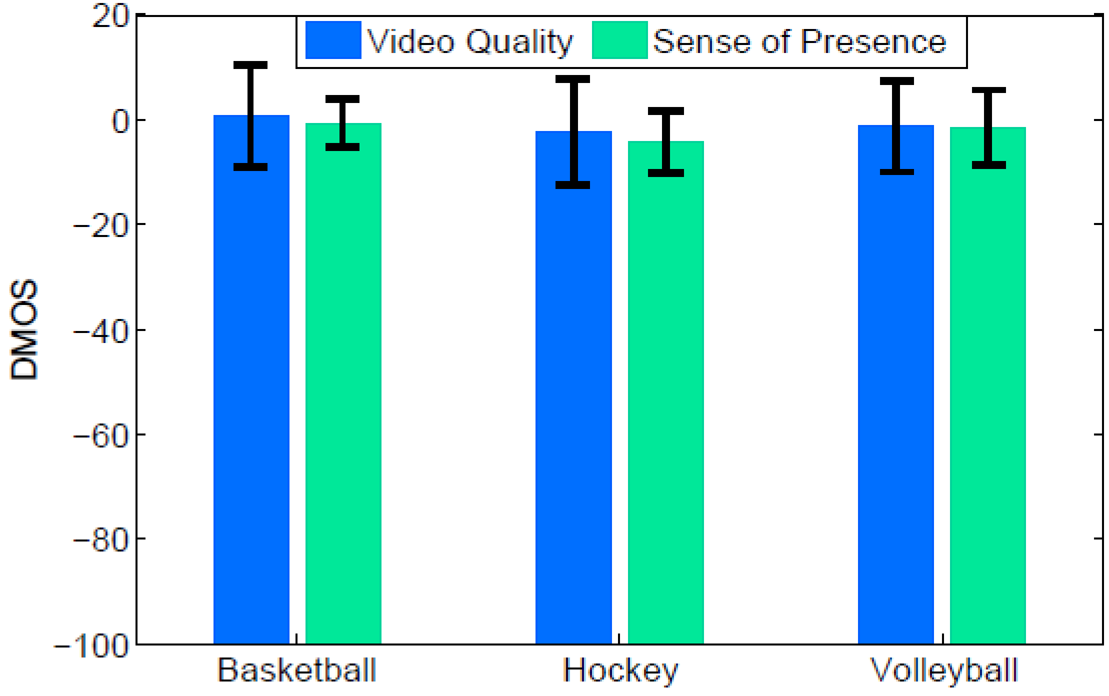
Subjective Evaluation

- Use the standard ITU continuous scale to rate both video quality and sense of presence
- The labels marked on the continuous scale are Excellent, Good, Fair, Poor, and Bad
- Error bars represent the standard deviation
- Low-textured hockey field makes it difficult to perform accurate feature matching and alignment



360 Content Comparison

- Calculate the DMOS (= score for our technique - score for original 360)
- Zero implies that the results are judged the same, while a negative DMOS implies that our result has less quality
- The only statistically significant difference reported (p-value < 0.05) is the sense of presence for hockey



Player Placement Analysis

- Use the originally captured wide-angle main feed as reference
- Measure the distance between the position of each copied player and its original position in the reference frame
- The displacement is highest for hockey, with a maximum around 10 pixels
- A displacement of 10 pixels in the panorama translates to a distance around 20 cm in a real field

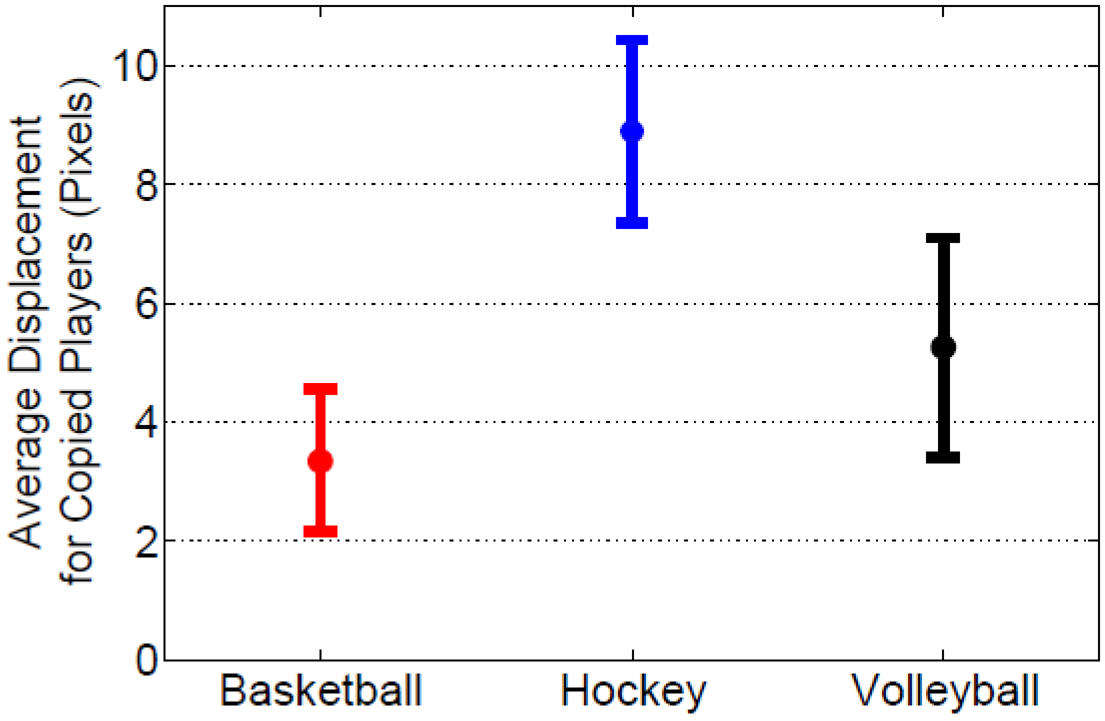


Figure 9: Average displacement of copied players for different games. Error bars represent the standard deviation.

Conclusion

- Present a technique for generating VR content for sports from common broadcast camera feeds
- Subjective evaluation results show that the proposed method is comparable with the original 360 content
- MOS ratings indicate that most participants experienced a strong sense of immersion

Thanks for listening

Any question?