National Technical University of Athens

Image,Video & Multimedia Systems Laboratory

An Optimal Framework for Summarization of Stereoscopic Video Sequences

Nikolaos D. Doulamis, Anastasios D. Doulamis, Yannis S. Avrithis, Klimis S. Ntalianis and Stefanos D. Kollias

IWSNCH3DI99

Introduction

- Recent increase in use of 3-D video.
  - More efficient visual representation.
  - Enhancement of multimedia communications.
  - Provides additional information (depth).
- Increase in storage & communication of visual information (images, video).
- Traditional 3-D sequences recorded sequentially => very hard to perform video browsing, content-based indexing and retrieval.

Problem Statement

- Representation of 3-D video sequences with a small number of *representative frames*.
- *Multiresolution RSST* algorithm applied for color and depth segmentation.
- Description of each frame by a *fuzzy feature vector*.
- Extraction of key frames by means of a *genetic algorithm*.
- Method applicable to any video sequence.

Depth Estimation (1)
Disparity Field

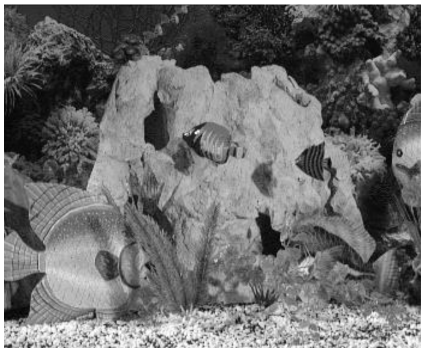
- Disparity vector $\mathbf{d}(x_1,y_1) = [d_x(x_1,y_1)\ d_y(x_1,y_1)]^T$ at (x_1, y_1) in camera 1 with respect to camera 2 is given by:

$$d_x = d_x(x_1,y_1) = x_2 - x_1 = f_1(Z)$$
$$d_y = d_y(x_1,y_1) = y_2 - y_1 = f_2(Z)$$

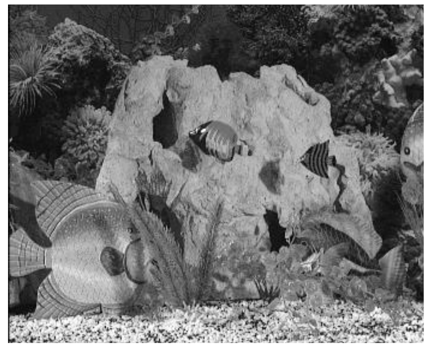
- (x_1, y_1) and (x_2, y_2) image points generated by the projection of a 3-D point **w** onto image planes I_1 and I_2.

Depth Estimation (2)

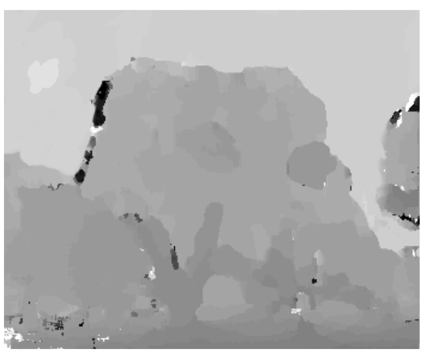
Disparity Field and Depth Map



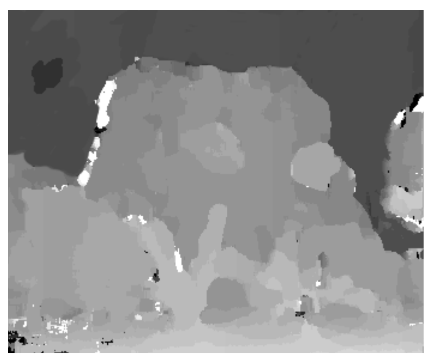
(a)



(b)



(c)



(d)

Aqua Sequence

(a) Right channel

(b) Left channel

(c) Horizontal disparity field

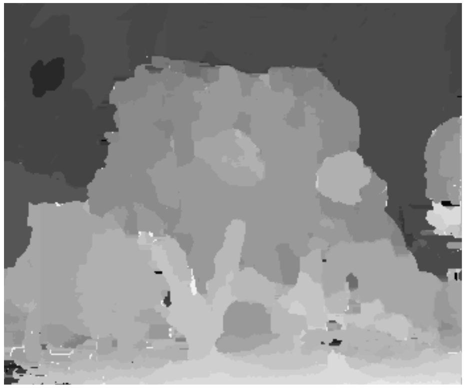
(d) Depth map

Depth Estimation (3)
Occluded Areas

- Wrong estimation of disparity and depth due to occlusion of background objects from foreground ones.
- Occlusion detection by locating regions of I_1 where disparity decreases with a slope approximately equal to -1.
- Occlusion compensation by keeping disparity constant and equal to the maximum disparity of the occluded area.

Depth Estimation (4)
Occlussion Detection & Compensation

(a)



(b)

Aqua Sequence

(a) Compensated horizontal disparity field

(b) Compensated depth map

IWSNCH3DI99

Object Segmentation (1)
Segmentation Algorithm

- Fully automated segmentation scheme.
- Use of M-RSST segmentation algorithm.
  - Multiresolution implementation of the RSST
  - Overcomes the computational complexity of the RSST
  - Reduces the number of very small segments

Object Segmentation (2)

Segmentation Results

(a)

(b)

Aqua Sequence

(a) Color segmentation

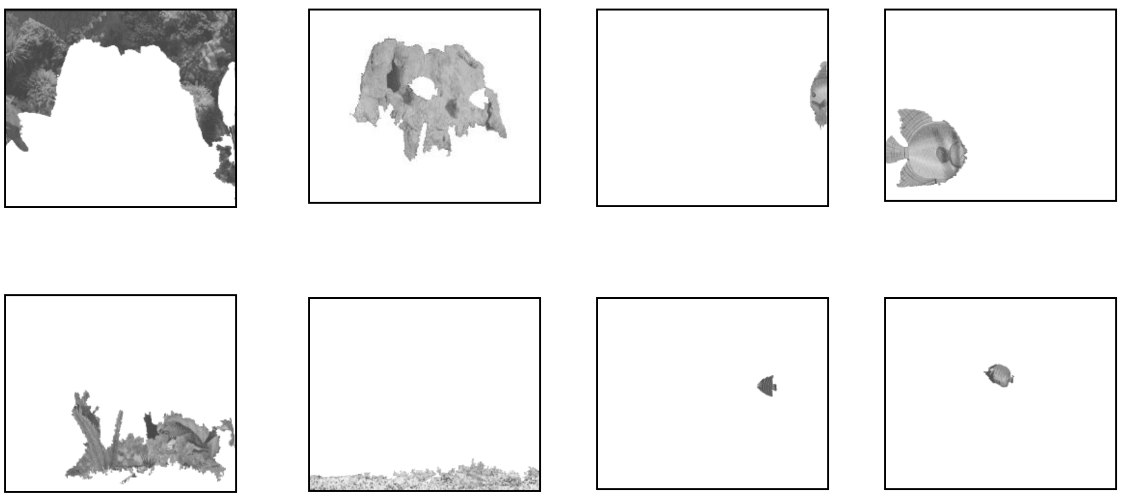
(b) Depth segmentation

Object Segmentation (3)
Color & Depth Segment Fusion

- Projection of color segments onto depth segments.
  - Video objects provided by depth segmentation are retained
  - Object boundaries given by color segments are accurately extracted
  - Each color segment associated to a depth segment so that their area of intersection is maximized

IWSNCH3DI99

Object Segmentation (4)
Object Segmentation Results



Aqua Sequence:

Eight meaningful objects extracted

Stereo Video Summarization (1)
Fuzzy Feature Vector Formulation

- Color and depth segments are classified into predetermined classes, forming a multidimensional histogram.
- A degree of membership allocated to each class resulting in fuzzy classification => reduction of the possibility of classifying two similar segments to different classes.

Stereo Video Summarization (2)
Shot Detection - Frame Extraction

- Shot cut detection algorithm applied to every stereo video sequence.
- For every shot, key frame extraction performed by locating frames with minimum correlation.
  - Use of a correlation measure $R_F(\mathbf{a})$:

$$R_F(\mathbf{a}) = R_F(a_1, \ldots, a_{K_F}) = \frac{2}{K_F(K_F - 1)} \sum_{i=1}^{K_F - 1} \sum_{j=i+1}^{K_F} (\rho_{a_i, a_j})^2$$

IWSNCH3DI99

Stereo Video Summarization (3)
Genetic Algorithm (1)

- Complexity of exhaustive search for the minimum value of R_F (**a**) extremely high => genetic algorithm is adopted.
- Sets of frames represented by chromosomes.
- Initial population of *P* chromosomes is randomly created and new generation populations are generated by applying several operations on existing chromosomes.

Stereo Video Summarization (4)
Genetic Algorithm (2)

- R_F (**a**) used as an objective function to estimate the performance of all chromosomes.
- New chromosomes produced through parent selection, crossover & mutation operations and inserted to the existing population while older chromosomes are removed.
- After several cycles the population converges to a optimal solution.

Experimental Results (1)

- Use of 3-D sequence “Eye to Eye” (25 min or 12,739 frames) for evaluation of the proposed scheme.
- For presentation purposes one shot and one frame every seven are shown.
- Extraction of 4 key frames out of 188 (stereo pairs).

Experimental Results (2)

Shot 38 from television program “Eye to Eye”

Experimental Results (3)

Frame 3805

Frame 3823

Frame 3848

Frame 3960

The extracted characteristic frames

Conclusions

- Stereo video summarization scheme presented, with promising results.
- Use of 3-D information (disparity and depth), M-RSST, fuzzy classification and genetic algorithm.
- Main contribution:
    - Proposal of the first summarization scheme for 3-D sequences