3D From an Image Sequence – Occlusions and Perspective

Amir Shmuel and Michael Werman
Department of Computer Science
The Hebrew University of Jerusalem
91904 Jerusalem, Israel

Abstract
This paper proposes an active solution to optimally recovering 3D scene depth from a sequence of pictures. The choice of the difference between camera locations for the different pictures in the sequence is described. The trade-off between short and long distances between successive locations of the camera is shown in terms of occlusions, perspective transformation and exact triangulation.

1 Introduction
In a former paper ([1]) we applied the concept of active vision in order to enhance depth determination. In [1] an iterative active algorithm is presented for extracting 3D from an image sequence. A mathematical model is proposed that includes the known 3D information in the scene: locations in space and their certainties. The scene is represented by a dense inverse depth map: a value $\frac{1}{\Delta x}$ for each location, where the subscript $i$ denotes the $i^{th}$ pixel. The knowledge of the scene is uncertain. After the completion of the $i^{th}$ iteration the knowledge of the inverse depth of the $i^{th}$ location is represented as a normal random variable:

$$\frac{1}{\Delta x_i} \sim N \left( \frac{1}{\Delta x_i^0}, \sigma_x^2 \right) \quad (\hat{X} \text{ for any } X \text{ means the estimated } X).$$

In [1] it is shown that for exact disparity measurements from motion, the optimal motion of the camera is often a translation within the image plane. The determination of an optimal direction $\theta$ of the camera motion within the image plane is discussed, while assuming that the length $T$ of the base-line is 1. Two natural questions arise:

1. Using the scene information already in hand, what is the optimal length $T$ of the next camera move?
2. Is it possible to use the already known scene 3D characteristics in order to refine the coming measurements?

2 Stereo Base-line Trade-off
It is well known that a long base-line enables reliable determination of inverse depth. It can be best shown using a 2D scene projected perspective onto 1D images. Consider a scene feature whose projections on the two stereo images are known inexactly, and lie in the ranges: $x_1 \pm \Delta x, x_2 \pm \Delta x$. The uncertainty range associated with inverse depth is

$$\frac{|x_1 - x_2| - 2\Delta x}{T} \leq \frac{1}{\hat{T}} \leq \frac{|x_1 - x_2| + 2\Delta x}{T} \quad (1)$$

A long base-line $T$ reduces this range. This observation is valid if the error range $\Delta x$ do not change as $T$ grows. Unfortunately, as the length of the base-line grows - it causes difficulties in the solution of the correspondence problem. Corresponding points in a pair of stereo images rarely have the same intensity values. The main causes for the different intensity images of a stereo pair are: photometric effects, occlusions, different perspective transformations, sensor noise, and discretization.

Differences caused due to photometric effects, originate in the different relative positioning of the camera, scene and illumination sources. The treatment of photometric effects is beyond the scope of this paper. Thus, Lambertian surface characteristics are assumed. The difference in location and view direction of stereo cameras often cause parts of one image to be occluded in the other image. Naturally, intensity changes occur as well. In addition to occlusion, the perspective projections of far scene features change less than those corresponding to close scene features, as the camera shifts. Therefore the intensity function undergoes distortion, even when no occlusions occur.

Among the five mentioned causes, the first three depend on the size of the difference between the cameras' locations and orientations. In the extreme case where there is no change in the camera position, no changes in the intensity are caused by photometric effects, occlusions and different perspective transformation. On the other hand, as the magnitude $T$ grows, these first three mentioned causes are likely to cause bigger changes in the intensity function. This explains the growing difficulties in finding the correspondences, as the length of the base-line gets bigger.

In terms of Eq. 1, the effect of a longer base-line on finding correspondences is expressed by a bigger measurement uncertainty $\Delta x$. Thus, there is a trade-off in the choice of the base-line. This trade-off demands a procedure for evaluating the gain against loss as a function of the length of the base-line. The remainder of this paper introduces an analysis of how stereo images differ from each other due to occlusions. This analysis is the basis for choosing an optimal base-line, by moving as far as possible while assuring that the differences between the images are not to significant.

3 Occlusion Probability
What is the probability that a specific location, belonging to an uncertain disparity map, will have a match in an image taken from another position?

The ability to compute occlusion probability may be used in two ways: choosing the next move of the camera, such that within the region of interest the probability of occlusion is low, and excluding disparity measurements corresponding to regions with high probability of occlusion.
Consider the case where a one-dimensional image camera, modeled as an ideal pin-hole camera, moves in a 2D scene (see Fig. 1, top-left). The focal length of the camera is \( f \), and the length associated with each pixel on the image line is \( p \). The current location of the camera focal point is \((X_F, Z_F)\), in a global coordinate system. Disparity information is given for each pixel of the current location of the camera. The information is uncertain, and modeled as a vector of independent random variables: \( \frac{1}{f_i} \sim f_i \), where \( f_i, 1 \leq i \leq N \), is a known density function. The problem to be solved is: what is the probability that the scene area corresponding to the \( i \)th pixel of the current image, will be occluded in an image taken from a new location \((X_F', Z_F')\)? The new positions that will be considered in this section are the consequence of any translational move of the camera.

We assume a continuous scene curve, though the same procedure can take place for discontinuous curves. It is convenient to approximate the scene depth function as the envelope of \( N \) polygons. The polygon corresponding to the \( i \)th pixel, whose image coordinate is \( z_i \), is composed of two rays and a segment. The rays are \( X_F + \left( z_i \pm \frac{1}{2} \right) Z \) where \( Z \) is a depth parameter. The segment closes each polygon in front. Its \( Z \) coordinate location is uncertain, and characterized by \( f_i \).

A certain location in the scene will be projected on the new image plane, if no other scene part obscures it. This observation justifies defining the occurrence of a match to some location, when a ray transmitted from the focal point through the area of a certain pixel, "hits" the relevant location. We define a match as a ray hitting the front segment of the considered polygon. The suggested method for defining a match is similar to methods used in computer graphics for removing hidden surfaces.

Define the following events

- \( A_{ij} \equiv \) The \( j \)th pixel in camera location \((X_F, Z_F)\) will have a match in camera location \((X_F', Z_F')\).

- \( A_{ij} \equiv \) The \( i \)th ray from camera location \((X_F, Z_F)\) matches the \( j \)th pixel in camera location \((X_F', Z_F')\).

In order that \( A_{ij} \) occurs, the \( i \)th ray should pass all parts of the surface that are prone to block it. Then, it should hit the matching zone of the \( j \)th pixel. In our terms, assuming that the different disparity values are independent

\[
\Pr(\text{Atj}) = \prod_{i=1}^{j-1} \Pr(\text{1 passes polygon m}) \cdot \Pr(\text{1 hits jth segment})
\]

The transformation of the last equation into an explicit computation is performed by means of geometric considerations, as described in [2]. It is shown that

\[
\Pr(A_j) = \sum_{i=1}^{N} \Pr(A_{ij}) - \sum_{i=1}^{N-1} \Pr(A_{ij} \cap A_{i+1,j})
\]

The complexity of computing the probability of occlusion corresponding to a certain location is \( O(N^2) \). The detailed solutions of the 2D and 3D scenes problems can be found in [2].

### 4 Experimental Results

The input of the experiment we show is composed of the uncertain disparity map at the bottom-left part of Fig. 1, that corresponds to the scene image at the top-right part of the same figure. The probability of each pixel location to be viewed from a new camera location was computed. The new simulated camera position was the result of shifting the camera 0.4 cm in 220° relative to the \( z \) axis. The results are presented in Fig. 1, bottom-right (bright gray level corresponds to high probability).

#### Figure 1: Scene reconstruction and definition of matching (top-left), scene image (top-right) and uncertain depth map (bottom-left), and the occlusions probability detected when simulating camera move in 220° (bottom-right).