Motion from Occlusions

César Silva **José Santos-Victor**[*]

Instituto de Sistemas e Robótica
Instituto Superior Técnico
Av. Rovisco Pais, 1
Lisbon
Portugal

Abstract. Finding correspondences is one of central problems in stereo and motion analysis. To perceive depth and motion, the human binocular vision uses surface and edge representations and performs correspondences between two or more images, captured along time. However, a non-correspondence, such as an occlusion, can play an important role in motion and depth interpretation. Anderson and Nakayama [1] have shown that an occlusion is one of the most powerful cues to perceive depth and motion, and influence the earliest visual stages of stereo matching.
On the other hand, in computer vision, the occlusions is often considered as artifacts, undesirable in many applications, mainly for the image motion computation and the stereo matching. Consequently a number of algorithms have been designed to handle occlusions in order to estimate multiple image motions or the disparity field [5, 3, 2, 6].
Nevertheless, in this paper our focus is not on the explicit estimation of image motion (or disparity) but on the role of occlusions in motion perception. Assuming a moving monocular observer, our goal here is to study the advantages of perceiving occlusion points. Therefore we intend to find a connection between the observable occlusions and the motion perception. We will see that occlusions are, per se, important cues for the motion segmentation of different moving objects or the egomotion estimation of a moving observer. First, we develop a sufficient condition for the existence of an occlusion; thereafter we study the relation between egomotion and occlusions, and finally we present some results with real image sequences.

1 A Definition for Occlusion Points

First of all, the occlusion phenomenon can be seen in a simple intuitive geometric way. In a motion sequence, it means that an occluding surface is moving in front of an occluded surface. Consequently there is a discontinuity in the perceived motion and a discontinuity in the depth. However, unless we impose prior models to the image motion field, to the 3D structure or to global image features, we can only decide about the existence of a local occlusion in two consecutive frames if the photometric properties change significantly in a local neighborhood. Thus, we can associate an occlusion point to a photometric value that perceptually "appears" or "disappears" between two consecutive frames, classified respectively as *emergent* or *submergent* occlusion points.
Therefore, an occlusion has to be studied as both a geometric and photometric phenomenon. Consequently we define occlusion points through a sufficient condition based some on a local photometric dissimilarity over time. This sufficient condition can be characterized rigorously for the continuous case, as follows:
Assume that $k = \binom{x}{t}$ and

$$K^+ = \{k : ||k|| = 1 \wedge t > 0\}$$
$$K^- = \{k : ||k|| = 1 \wedge t < 0\}\,,$$

[*] {cesar,jasv}@isr.ist.utl.pt

where x and t represent the spatial and temporal coordinates of an image signal (see Figure 1).
Assume that $f(k)$ is a photometric measure of the image in k (for example the brightness value). The point k_0 is an emergent[2] occlusion point if

$$\exists k^+ \in K^+ : \forall k^- \in K^-, \quad \lim_{\gamma \to 0^+} f(k_0 + \gamma k^+) \neq \lim_{\gamma \to 0^+} f(k_0 + \gamma k^-). \tag{1}$$

Figure 1 illustrates the meaning of the sufficient condition designed here with a simple example of an emergent occlusion point. Figure 1(b) shows that there is a vector k^+ (with positive t) associated to a color (red) which does not exist on the semi-plane $t < t_0$, in a neighborhood of the occlusion point (x_0, t_0).

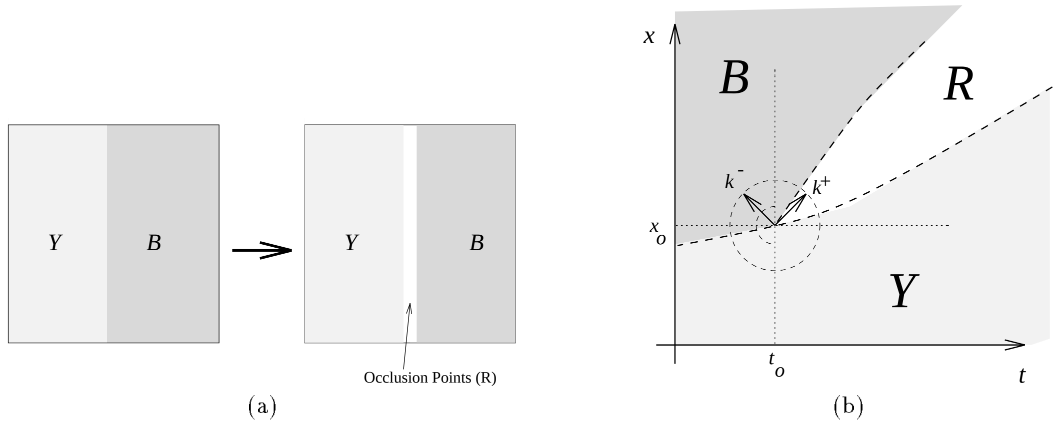


Fig. 1.: Emergent Occlusion Point: (a) The red surface (R) appears between the blue (B) and the yellow (Y) surfaces; (b) Image motion over time t for a horizontal line parameterized by x.

In order to guarantee the applicability of the sufficient condition in the discrete case, we have to define a dissimilarity function. We adopted a function developed by Tomasi and Manduchi [7] that originally was applied to a single image and it was used to smooth images preserving the photometric discontinuities. We changed this function in order to measure the similarity between two consecutive frames.
Suppose that the pixel x_0 in frame t_0 is characterized by a photometric value $f(x_0, t_0)$. The problem to solve is to verify the existence of a similar photometric value (possibly through a perceptually meaningful way) within a given region in frame t_1. We start by defining a function $S(x_0, t_0, x_1, t_1)$ that compares the similarity of $f(x_0, t_0)$ to $f(x_1, t_1)$:

$$S(x_0, t_0, x_1, t_1) = e^{-\frac{1}{2}\left(\frac{\|f(x_1,t_1) - f(x_0,t_0)\|}{\sigma}\right)^2}, \tag{2}$$

where σ^2 corresponds to the variance of the associated gaussian filter. Next, we apply this similarity function to all pixels x_1 in a neighborhood $V(x_0)$ around the pixel x_0 in frame t_1:

$$f_{t_1}(x_0, t_0) = \frac{\sum_{x_1 \in V(x_0)} f(x_1, t_1) \cdot S(x_0, t_0, x_1, t_1)}{\sum_{x_1 \in V(x_0)} S(x_0, t_0, x_1, t_1)}. \tag{3}$$

If $\sigma \to 0$, this function yields (except for some pathological configurations) the photometric value $f(x^*, t_1)$, $x^* \in V(x_0)$, that is closest to $f(x_0, t_0)$. There exists dissimilarity (and therefore (x_0, t_0) is an occlusion point), if $\|f_{t_1}(x_0, t_0) - f(x_0, t_0)\|$ is above a certain acceptable threshold T. This threshold and the variance σ^2 defined in function S depend on perceptual criteria using the photometric range of the images. Notice that if an occlusion point x_0 is detected, then it is an emergent occlusion point if $t_1 < t_0$ or submergent if $t_1 > t_0$.

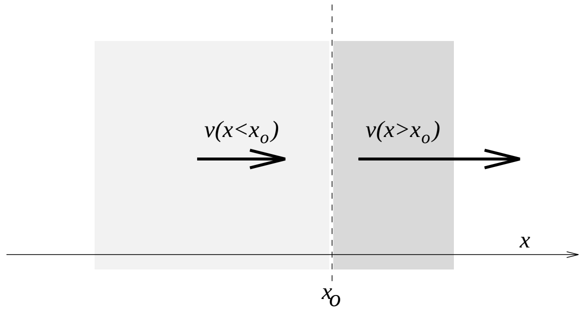
[2] If we exchange the sign of the superscripts of the variables, a similar condition for submergent occlusion points is obtained.

2 Egomotion Perception by using Occlusion Points

In this section we address the problem of egomotion for a moving monocular observer with arbitrary translation and rotation, in a scene with static objects. First of all, the rotation does not produce occlusion points in a perspective camera model, because in this case the image motion does not depend on the depth properties of the scene. Consequently, only the translation produces occlusion points. The translation of an observer can be identified by the projection of the linear velocity on the image plane, known by the Focus of Expansion (FOE).
In order to explore the relation between the FOE and the behavior of the occlusions, for now let us consider a single scanline camera to simplify the problem.
Assume that x parameterizes the scanline defined before and $v(x)$ describes the image velocity along the line. The flow $v(x)$ can be represented as a function of the camera motion parameters [4], as follows:

$$v(x) = \frac{W}{Z(x)}(x - \alpha) + r(x), \tag{4}$$

where $r(x)$ is the motion component due to the camera rotation, α is the FOE projection on the scanline considered, W is the camera velocity component along the optical axis (let us assume that it is positive), and finally $Z(x)$ is the depth of the corresponding 3D point.
If x_0 is an occlusion point then there is a discontinuity in $v(x_0)$ — this means that $v(x_0^-) \neq v(x_0^+)$. On the other hand, if $v(x_0^-) < v(x_0^+)$ then x_0 is an emergent occlusion point (see Figure 2(a)); if $v(x_0^-) > v(x_0^+)$ then x_0 is an submergent occlusion point.
If $Z(x_0^-) > Z(x_0^+)$, then the occluding surface is on the right of the occluded surface (because the occluding surface is naturally nearer than the occluded one). If $Z(x_0^-) < Z(x_0^+)$, then the occluding surface is on the left of the occluded surface (see Figure 2(b)).
Figure 2 illustrates these properties with a simple example, where the occluding surface is on the left of the occluded surface ($Z(x_0^-) < Z(x_0^+)$), and x_0 is an emergent occlusion point ($v(x_0^-) < v(x_0^+)$).
When an occlusion is observed, it can be classified within four classes which consist of the combination of getting a right or left occluding surface and an emergent or submergent occlusion point.



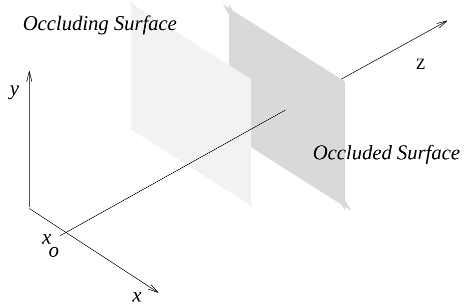


Fig. 2.: Example of an occlusion situation: (a) x_0 is an emergent occlusion point; (b) the projected occluding surface is on the left of x_0.

To classify an occlusion as emergent or submergent, we rely on the method described in Section 1. This alleviates the need to explicitly determine the local optic flow $v(x_0^-)$ and $v(x_0^+)$. To determine whether we have a left or right occluding surface, we monitor the photometric changes occurring in time at the left and right side of the occlusion point. The advantage is that we no longer need to know $Z(x_0^-)$ or $Z(x_0^+)$ to reason about the nature of the occlusion. The method we use seeks the image contour closest to the occlusion point, that preserves both photometric and geometric properties over time, thus belonging to the occluding surface.
We define the following notation to classify the occlusion point, based on functions $p_1(x_0)$ and $p_2(x_0)$:
- $p_1(x_0) = 1$ or $p_1(x_0) = -1$ if the occlusion point x_0 has respectively a left or a right occluding surface;
- $p_2(x_0) = 1$ or $p_2(x_0) = -1$ if x_0 is respectively an emergent or a submergent occlusion point;
- $p_1(x_0)$ and $p_2(x_0)$ are zero if x_0 is not an occlusion point.

By using the inequalities above and the equation (4), we can determine whether the FOE (α) is located to the left or right of the occlusion point x_0:

$$\begin{cases} p_1(x_0) \cdot p_2(x_0) = 1 \Rightarrow (v(x_0^-) - v(x_0^+))(Z(x_0^-)^{-1} - Z(x_0^+)^{-1}) < 0 \Rightarrow x_0 < \alpha \\ p_1(x_0) \cdot p_2(x_0) = -1 \Rightarrow (v(x_0^-) - v(x_0^+))(Z(x_0^-)^{-1} - Z(x_0^+)^{-1}) > 0 \Rightarrow x_0 > \alpha \end{cases}$$

This shows that classifying the occlusion point x_0 corresponds to detecting its location relatively to the FOE. In order to detect automatically the location of the FOE, we designed a function that integrates, along x, the value of the dissimilarity relations from (3) taking into account $p_1(x) \cdot p_2(x)$. This function can be described for the discrete case as follows:

$$F(x,t) = \sum_{\zeta=-\infty}^{x} p_1(\zeta)p_2(\zeta)\left(\|f_{t+1}(\zeta,t) - f(\zeta,t)\| + \|f_{t-1}(\zeta,t) - f(\zeta,t)\|\right)$$

where $t-1$ and $t+1$ correspond to the previous and the next frames. This function decreases if x is on the right of the FOE and increases if it is on the left of the FOE. Thus the FOE is located at the absolute maximum of $F(x,t)$. By integrating the information over the image, the method becomes more robust to eventual false occlusion detections.

3 Results

In this Section, we will apply the occlusion detection to three image sequences. The first sequence (the Lock Sequence, Figure 3) shows the performance of the dissimilarity function in order to find the occlusions points. The Focus of Expansion is roughly at the center of the images and the occlusions found on the boundaries of the lock are of the emergent type, as expected. If one passes an arbitrary line along the image, the relation between the occlusions and the FOE location can be observed.

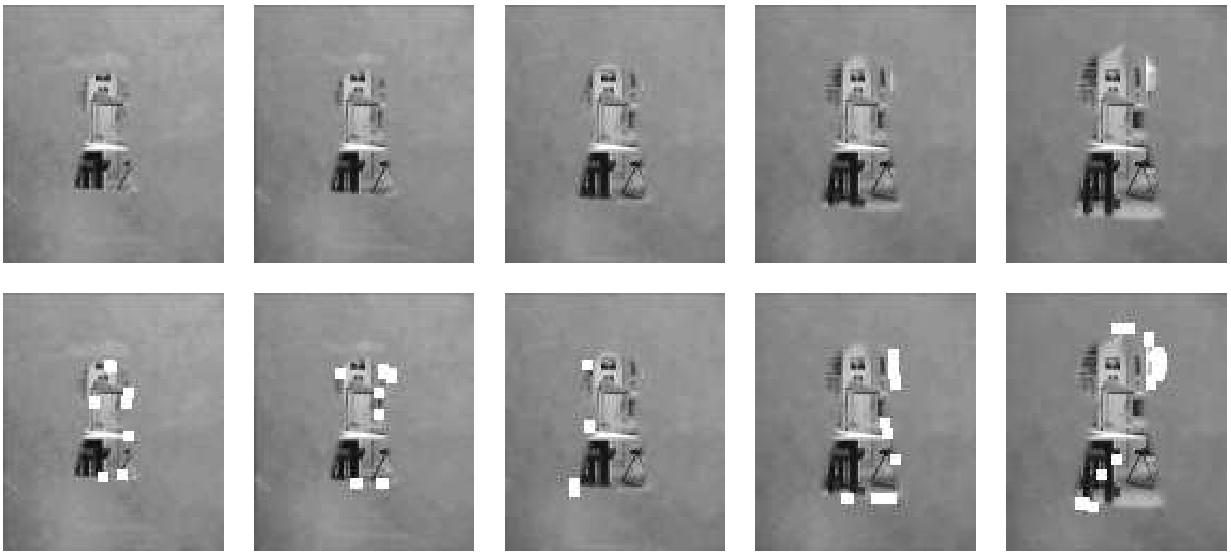


Fig. 3.: Top: Five sequential samples of the Lock Sequence; Bottom: Associated emergent occlusion points, represented by white squares.

The second sequence (the Penguin Sequence, Figure 4) was performed with a static camera with the Penguin moving to the right. In this experiment we see that occlusions can be used for the segmentation of a moving object. The function $F(x,t)$ was computed integrating the information included in the set of all horizontal scanlines. Then we can notice both the occlusion boundaries of the Penguin and its velocity direction, through the decreasing behavior of the function.

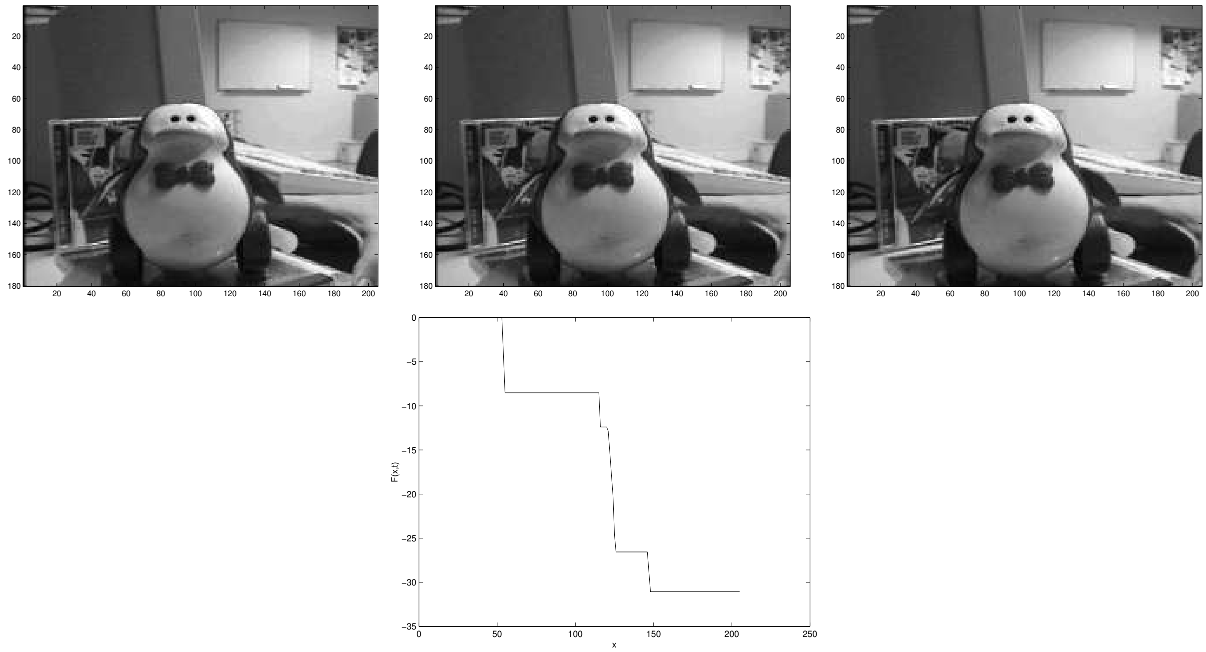


Fig. 4.: Top: Three sequential samples of the Penguin Sequence. Bottom: the function $F(x, y)$ along x. Notice that the decreasing steps correspond to penguin occlusion boundaries.

The third sequence (the Tree Sequence, Figure 5) consists in a leftward moving camera. Since the FOE is on left of the image plane (at infinity), the function F has an decreasing behavior.
The photometric parameter used was the brightness value, in a range of 0-255. In all experiments presented before we applied a dissimilarity function with $\sigma = 5$ and the occlusion detection threshold $T = 5$. These values were chosen empirically according to the global distribution of the brightness along the sequences. In future, we plan to define automatically σ and T by using local properties of the image.

4 Conclusions

In this paper we study the importance of the occlusions for motion detection. Based on a theoretical framework for the definition of occlusions in the continuous case, we develop a dissimilarity function for the discrete case, using local photometric and geometric properties of the image. Assuming a moving monocular camera, we show that the occlusion classification is equivalent to the detection of the translational direction. Thus, we design a method to recover egomotion information, according to the following observations:

- Occlusions are important cues for the motion segmentation and the egomotion perception.
- In a moving camera, only translation produces occlusion points. Therefore, the rotation does not influence the translational estimation.
- To detect translation, no special models for motion or structure are needed.
- The translation can be detected even if its projection is outside the image field.
- Naturally, without occlusion points the approach is not applicable.

A number of experiments with real images have been presented, for various kinds of motion, that illustrate the capabilities of our approach.
As future work, we plan (1) to extend the method for color images, developing an appropriate photometric function; (2) to incorporate the occlusion information of all image directions in a global function in order to estimate robustly the FOE location; (3) to study a scale-space based approach to define automatically the fixed parameters of the method (size of masks, filtering parameters, thresholds). We intend also (4) to use the occlusion cues for a navigation system, associated with other sources of information.

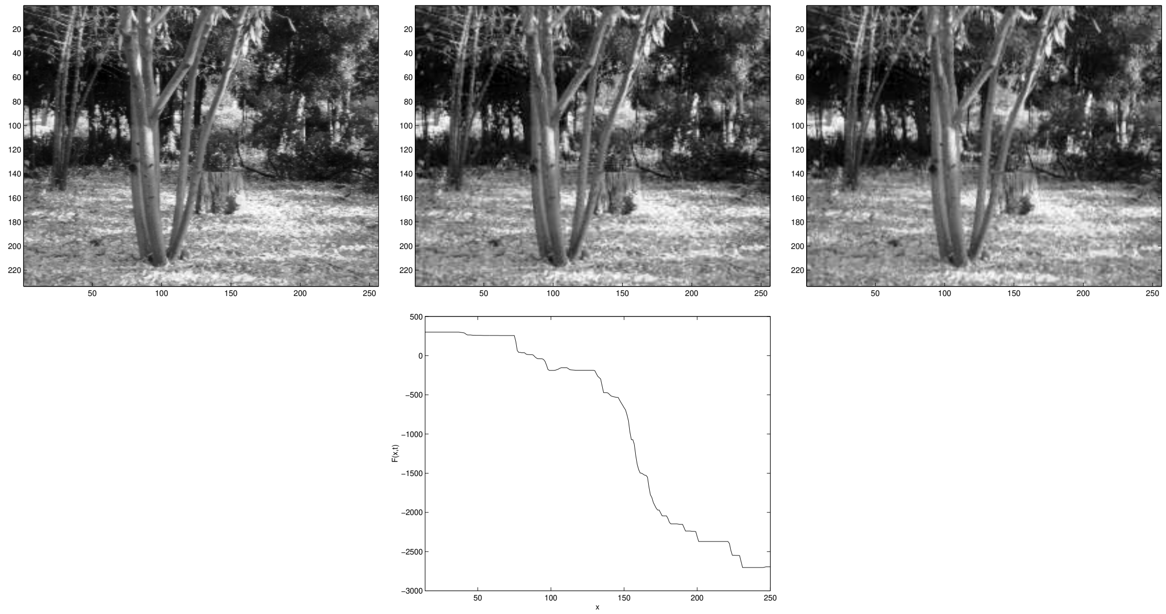


Fig. 5.: Three sequential samples of the Tree Sequence. As expected, the function $F(x, y)$ has a decreasing behavior with a maximum at the beginning of the x axis (then the FOE is on the left of the image field).

Acknowledgments

This work was partially funded by the EU ESPRIT-LTR proj. 30185.NARVAL and a grant from the PRAXIS-XXI research program.

References

1. B. L. Anderson and K. Nakayama. Toward a general theory of stereopsis: Binocular matching, occluding contours, and fusion. *Psych. Review*, (101):414–445, 1994.
2. S. Beauchemin, A. Chalifour, and J. Barron. Discontinuous optical flow: Recent theoretical results. In *Vision Interface (VI97)*, Kelowna, B. C., May 1997.
3. U. R. Dhond and J. K. Aggarwal. Stereo matching in the presence of narrow occluding objects using dynamics disparity search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(7):719–724, July 1995.
4. B.K.P. Horn and B. Shunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.
5. Y. Ohta and T. Kanade. Stereo by intra-and-inter-scanline search using dynamic programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(2):139–154, March 1985.
6. W. B. Thompson. Exploiting discontinuities in optical flow. *International Journal of Computer Vision*, 30(3):163–174, 1998.
7. C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proc. of the Sixth International Conference on Computer Vision*, Bombay, India, January 1998.