Motion from Occlusions*

César Silva, José Santos Victor

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

Abstract

In computer vision, occlusions are almost always seen as undesirable singularities that pose difficult challenges to image motion analysis problems, such as optic flow computation, motion segmentation, disparity estimation, or egomotion estimation. However, it is well known that occlusions are extremely powerful cues for depth or motion perception, and could be used to improve those methods.

In this paper, we propose to recover camera motion information based uniquely on occlusions, by observing two specially useful properties: occlusions are independent of the camera rotation, and reveal direct information about the camera translation.

We assume a monocular observer, undergoing general rotational and translational motion in a static environment. We present a formal model for occlusion points and develop a method suitable for occlusion detection. Through the classification and analysis of the detected occlusion points, we show how to retrieve information about the camera translation (FOE). Experiments with real images are presented and discussed in the paper.

Keywords: Egomotion estimation; occlusion analysis; image motion analysis; stereo matching

1. Introduction

Finding correspondences is one of the 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 occlusion is one of the most powerful cues to perceive depth and motion, and influence the earliest visual stages of stereo matching. Figure 1 illustrates this idea with samples of a well-known image sequence, where the occlusions (together with the image motion) give a clear perception of depth and camera motion.

Nevertheless, in computer vision, occlusions are 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 either multiple image motions [12,2,10,7] or the disparity field of a stereo pair [8,3,4,6].

In this paper we do not focus on the explicit estimation of image motion (or disparity) but rather, on the role played by occlusions in motion perception. Assuming a moving monocular observer, we study the relation between the observable occlusions and the camera motion. Thus, we will show how the occlusions provide, per se, fundamental information for motion estimation.

To approach this problem, we first develop a sufficient condition for the existence of an occlusion, and thereafter a direct relation between camera translation and occlusion classification is presented. Finally, we report some results with

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

Email addresses: {cesar, jasv}@isr.ist.utl.pt

Figure 1. Three sequential samples of the Flower Garden Sequence. The camera is going to the right. Occlusions and image motion give both a depth and camera motion perception.

real image sequences.

2. A Definition for Occlusion Points

In the previous section, we have argued that occlusions convey important information about the camera motion and scene structure. In this section we propose a formal definition of occlusion points and a methodology for detecting occlusions in image sequences.

Geometrically, an occlusion is caused by an occluding surface moving in front of an occluded surface. Additionally, if the observer is moving in a static environment, occlusions correspond to discontinuities both in the perceived motion and depth. However, unless we impose prior models to the image motion field, 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 point.

Hence, an occlusion has to be studied both as a geometric and photometric phenomenon. We propose to define occlusion points through a sufficient condition based on a local photometric dissimilarity over time, with precise geometric properties. This sufficient condition can be characterized rigorously for the continuous case.

First of all, we denote the spatial and temporal coordinates of an image sequence (see Figure 2(a)(b)) by x and t, represented by a vector $k = \binom{x}{t}$, where x is the spatial coordinate of an arbitrary scanline of the image. Additionally let us define the following auxiliary sets of space-time coordinates, representing two halves of a circle in x-t-space[2]:

$$\begin{aligned} K^+ &= \{k : \|k\| = 1 \wedge t > 0\} \\ K^- &= \{k : \|k\| = 1 \wedge t < 0\} \end{aligned}$$

As we have already discussed, an occlusion point corresponds to photometric values that *appear* and *disappear* between frames. Let $f(k)$ denote such photometric measure of the image in k (for example the brightness value). Based on this notation, we present the following sufficient condition for the existence of an occlusion.

The point $k_0 = (x_0,\ t_0)$ is an emergent occlusion if

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

Similarly, k_0 is a submergent occlusion point if

$$\begin{aligned} &\exists k^- \in K^- : \forall k^+ \in K^+, \\ &\quad \lim_{\gamma \to 0^+} f(k_0 + \gamma k^+) \neq \lim_{\gamma \to 0^+} f(k_0 + \gamma k^-). \end{aligned} \tag{2}$$

Figure 2(a) illustrates the meaning of the sufficient condition proposed here, with a simple ex-

[2] Assume for simplicity that the units of x and t are meters and seconds respectively.

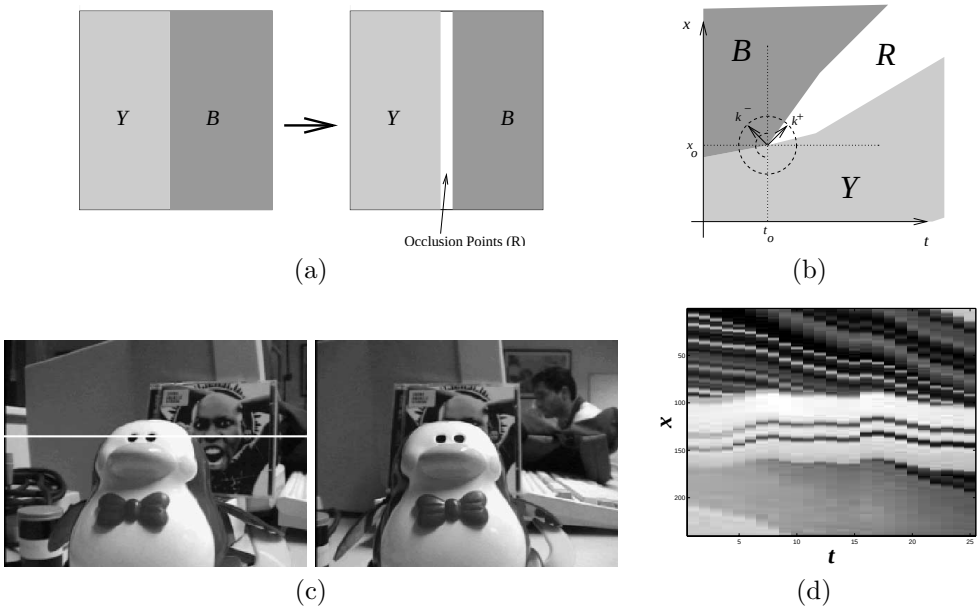


Figure 2. (a) Emergent Occlusion Point: Surface R appears between surfaces Y and B; (b) Image motion over time t, for a horizontal line parameterized by x. (c) Example of a real sequence with translation and rotation; (d) Image motion over time, for the horizontal scanline selected in the left pictures.

ample of an emergent occlusion point. The surface R appears between surfaces Y and B. Considering a given horizontal scanline, the surface R emerges at the point $(x_0,\ t_0)$, as shown in Figure 2(b). According to the condition defined before, this point is an emergent occlusion point, because there is a vector k^+ (with positive t) associated to a region (R), which photometric value does not exist on the half-plane $t < t_0$, in a neighborhood of the point (x_0, t_0). Figure 2(c) shows an example of a real sequence, where the submergent and emergent occlusion surfaces are visible on the temporal evolution of a given horizontal scanline (Figure 2(d)).

This sufficient condition is useful as formal model for a generic occlusion definition. However, in order to guarantee its applicability in the discrete case, we have to define more carefully the associated inequality relation. Consequently, we have developed a dissimilarity criterion inspired on a function developed by Tomasi and Manduchi [11] that was originally designed to smooth a single image, preserving the photometric discontinuities. We have changed this function in order to measure the similarity between two consecutive frames.

Suppose that 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 (preferably 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 $f(x_1, t_1)$:

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

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 pixel x_0 in frame t_1:

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

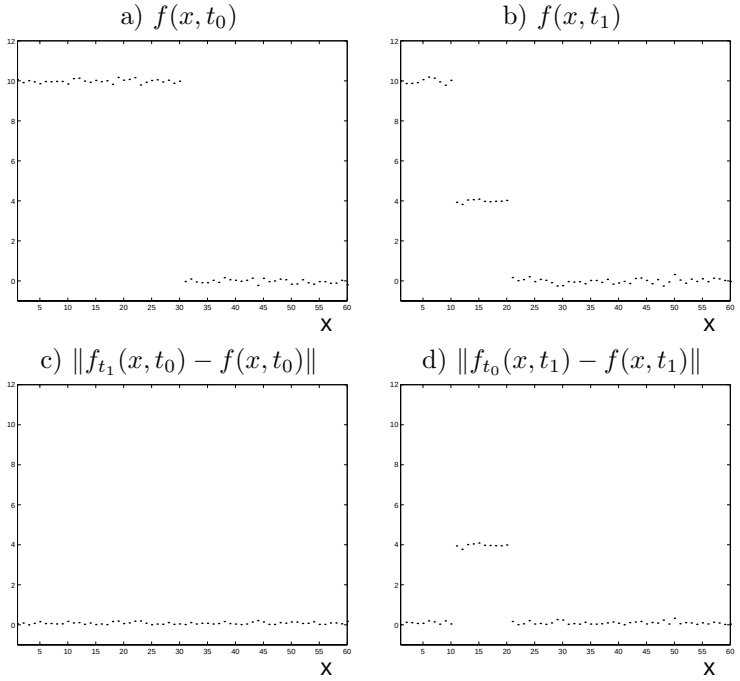


Figure 3. (a) Brightness function of a scanline in frame t_0; (b) Brightness function of the same scanline in frame t_1, after applying a shift in x, adding some gaussian noise in the brightness axis and introducing new brightness information (simulating an occlusion). (c-d) Computation of $\|f_{t_1}(x, t_0) - f(x, t_0)\|$ and $\|f_{t_0}(x, t_1) - f(x, t_1)\|$ for all x in frames t_0 and t_1 respectively. We assumed that $V(x)$ is the set of points within the interval $[x - 20 \;\; x + 20]$ and $\sigma = 1$.

When $\sigma \rightarrow 0$, 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 based on the photometric range of the images.

Figure 3 illustrates the performance of this dissimilarity criterion in detecting the presence of an occlusion between two simple functions (Figures 3a-b) which differ additionally by a translation. By subtracting directly both functions, we cannot detect easily the occlusion, because the difference is affected indistinctly by both occlusions and translation. On the contrary, by using the proposed approach, we can detect exactly the dissimilar region which corresponds to the occlusion region. Hence, $\|f_{t_1}(x, t_0) - f(x, t_0)\|$ (Figure 3c) is almost zero for all x, meaning that all photometric information in frame t_0 is present in frame t_1. On the other hand, the computation of $\|f_{t_0}(x, t_1) - f(x, t_1)\|$ (Figure 3d) shows that there is a region in frame t_1 which is dissimilar from the information present in frame t_0, revealing then an occlusion region. Furthermore, if the frame t_1 appears temporally after the frame t_0 ($t_1 > t_0$) then we can conclude that the detected occlusion is emergent, otherwise ($t_1 < t_0$) the occlusion is submergent. Notice that the occlusion detection only becomes effective if an adequate threshold is applied (in the example, $T = 1$ is an acceptable value).

3. Occlusions and Egomotion Perception

In the previous section we have proposed a formal definition of *emergent* and *submergent* occlusion points, together with a photometric criterion for their detection.

In this section we analyze how these occlusion points can be used to retrieve information regarding the observer's 3D motion (egomotion). We consider a monocular observer under a perspective camera model, moving with arbitrary translation and rotation, in a scene with static objects.

Associated to the egomotion estimation problem, one observes one of the most important properties of the occlusions:

Property 1 — *The camera rotation does not produce occlusion points. Consequently, the occlusion points are uniquely due to the camera translation.*

Property 1 states a well known fact. Only the translational part of the image motion depends on the scene depth. As occlusions are produced by depth discontinuities, only the translational component of the camera motion will give rise to occlusion effects.

Notice that one of the most important difficulties when estimating the camera motion using optic flow consists in decoupling the effects of translation from those of rotation [9]. This problem does not exist when considering occlusions.

The translation of an observer is usually 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, 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 that line. The flow $v(x)$ can be represented as a function of the camera motion parameters [5], as follows:

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

where $r(x)$ is the motion component due to the camera rotation, x_{FOE} 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.

Assuming that x_0 is an occlusion point, one observes a discontinuity in $v(x_0)$ and a discontinuity in $Z(x_0)$ — this means $v(x_0^-) \neq v(x_0^+)$ and $Z(x_0^-) \neq Z(x_0^+)$ respectively. However these discontinuities have a different physical meaning as described by the following two properties:

Property 2 : Emergent/Submergent Occlusion
When x_0 is an emergent occlusion point, $v(x_0^-) < v(x_0^+)$; when x_0 is a submergent occlusion point, $v(x_0^-) > v(x_0^+)$.

Property 3 : Left/Right Occlusion
If $Z(x_0^-) > Z(x_0^+)$ then the occluding surface is on the right[3] 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.

Figure 4 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 ($v(x_0^-) < v(x_0^+)$).

In summary, 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.

In the following property we show the direct relation between the camera translation (measured on the scaline by x_{FOE}) and the occlusion classification presented before.

Property 4 : Fundamental Relation between Camera Translation and Occlusions

- *An occlusion point x_0 is on the right of x_{FOE} if either (1) x_0 is emergent and the occluding surface is on the left side, or (2) x_0 is submergent and the occluding surface is on the right side.*

[3] stipulating x_0^+ on the right of x_0^-.

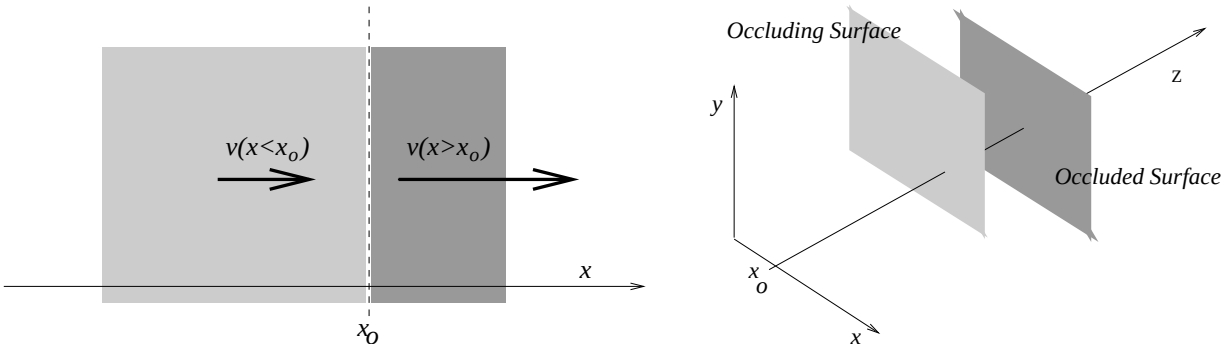


Figure 4. Example of an occlusion situation: (a) x_0 is an emergent occlusion point; (b) the projected occluding surface is on the left of x_0.

- *The occlusion point x_0 is on the left of x_{FOE} if either (1) x_0 is emergent and the occluding surface is on the right side, or (2) x_0 is submergent and the occluding surface is on the left side.*

Proof: To prove the property described above, let us define a notation for the classification of occlusion points, based on functions $\mathcal{L}(x_0)$ and $\mathcal{E}(x_0)$ described as follows:

- $\mathcal{L}(x_0) = 1$ or $\mathcal{L}(x_0) = -1$ if the occlusion point x_0 has respectively a left or a right occluding surface;
- $\mathcal{E}(x_0) = 1$ or $\mathcal{E}(x_0) = -1$ if x_0 is respectively an emergent or a submergent occlusion point;
- $\mathcal{L}(x_0)$ and $\mathcal{E}(x_0)$ are zero if x_0 is not an occlusion point.

By using Properties (2, 3) and Equation (4), then we can determine whether the FOE (x_{FOE}) is located to the left or right of the occlusion point x_0:

$$\mathcal{L}(x_0) \cdot \mathcal{E}(x_0) = 1 \Rightarrow$$
$$\Rightarrow \left(v(x_0^-) - v(x_0^+)\right)\left(Z(x_0^-)^{-1} - Z(x_0^+)^{-1}\right) < 0 \Rightarrow$$
$$\Rightarrow x_0 < x_{\text{FOE}}$$

$$\mathcal{L}(x_0) \cdot \mathcal{E}(x_0) = -1 \Rightarrow$$
$$\Rightarrow \left(v(x_0^-) - v(x_0^+)\right)\left(Z(x_0^-)^{-1} - Z(x_0^+)^{-1}\right) > 0 \Rightarrow$$
$$\Rightarrow x_0 > x_{\text{FOE}}$$

QED

This result shows that classifying the occlusion point x_0 corresponds to detecting its location relatively to a projection of the FOE.

Moreover the occlusions are not affected by the camera rotation as described by Property 1. This is a huge advantage when compared to other approaches that use the optic flow to estimate the egomotion, where decoupling the rotation and translation components is a difficult problem.

However it remains the question about the algorithmic procedure to classify the occlusion point. In fact, the occlusion classification could be performed by the optic flow and depth description in a certain neighborhood. Assuming that both the optic flow and depth are unknown (and hardly computable), we propose to use exclusively the dissimilarity function developed in Section 2.

First of all, remind that Section 2 describes a method to classify an occlusion as emergent or

submergent. This alleviates the need to explicitly determine the local optic flow $v(x_0^-)$ and $v(x_0^+)$. Secondly, to determine whether we have a left or right occluding surface, we monitor the temporal photometric changes at the left and right side of the occlusion point. The obvious 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. An alternative equivalent procedure consists of studying the evolution of points which do not preserve their photometric properties over time, thus belonging to the occluded surface. Notice that the complete occlusion classification can rely uniquely on the dissimilarity criterion presented before.

In order to detect automatically the location of the FOE projection along a scanline, we designed a function that integrates, along x, the value of the dissimilarity relations from (3) taking into account $\mathcal{L}(x) \cdot \mathcal{E}(x)$. This function can be described for the discrete case as follows:

$$\begin{aligned} F(x,t) &= \sum_{\zeta=-\infty}^{x} \mathcal{L}(\zeta)\mathcal{E}(\zeta)d(\zeta,t) \\ d(\zeta,t) &= \|f_{t+1}(\zeta,t) - f(\zeta,t)\| + \\ &+ \|f_{t-1}(\zeta,t) - f(\zeta,t)\| \end{aligned}$$

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.

4. Results

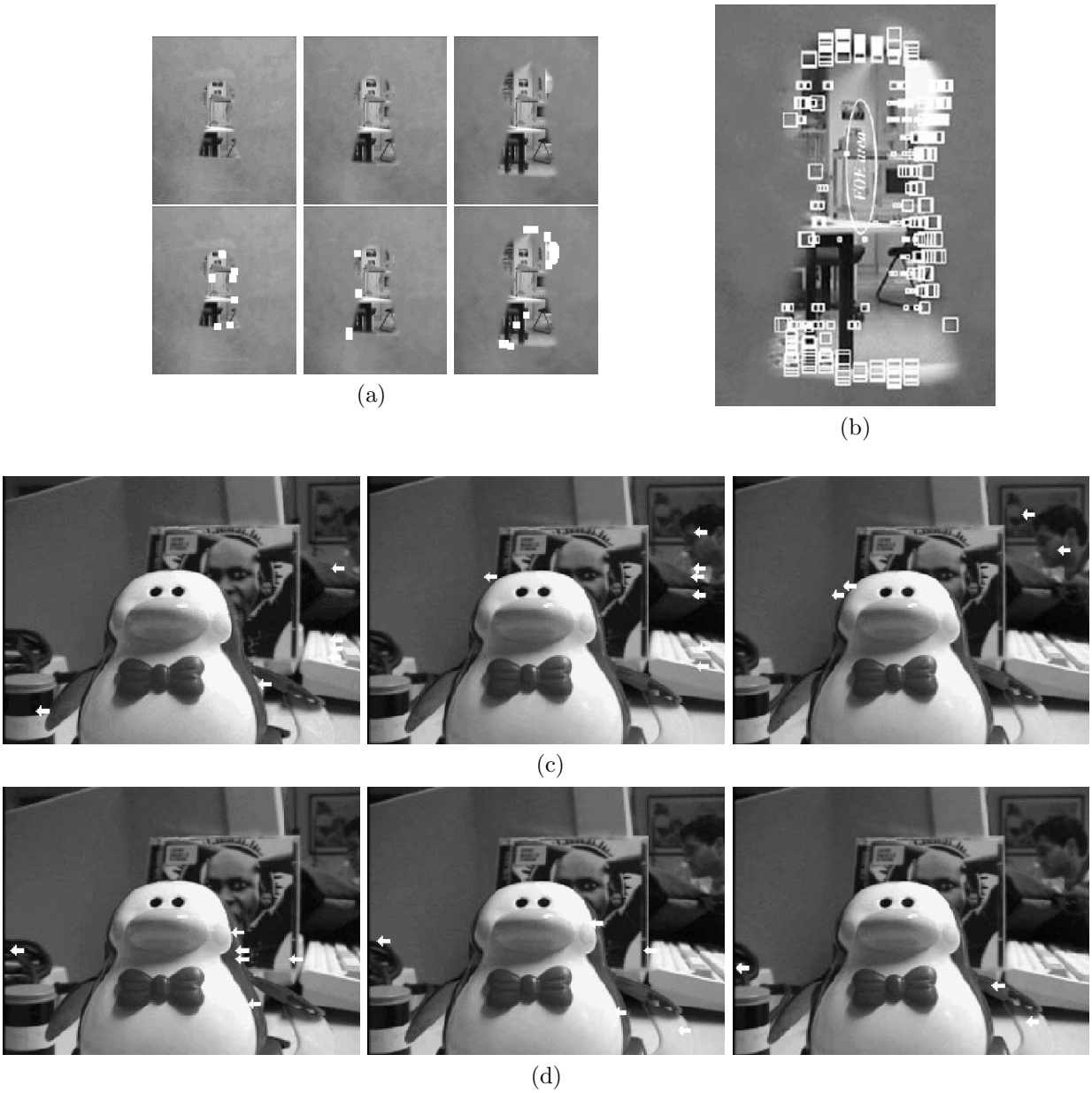
In this Section, we apply the occlusion detection and classification process to four image sequences. The first sequence (the Lock Sequence, Figure 5(a)) shows the performance of the dissimilarity function in order to find the occlusion points. The Focus of Expansion is roughly at the center of the image and emergent occlusions are found on the boundaries of the lock hole, as expected. Considering an arbitrary line along the image, the relation between the occlusions and the FOE location can be observed: first the occlusions are emergent, second the occluded points are inside the lock hole (or the occluding surface is outside), thus completing the occlusion classification and indicating that the FOE is somewhere in a restricted area at the center of the image (Figure 5(b)).

The second sequence (the Penguin Sequence, Figure 5(c)) was performed with static objects and a leftward moving camera with rotation. In this experiment we show how the occlusion points are immune to rotational contamination of the image motion. In Figure 5(c) we present three samples of the sequence where the emergent occlusions appear mainly on the left side of the penguin whereas the submergent occlusions disappear on the right side (Figure 5(d)). This classification indicates that the FOE is on left of each occlusion point detected.

The third sequence (Figure 6-left) 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 (summing the contributions of all $F(x,\ t)$ over the vertical coordinate). Notice both the occlusion boundaries of the penguin and its velocity direction, given by the decreasing behavior of the function.

The last sequence (the Tree Sequence, Figure 6-right) consists in a leftward moving camera, with a large number of occlusions. Since the FOE is on the left of the image (at infinity), the function F has a decreasing behavior.

The photometric parameter used here was the brightness value, in a range of 0-255. In all experiments, we applied the same 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 the future we plan to define σ and T automatically using local properties of the image.



(a)

(b)

(c)

(d)

Figure 5. (a) Three sequential samples of the Lock Sequence and associated emergent occlusion points (white squares). (b) Last frame of the Lock Sequence with previously detected emergent occlusion points superimposed on the image (more recent ones represented by larger squares). The ellipse in the figure illustrates qualitatively the expected region for the FOE location. (c) Samples of the Penguin Sequence, with associated emergent occlusion points. Each occlusion point found indicates the same FOE direction (arrow direction) given by its classification. (d) Same images with submergent occlusion points.

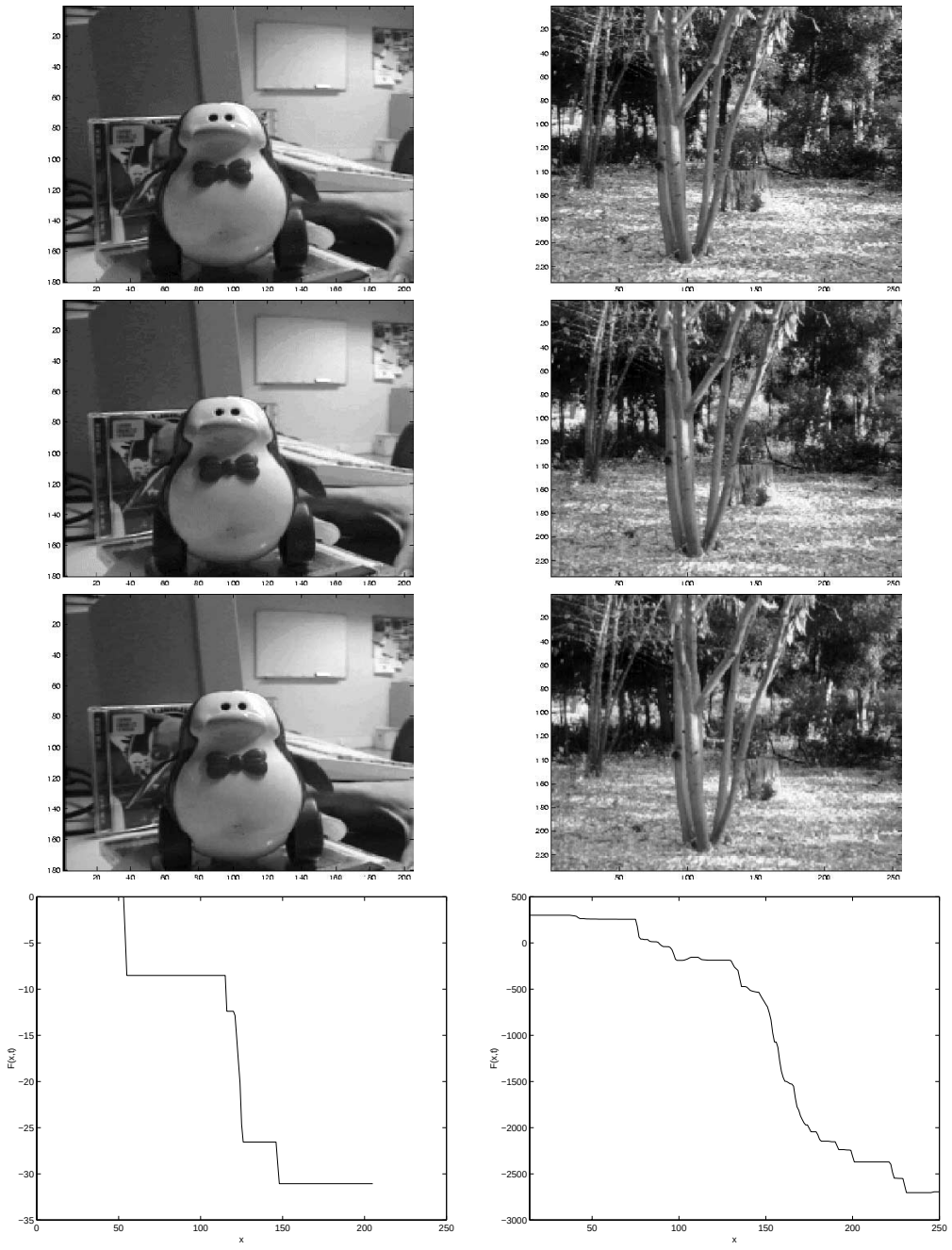


Figure 6. Left: Three sequential samples of the Penguin Sequence — on the bottom the function $F(x, y)$ along x (notice that the decreasing steps correspond to penguin occlusion boundaries). Right: Three sequential samples of the Tree Sequence. As expected, the function $F(x, y)$ (bottom) has a decreasing behavior whith a maximum at the beginning of the x axis.

5. Conclusions

In this paper we have studied the importance of the occlusions for motion detection. Based on a theoretical framework for the definition of occlusions in the continuous case, we developed 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 a translational direction. Thus, we design a method to recover egomotion information, according to the following observations:

- Occlusions are extremely important cues for the egomotion perception.
- With a moving camera, only translation produces occlusion points. Therefore, the rotation does not influence the translational estimation.
- To detect the camera translation, no special models for motion or structure are needed.
- The camera translation can be detected even if its projection is outside the image field.
- The occlusion classification can be performed by using uniquely dissimilarity criteria (more robust than similarity criteria).

A number of experiments with real images have been performed, for various kinds of motion, that illustrate the capabilities of our approach.

As future work, we plan to extend the method for color images, developing an appropriate photometric function, and to incorporate the occlusion information of all image directions in a global function in order to estimate robustly the FOE location. We intend also to use the occlusion cues for a navigation system, associated with other sources of local and global information.

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. D. Geiger, B. Landendorf, and A. Yuille. Occlusions and binocular stereo. *Int. Journal of Computer Vision*, 14(3):211–226, 1995.
5. B.K.P. Horn and B. Shunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.
6. H. Ishikawa and D. Geiger. Occlusions, discontinuities, and epipolar lines in stereo. In *Proc. of the 5th European Conference on Computer Vision*, Germany, June 1998.
7. H. Nagel and A. Gehrke. Spatiotemporally adaptive estimation and segmentation of offields. In *Proc. of the 5th European Conference on Computer Vision*, Germany, June 1998.
8. 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.
9. C. Silva and J. Santos-Victor. Robust egomotion estimation from the normal flow using search subspaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(9):1026–1034, September 1997.
10. W. B. Thompson. Exploiting discontinuities in optical flow. *International Journal of Computer Vision*, 30(3):163–174, 1998.
11. 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.
12. J. Wang and E. Adelson. Layered representation for motion analysis. In *Proc. of the IEEE Comp. Vis. and Pat. Rec. (CVPR)*, New York, USA, June 1993.