GEOMETRY-BASED ESTIMATION OF OCCLUSIONS FROM VIDEO FRAME PAIRS

Serdar Ince and Janusz Konrad

Department of Electrical and Computer Engineering, Boston University
8 Saint Mary's St., Boston, MA 02215

ABSTRACT

The knowledge of occlusions and newly-exposed areas, a natural consequence of changing object juxtaposition in a 3-D scene, can be effectively used to improve video coding efficiency, video rate conversion quality and view interpolation fidelity. Although various occlusion estimation methods have been proposed to date, most of them are not robust or are computationally complex. In this paper, we study two simple, well-known occlusion estimation methods, one based on a photometric mismatch between two frames of an image sequence, while the other based on a geometric mismatch. We demonstrate their weaknesses and propose a new geometric method that exhibits good robustness to noise in the data while maintaining low computational complexity.

1. INTRODUCTION

Occlusion effects occurring in image sequences are a natural consequence of changing object juxtaposition in a 3-D scene. These effects result in parts of an image frame disappearing in the following frame(s), known as *occlusion areas*, or appearing in the following frame(s), known as *newly-exposed areas*. Both types of areas play a very important role in motion estimation from dynamic imagery and in disparity estimation from stereo or multiview imagery; for frame points in occlusion areas forward motion is undefined (those points disappear in the next frame). Similarly, for frame points in newly-exposed areas backward motion is not defined. Consequently, motion parameters should not be computed for image points belonging to either type of area, as they are meaningless. However, and this is the second observation, most motion estimation algorithms employ some form of regularization (explicit motion smoothness prior, block-based motion model, intensity matching over a window, etc.). Since in occlusion and newly-exposed areas motion parameters are undefined, regularization should be disallowed between image points from those areas and neighboring points with well-defined motion. In order to achieve this, occlusion and newly-exposed areas must be explicitly known. Similar observations apply to disparity estimation.

Estimation of occlusion and newly-exposed areas is an inverse problem and, as such, is ill-posed. Most of occlusion/ newly-exposed area estimation methods rely on 3 or more frames to make decision about individual image points [1, 2, 3, 4, 5]. Such methods compare intensity consistency between the current and previous frame(s) with that between the current and future frame(s). In general, this improves reliability of occlusion estimation but requires larger buffers and is more complex computationally. Methods have been proposed that estimate newly exposed areas from two image frames only. However, such methods based on photometric detection mechanism (intensity mismatch) [6, 7] are not reliable, while those based on geometric mechanism (motion vector mismatch) [8] although more reliable under high PSNR conditions still fail on noisy data.

In this paper, we propose a simple method for the detection of occlusion and newly-exposed areas that is based on geometric properties of the motion field. The method is applicable to any motion field derived from an image pair. Its principle is based on the observation that the regular grid in the reference image plane, at which the motion vectors are anchored, forms an irregular grid in the target image plane after motion compensation. Since the target image will contain no motion-compensated projections in the newly-exposed areas, such areas can be easily detected. We present a simple neighborhood test to detect newly-exposed pixels and we compare our approach with standard photometry- and geometry-based methods.

2. PHOTOMETRY-BASED ESTIMATION OF OCCLUSIONS

The usual assumption in estimation of occlusions from two frames, is excessive intensity matching (motion-compensated prediction) error observed; reference-frame pixels that disappear cannot be accurately matched in the target frame and thus induce significant errors. Let $I_1[\mathbf{x}]$ denote intensity of the first frame of a sequence at spatial position $\mathbf{x}$, and $I_2[\mathbf{x}]$ – similar intensity in the second frame. If $\mathbf{d}_f$ denotes a forward motion (disparity) field anchored on the sampling grid of frame #1 (reference) and pointing to the target frame #2, while $\mathbf{d}_b$ denotes a backward motion field, then the corresponding motion-compensated prediction errors at $\mathbf{x}$ are:

$$\begin{aligned} \varepsilon_f[\mathbf{x}] &= I_1[\mathbf{x}] - I_2[\mathbf{x} + \mathbf{d}_f[\mathbf{x}]], \\ \varepsilon_b[\mathbf{x}] &= I_2[\mathbf{x}] - I_1[\mathbf{x} - \mathbf{d}_b[\mathbf{x}]]. \end{aligned}$$

The usual occlusion detection methods then declare a pixel in the reference frame as being occluded in the target frame if $|\varepsilon_f| > \Theta$ for frame #1 and $|\varepsilon_b| > \Theta$ for frame #2. Note that although newly-exposed areas cannot be detected by this mechanism (pixels are not visible), effectively the occluded areas in frame #2 (computed using $\mathbf{d}_b$) are in fact the newly-exposed areas for frame #1.

3. GEOMETRY-BASED ESTIMATION OF OCCLUSIONS – TRADITIONAL APPROACH

An alternative, to the photometric detection of occlusion areas, is a geometric detection. Such a detection is based on the assumption that a mismatch of forward and backward motion vectors is due to disappearing image areas. In particular, the following vector

This work was supported by the National Science Foundation under grant ECS-0219224. E-mail: {ince,jkonrad}@bu.edu

matching errors have been used in the past to detect the occlusion and newly-exposed areas in reference frame I_1 [8]:

$$\begin{aligned}\rho_f[\mathbf{x}] &= \|\mathbf{d}_f[\mathbf{x}] - \mathbf{d}_b[\mathbf{x} + \mathbf{d}_f[\mathbf{x}]]\|, \\ \rho_b[\mathbf{x}] &= \|\mathbf{d}_f[\mathbf{x} - \mathbf{d}_b[\mathbf{x}]] - \mathbf{d}_b[\mathbf{x}]\|.\end{aligned}$$

By comparing the above errors with a threshold, decision can be made as to whether pixel at position $\mathbf{x}$ in I_1 is occluded ($\rho_f > \Delta$) or exposed ($\rho_b > \Delta$). For increased robustness, this decision can be averaged over a window, however this will sacrifice resolution of the method.

4. GEOMETRY-BASED ESTIMATION OF OCCLUSIONS – NEW APPROACH

We propose occlusion/newly-exposed area detection based on another geometric principle. A typical motion field computed under some form of spatial regularization will lead to converging motion vectors originating in occlusion areas of the reference frame (area A in Fig. 1). Such vectors cannot provide a good intensity match and assume compromise coordinates with respect to the neighboring vectors from, e.g., a moving object and static background. This convergent behavior is a compromise between the lack of intensity match and spatial smoothness enforced, and potentially leads to multiple vectors pointing to the same location in the target frame. This might suggest that a high spatial density of motion-compensated positions in the target frame (I_2 in Fig. 1) is indicative of an occlusion area. However, in practice, it turns out that results are very sensitive the selected density threshold. On the other hand, pixels in the target frame that did not exist in the reference frame (newly-exposed pixels in area B) have no relationship with the reference frame and, as such, cannot be pointed to by forward motion vectors. Thus, areas in the target frame that are void of motion-compensated projections can be relatively easily detected. This is the basis of the proposed detection algorithm.

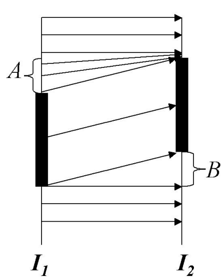


Fig. 1. Simple occlusion process and typical motion field; A – area to be occluded, B – area newly exposed.

The detection algorithm is very simple, and is equally applicable to occlusion detection if I_2 is the reference frame and I_1 is the target frame. Let Λ be a 2-D sampling lattice for I_1 and I_2 limited to the domain of each image. This is unlike the standard definition of a lattice that does not constrain its extent. Also, let $\mathcal{S}$ be a set irregular spatial positions in I_2 obtained by motion compensation of pixels from I_1, i.e., $\mathcal{S} = \{\mathbf{y} : \mathbf{y} = \mathbf{x} + \mathbf{d}_f[\mathbf{x}], \mathbf{x} \in \Lambda\}$. Note that $card\{\Lambda\} \geq card\{\mathcal{S}\}$ because certain points from I_1 may project to the same location in I_2. Define an indicator function as follows:

$$\xi_i(\mathbf{x}) = \begin{cases} 1, & ||\mathbf{x} - \mathbf{z}_i|| \leq r \\ 0, & otherwise \end{cases} \quad \mathbf{x} \in \Lambda, \mathbf{z}_i \in \mathcal{S},$$

For each lattice point $\mathbf{x}$ and irregular point $\mathbf{z}_i$, both in I_2, $\xi_i(\mathbf{x})$ is 1 if $\mathbf{z}_i$ is within a disk of radius r from $\mathbf{x}$. By accumulating $\xi_i(\mathbf{x})$:

$$M(\mathbf{x}) = \sum_{i=1}^{card\{\mathcal{S}\}} \xi_i(\mathbf{x}),$$

we measure the local density of motion-compensated projections at each $\mathbf{x} \in \Lambda$, and by thresholding $M(\mathbf{x})$ we find which areas of I_2 exhibit the lowest density of such projections: $\mathbf{x}$ is declared newly-exposed if $M[\mathbf{x}] < \Psi$, i.e., if sufficiently few irregular points are in the vicinity of $\mathbf{x}$. We use r=2, but we test a range of values of Ψ. For areas where the motion field $\mathbf{d}_f$ is uniformly translational (regular projections), $M(\mathbf{x}) = 13$ for $r = 2$. At $M(\mathbf{x}) = 6$ more than half of the projections are missing suggesting vicinity of a newly-exposed area.

Since it is easier to find regularly-spaced neighbors than those spaced irregularly, the algorithm is implemented differently in practice. For each projection $\mathbf{z}_i \in \mathcal{S}$, its neighbors $\mathbf{x} \in \Lambda$, such that $\|\mathbf{z}_i - \mathbf{x}\| \leq r$, are found, and each neighbor's counter is incremented by 1. After all $\mathbf{z}_i$ have been scanned, each counter contains the number of projections within distance of r.

5. EXPERIMENTAL RESULTS

In all the results shown, motion was computed using 8×8 block matching under spatial regularization (neighboring blocks are encouraged to have similar motion vectors). The resulting vector fields are diffuse on the occlusion side and have sharp boundary on the newly-exposed side of the moving object (left column of Fig. 2 and middle row in Fig. 4). In Fig. 2, we show results of experiment with synthetic motion of natural intensities. We measure the accuracy of detection of occluded and newly-exposed areas using symmetric difference between the ground-truth pixels and the detected pixels (union of false-positives and misses), shown in the center column of Fig. 2 as a function of a threshold parameter for each detection method (Θ, Δ or Ψ). For each method, we show detection result for parameter value with the lowest detection error.

Clearly, the photometric approach provides a globally-correct result that is locally very fragmented; extension of the method to a window instead of single pixel would solve this but at the cost of significant resolution loss. The two geometric approaches result in similar occlusion/newly-exposed area descriptors, but the one based on vector difference leaves gaps in otherwise compact regions. In terms of the detection error the new geometric approach outperforms the traditional one by close to 10%. As shown in Fig. 3, the photometric approach performs very poorly under noisy conditions. This is not unexpected since the detection is based directly on (noisy) intensities; using a window, again, would sacrifice resolution. The traditional geometric approach also fails in the presence of noise; disregarding the effects at image boundaries (vectors are incorrect due to the selected image boundary handling), the new method results in much more accurate estimates.

We also applied the proposed method to some well-known test sequences. As can be seen in Fig. 4, relatively accurate occluded and newly-exposed areas were obtained on *Flowergarden* and *Map* using this very simple, fast method. The results are not as accurate on *Tsukuba* and *Teddy* because of their relative complexity; the detected areas are in correct positions but are very fragmented. The accuracy of detection results is directly related to the quality of computed motion; better results should be possible with more sophisticated motion estimation than block-based.

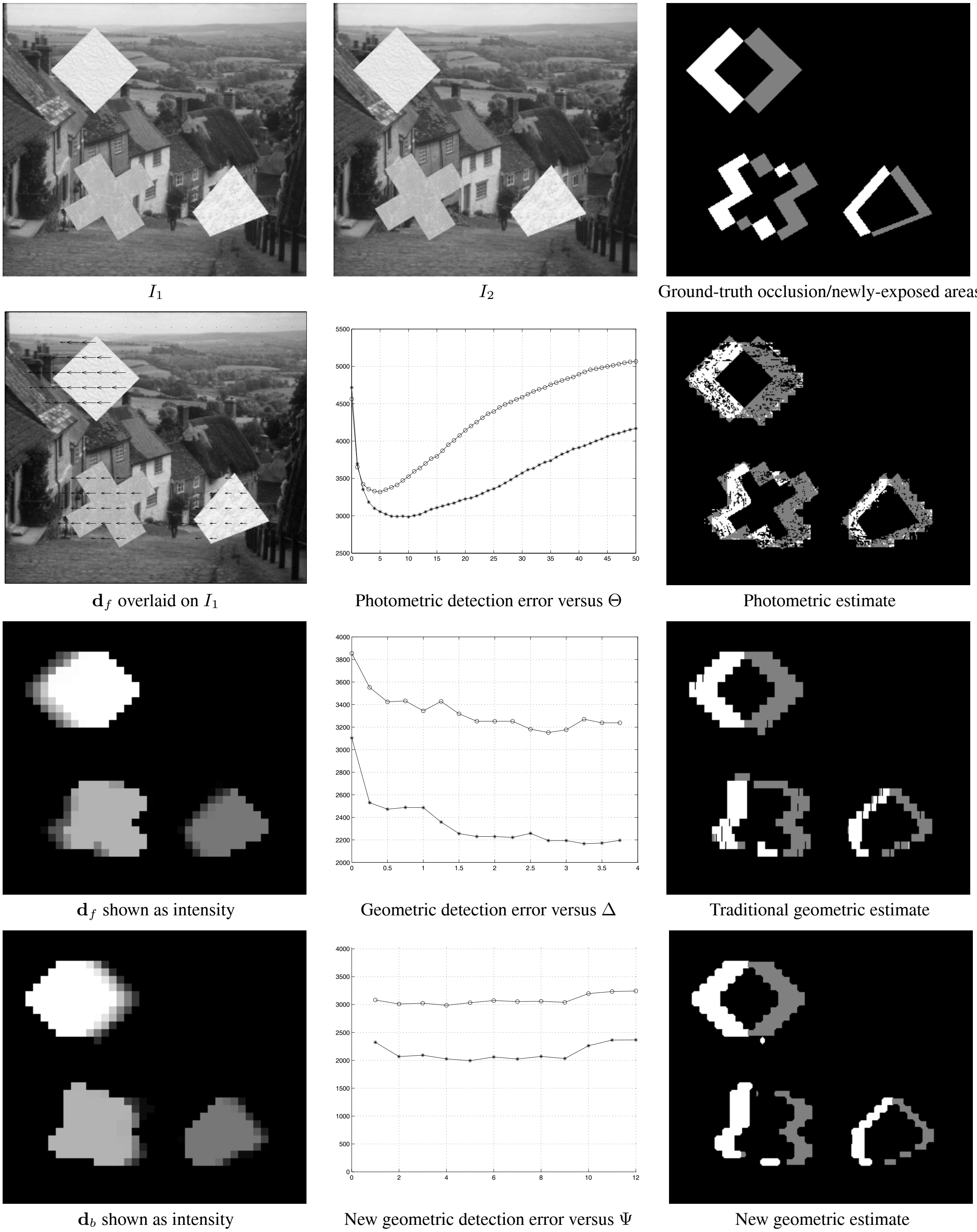
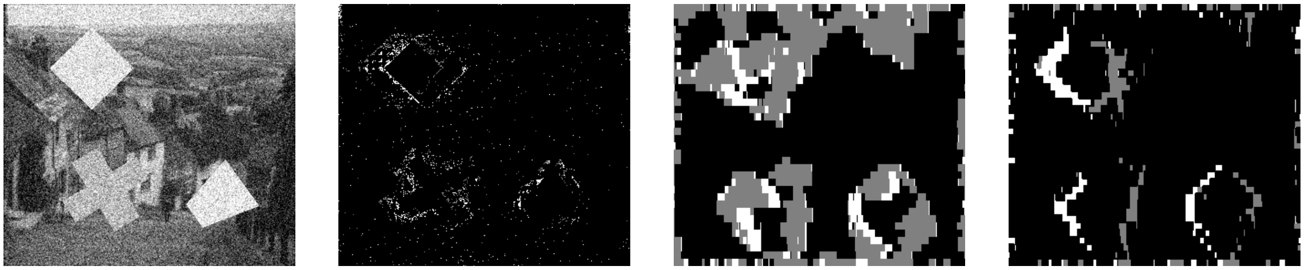


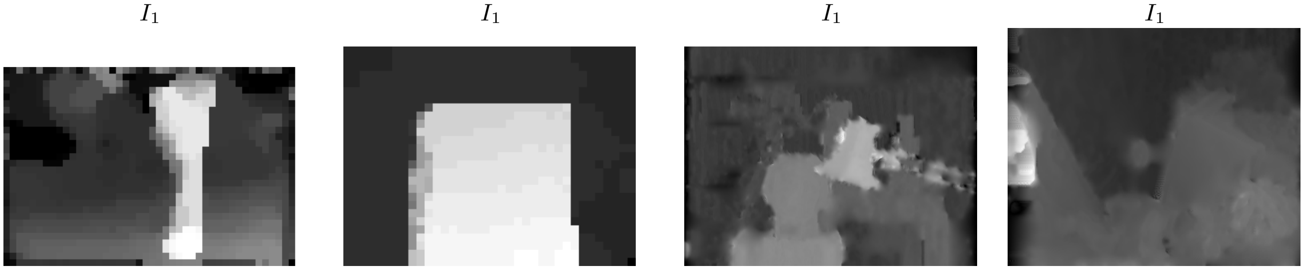
Fig. 2. Occlusion estimation results for a natural-texture, synthetic-motion sequence. In the middle column, two error plots are included, one for detection frome #1 to frame #2, and the other – from frame #2 to frame #1. In the right column, white denotes occluded area, and gray denotes newly-exposed area.



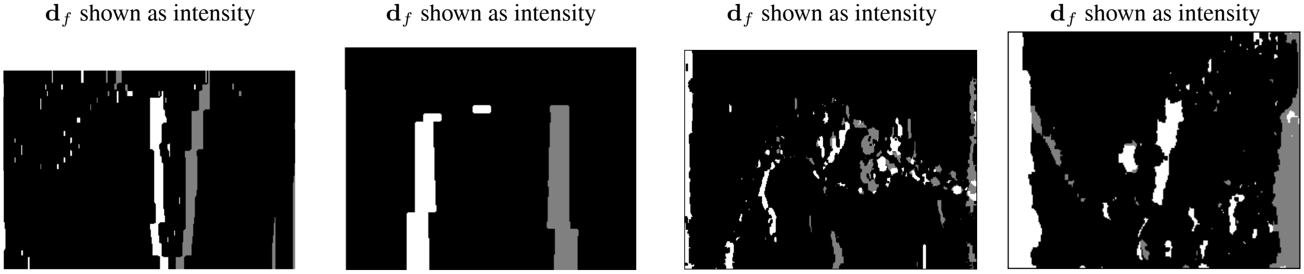
I_1 + noise | Photometric estimate | Traditional geometric estimate | New geometric estimate

Fig. 3. Results for the synthetic motion sequence with additive white Gaussian noise with standard deviation σ=36 (PSNR=17.44dB).

I_1 | I_1 | I_1 | I_1



$\mathbf{d}_f$ shown as intensity | $\mathbf{d}_f$ shown as intensity | $\mathbf{d}_f$ shown as intensity | $\mathbf{d}_f$ shown as intensity



New geometric estimate | New geometric estimate | New geometric estimate | New geometric estimate

Fig. 4. Occlusion estimation results for four well-known test sequences *Flowergarden*, *Map*, *Tsukuba* and *Teddy* (Last three test sequences are available at `www.middlebury.edu/stereo/`).

6. REFERENCES

[1] R. Depommier and E. Dubois, "Motion estimation with detection of occlusion areas," in *Proc. IEEE Int. Conf. Acoustics Speech Signal Processing*, Mar. 1992, pp. III.269–III.272.

[2] M. Chahine and J. Konrad, "Estimation and compensation of accelerated motion for temporal sequence interpolation," *Signal Process., Image Commun.*, vol. 7, no. 4–6, pp. 503–527, Nov. 1995.

[3] S.-I. Iu, "Robust estimation of motion vector fields with discontinuity and occlusion using local outliers rejection," *J. Vis. Commun. Image Represent.*, vol. 6, no. 2, pp. 132–141, June 1995.

[4] K.P. Lim, A. Das, and M.N. Chong, "Estimation of occlusion and dense motion fields in a bidirectional Bayesian framework," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 5, pp. 712–718, May 2002.

[5] M. Ristivojević and J. Konrad, "Joint space-time image sequence segmentation: object tunnels and occlusion volumes," in *Proc. IEEE Int. Conf. Acoustics Speech Signal Processing*, May 2004, vol. III, pp. 9–12.

[6] R. Thoma and M. Bierling, "Motion compensating interpolation considering covered and uncovered background," *Signal Process., Image Commun.*, vol. 1, pp. 191–212, 1989.

[7] J.N. Driessen and J. Biemond, "Motion field estimation for complex scenes," in *Proc. SPIE Visual Communications and Image Process.*, 1991, pp. 511–521.

[8] E. Izquierdo, "Stereo matching for enhanced telepresence in three-dimensional videocommunications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 4, pp. 629–643, Aug. 1997.