Computational Analysis of Binocular Half-Occlusions

Mikhail Sizintsev

Richard P. Wildes

Technical Report CS-2005-12

September 15, 2005

Department of Computer Science

4700 Keele Street North York, Ontario M3J 1P3 Canada

Computational analysis of binocular half-occlusions

Mikhail Sizintsev
Richard P. Wildes
Department of Computer Science and Engineering
and the Centre for Vision Research
York University
Toronto, Ontario M3J 1P3
Canada

Computational analysis of binocular half-occlusions

Mikhail Sizintsev
Richard P. Wildes
Department of Computer Science and Engineering
and the Centre for Vision Research
York University
Toronto, Ontario M3J 1P3
Canada

Abstract

Binocular half-occlusions arise as a foreground surface obscures a portion of a background surface to one of the two images in a binocular pair. Half-occlusions occur at the projections of surface boundaries; therefore, they are a potential source of information in delineating 3D objects. In this paper, a novel integration of two sources of information (constraints) is proposed for binocular half-occlusion detection. The first constraint comes as an analytic relationship between the change in disparity across a half-occluded region and region width. The second constraint comes in terms of expected values of binocular match scores in the vicinity of half-occlusions. To allow for uncertainty in inference, these two constraints have been combined in a Bayesian formulation. A corresponding algorithm for half-occlusion detection, given maps of binocular disparity and match score has been implemented in software and empirically evaluated on seven standard stereo pairs. For cases where direct comparisons are possible, the developed method outperforms previous approaches to half-occlusion detection.

1 Introduction

1.1 Motivation

Spatially displaced views of the world support the recovery of the 3D geometry of an imaged scene. Of possible viewing configurations, binocular imaging has been a particularly well researched situation as it provides the minimal multiview situation. Further, since binocular imaging reflects biological design, there is potential for cross fertilization between research in artificial and natural binocular stereo.

Significant strides have been made in the investigation of artificial [4, 36] as well as natural [18] binocular stereo. Still, outstanding problems persist: From a computational point of view areas of particular concern include poor speed-accuracy trade-offs, reliance on precise a priori calibration and poor reconstruction in the vicinity of 3D object boundaries. In this chapter, the concern is with improving reconstruction in the vicinity of 3D object boundaries. While humans appear to be able to make precise depth estimates in such situations [18], artificial systems are not capable of similar performance [4, 36]. Improved computational analysis of reconstruction near 3D object boundaries will shed light on how natural systems operate and also make 3D information practical for a variety of artificial vision system applications, e.g., precision robot manipulation, shape-based recognition and new view synthesis. A valuable source of information about 3D boundaries is provided by half-occlusion, where one view in a pair sees surface features that are occluded to the other view, see Figure 1. Half-occlusion usually occurs around object boundaries and other 3D scene discontinuities. As such, these points have great potential to aid in accurate reconstruction near 3D boundaries.

As early as Euclid, the basic geometric relationship that gives rise to half-cclusion was documented [7]. Further, the potential perceptual significance of binocular half-occlusion has been known at least since the time of Leonardo Da Vinci [34]. Much more recently, the fact that humans actually do exploit half-occlusions in making depth inferences was documented [25]. Subsequently, a great number of psychophysical studies of half-occlusion have supported their use by humans (see, e.g., [18] for review); however, the enabling computations remain unclear.

To forward understanding of half-occlusions, this paper presents research that distills their analysis and detection from other aspects of binocular stereo. That is, constraints are isolated that are indicative of half-occlusions, while making only generic assumptions about other aspects of the problem, e.g., calibration, correspondence and reconstruction. The two constraints of study concern the change of disparity across a region of half-occlusion and the expected quality of match within a region of half-occlusion. It is argued that these two constraints are complementary and their spatial conjunction is highly indicative of half-occlusion. Moreover, they encompass information that has been exploited across a significant number of previous approaches, that variously consider measurements of disparity gradient and poor match quality as indicative of half occlusion. The practical efficacy of the combined constraints for half-occlusion detection is illustrated via post-processing applied to the output of a simple area-based stereo matcher, as it allows the utility of the constraints to be studied apart from other aspects of binocular stereo and facilitates comparison to a recent empirical study of half-occlusion detection methods [12]. More generally, however, the constraints could be employed within other approaches to binocular analysis, e.g., as an

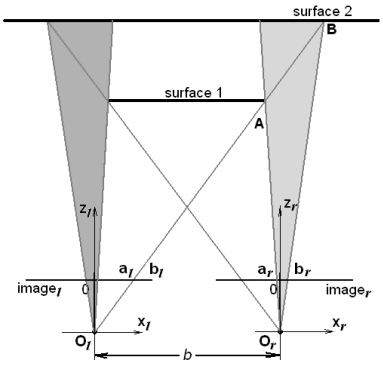

Figure 1: Geometry of Binocular Half-Occlusion. Interposition of surface 1 between surface 2 and the images causes portions of surface 2 that intersect the shaded regions to be visible to only one of the two views, i.e., half-occluded.

integral component of local or global correspondence algorithms.

1.2 Related research

Early work on computational stereo ignored half-occlusion or treated it as noise in the matching process [1]. Subsequently, a number of approaches to dealing with half-occlusions have emerged (see, e.g., [12, 4] for reviews and empirical comparison). Significantly, empirical comparison suggest that none of the compared approaches reliably identify the majority of half-occlusions while keeping false positives modest and no one approach is superior to the others.

Several more recent contributions to the literature of half-occlusions can be noted. The use of adaptive spatial support for match windows can ameliorate issues arising in attempts to match half-occluded regions by shaping windows to avoid poorly defined matches [21]; recent extensions increase the accuracy and efficiency of such processing [16, 13, 41]. Further, with more than two views, such adaptive processing can define which scene points are visible to which views [22, 30]. Other recent additions to the literature are based on the expected behavior of disparity gradient in the vicinity of half-occlusions [19, 15, 16, 38], e.g., the fact that occlusions in a left-based disparity map correspond to occluded regions in the right-based disparity map and vice versa (the "occlusion constraint"). The importance of disparity gradient as a constraint on allowable stereo matches has been known for some time (e.g., [6, 31]); however, its specific interpretation in terms of half-occlusion is relatively recent. Yet another approach rejects matches that are ambigous (in having rival candidates of similar cost) to diagnose occlusion [35]. Occlusion detection also has been bolstered by constraining occlusion boundaries to align with those of uniform colour segments [11]. Another recent addition to the literature involves interleaved processes of layered disparity estimation and assignment to layers, with the option of pixel assignment to no layer, so that half-occlusions are dealt with as assignment outliers [27]. Interleaved calculation of correspondences and occlusions also has been cast within an expectation maximization framework [10], with high cost matches rejected as arising from occlusion [37]. Finally, previous Bayesian approaches to binocular stereo have been documented (e.g., [39, 2, 14, 33, 37]), which to varying degrees

have dealt explicitly with half-occlusions.

In the light of previous efforts, the major contributions of the present research are as follows. (i) A novel, complementary combination of constraints for binocular half-occlusion detection is presented. The constraints arise from consideration of disparity change and match score in the vicinity of half-occlusions. While previous research has considered instances of both types of constraint for half-occlusion detection, the current work appears to be the first to recognize explicitly their complementarity to yield a combined approach. (ii) The isolated constraints are embodied in a Bayesian estimation framework, which has been implemented in software with qualitative and quantitative evaluation on a standard set of seven binocular test sets. The results show that, where direct comparison is possible, the developed method outperforms all others presented in a recent empirical comparison of half-occlusion detection methods [12].

1.3 Chapter outline

Section 1 of this chapter has served to establish motivation for the study of binocular half-occlusion and place the work in the context of previous research. Section 2 will detail the technical approach that has been developed for half-occlusion analysis and recovery. Section 3 will document empirical evaluation of the developed approach. Finally, Section 4 will provide a brief summary and conclusions.

2 Technical approach

2.1 Overview

In this section of the chapter the developed approach to binocular half-occlusion detection is delineated. The presentation begins by isolating two constraints that, when taken in tandem, tend to be highly indicative of half-occlusions. The first constraint comes in terms of an analytic relationship between the difference in disparities on either side of a half-occluded region and the region width. The second constraint stems from the observation that a region that is visible to one view but occluded to the other view has no physically defined match; hence, any attempted match is likely to have a poor match score. (e.g., low correlation value). To deal in a principled fashion with uncertainties and the combination of the isolated constraints, half-occlusion inference is cast in a Bayesian formulation to yield a probabilistic approach to detection. The section culminates by presenting the details of the resulting estimation procedure that, given disparity and match score maps that result from binocular image matching, produces a map of probability of half-occlusion.

2.2 Image formation

The operative geometric model of image formation is expressed in terms of Fig. 1, which shows a top down view of parallel axis (or otherwise rectified) binocular images formed under perspective projection with, e.g., left and right Euclidean coordinate systems defined at the centers of projection, O_l and O_r (resp.), and separated by baseline, b. The Z-axes

are taken parallel to the optical axes and increasing toward the orthogonal image planes, located at distance $f = 1$ along these axes. X-axes are parallel to the stereo baseline, increasing to the right and Y-axes are mutually orthogonal to the X and Z axes to yield right handed systems. Let world points be given as $\mathbf{A} = (X, Y, Z)$ and subscripts l and r used to reference points to the left and right coordinates systems, respectively, e.g., $\mathbf{A}_l$ references $\mathbf{A}$ to the left system. Image coordinates are similarly denoted using lower case letters; further, since ensuing developments concentrate on relationships along horizontal scan lines, image coordinates will be restricted correspondingly, so that perspective yields, e.g., $\mathbf{a}_l = \frac{X_l}{Z_l}$ as the left image coordinate of $\mathbf{A}$. Given the binocular imaging model, the right image coordinate for $\mathbf{A}$ is given as $\mathbf{a}_r = \frac{X_r}{Z_r} = \frac{X_l - b}{Z_l}$. Correspondingly, disparity (right-based) is given as

$$d_r(\mathbf{A}) = \mathbf{a}_l - \mathbf{a}_r = \frac{b}{Z}.$$

Notice that for surfaces of constant Z (fronto-parallel surfaces), disparity is constant.

2.3 Constraints

Two sources of constraint for half-occlusion detection are considered, which will be shown to be complementary in nature. The first constraint relates disparities that arise from the binocularly visible parts of the surfaces that border the half-occluded region. The second constraint is concerned with a property of the half-occluded region itself.

The first constraint comes in terms of an analytic relationship between the difference in disparity on either side of a half-occluded region and the region width. This relationship can be derived with reference to the geometry illustrated in Figure 1. With respect to the right image, consider the shaded region on the right side of the figure and let world point $\mathbf{A}$ be the left-most point that is binocularly visible, while the world point $\mathbf{B}$ is the right-most half-occluded point (visible only to the right image). The width of the half-occluded region projected to the right image is given as

$$\Omega_r^w(\mathbf{B}, \mathbf{A}) = \mathbf{b}_r - \mathbf{a}_r. \tag{1}$$

The disparity values for points $\mathbf{A}$ and $\mathbf{B}$ are

$$\begin{aligned} d_r(\mathbf{A}) &= \mathbf{a}_l - \mathbf{a}_r \\ d_r(\mathbf{B}) &= \mathbf{b}_l - \mathbf{b}_r \\ &= \mathbf{a}_l - \mathbf{b}_r, \end{aligned}$$

with $\mathbf{b}_l = \mathbf{a}_l$ because $\mathbf{A}$ and $\mathbf{B}$ lie along the same line through O_l, the left optical center, by construction. Correspondingly, the change in disparity across the half-occluded region is given as

$$\begin{aligned} \Delta d_r(\mathbf{B}, \mathbf{A}) &= d_r(\mathbf{B}) - d_r(\mathbf{A}) \\ &= \mathbf{a}_l - \mathbf{b}_r - (\mathbf{a}_l - \mathbf{a}_r) \\ &= \mathbf{a}_r - \mathbf{b}_r \end{aligned} \tag{2}$$

Now, taking the ratio of disparity change (2) to occlusion width (1) it is found that

$$\frac{\Delta d_r(\mathbf{B}, \mathbf{A})}{\Omega_r^w(\mathbf{B}, \mathbf{A})} = \frac{\mathbf{a}_r - \mathbf{b}_r}{\mathbf{b}_r - \mathbf{a}_r} = -1. \tag{3}$$

Interestingly, it is seen that this ratio is equal to the disparity gradient limit [6]. Further consideration of the geometry illustrated in Figure 1 shows that relationship (3) between disparity change and occlusion width also holds for regions visible only to the left view of a binocular pair. In the following, (3) will be referred to as the *disparity-change/width constraint*.[1] (c.f., [15] where this constraint is used, albeit differently, to define legal state transitions for dynamic programming disparity estimation).

The second constraint utilized for half-occlusion detection is based on the observation that, by definition, a region of half-occlusion, e.g., as delimited by $\mathbf{A}_+$ and $\mathbf{B}$ in Figure 1, projects to only one of the two images in a pair. Therefore, in attempting to match points, x, in a region visible to one image only, (e.g. the region spanned by $\mathbf{a}_{r+}$ and $\mathbf{b}_r$), the match goodness resulting from match score, $\rho(x)$, is expected to be poor, since there is no physically defined match, i.e.,

$$\rho(x \in \mathcal{R} = occl) \in \{poor\ match\ scores\}, \tag{4}$$

with $\mathcal{R} = occl$ symbolizing that a region, $\mathcal{R}$, is half-occluded and $\{poor\ match\ scores\}$ match scores indicative of poor performance for a given matcher. In the following, this constraint will be referred to as the *poor match constraint* and will be used in tandem with the disparity-change/width constraint to detect half-occlusions.

Prior to embedding the isolated constraints in a method for half-occlusion detection, it is worth briefly noting their relationships to other approaches (see, e.g., [12, 4] for extended discussion of alternatives and note that for space reasons explicit citations in the remainder of this section focus on more recent contributions). The disparity-change/width constraint is derived by an analysis that explicitly considers both the occluding and occluded surfaces. Therefore, it is most closely related to other approaches that consider the disparities of both occluded and occluder. The "occlusion constraint" states that a discontinuity in disparity in a right-based disparity map corresponds to a half-occluded region in a left-based disparity map and vice versa, e.g., [19, 16, 38]. In comparison, the disparity-change/width constraint is more precise than the occlusion constraint in that it gives a particular relationship between disparity change and occlusion region width; furthermore, it is defined with respect to one view only, which may make it more natural to exploit without performing two-way matching.

The "ordering constraint" also considers disparity of occluder and occluded, as it imposes strict ordering of matched points in left and right images (essential to dynamic programming-based matchers, e.g., [9]) and as a result can disallow matching in half-occlusion regions. However, ordering can be violated in physically realizable view conditions (e.g., involving thin objects [23]), in particular, in situations that do not involve half-occlusion. With respect to any given point in a binocularly viewed scene, the "forbidden zone" is that which contains points that will appear as violations of the ordering constraint [24], e.g. the zone bounded by lines through $\mathbf{A}, \mathbf{a}_l$ and $\mathbf{A}, \mathbf{a}_r$ in Fig. 1. The loci of points that yield the value of -1 for the disparity gradient lie along a boundary of the forbidden zone [43], e.g., the line through $\mathbf{A}$, $\mathbf{a}_l$ (and hence $\mathbf{B}$). The disparity-change/width constraint captures a subset of a foreground

[1] While definition of the disparity-change/width constraint appeals to the disparity of a half-occluded point, e.g., $\mathbf{B}$, this should not pose a problem in practice: Let subscript + applied to a point refer to a point immediately to the right, e.g., $\mathbf{B}_+$ refers to the point immediately to the right of $\mathbf{B}$. If the surface about $\mathbf{B}$ is taken as locally fronto-parallel, then its disparity is constant in that local region and can be estimated from, e.g., $\mathbf{B}_+$, which by definition is binocularly visible.

point's (e.g. $\mathbf{A}$'s) forbidden zone as delimited by a background point (e.g. $\mathbf{B}$) that lies along the foreground point's forbidden zone boundary, e.g. that portion of the forbidden zone that intersects the region bounded by lines through $\mathbf{A}$, $\mathbf{a}_r$ and $\mathbf{B}$, $\mathbf{b}_r$, i.e., that portion relevant to labeling the interval $[\mathbf{a_{r+}}, \mathbf{b}_r]$ as half-occluded. While the disparity-change/width constraint and disparity gradient limit share the same critical value, -1, for developments in this paper the disparity-change/width formulation and nomenclature are better suited as they provide an explicit link to occlusion width, which is exploited in detection.

Other approaches that explicitly consider both surfaces involved in half-occlusion are "bimodality tests", which rely on the observation that histogrammed disparity in the vicinity of half-occlusions can show bimodal distributions as both foreground and background surfaces are captured. Again, the disparity-change/width constraint is tighter, explicitly stating the relationship between disparity values of the surfaces which are covered by the aggregation window.

The poor match constraint arises from consideration of expected match score within a half-occluded region. Previously, such consideration has been used in half-occlusion detection in two ways: (i) Unidirectionally defined (e.g., right-to-left) match scores are examined for patterns indicative of match failure; in some cases patterns of interest involve rapid change in match score. Some global match methods (e.g. dynamic programming, graph cuts [4, 36]), make use of match scores to set occlusion cost. A recent cooperative matcher [44] (an extension of [28]) also uses poor matches to filter out half-occlusions (as well as other matching errors). Poor matches defined by colour differences at aligned image locations also have been used to diagnose half-occlusion [37]. The poor match constraint is an instance of this type of approach as it simply looks for locally bad matches. (ii) Inconsistencies between bidirectional matches are detected, i.e., "left-right checking", a method that requires two matching processes and therefore more expensive that unidirectional approaches. While both approaches can detect half-occlusions, they are not specific to this situation; rather, they more generally diagnose problems in matching (e.g., arising from various sources of noise).

Significantly, the two broad classes of approach to half-occlusion detection discussed in the previous paragraphs are complementary: Methods based on analysis of half-occlusion geometry predict the relationship between disparities that arise on either side of a half-occluded region; whereas, methods based on considerations of match quality predict what is expected within a region of half-occlusion. From this perspective, the present work encompasses a wide range of approaches (including all five compared in [12]), even as it yields a method that is more specific to half-occlusion than other approaches, which often are more generally aimed at diagnosing errors in matching. In the following these observations will be exploited to yield a method for half-occlusion detection that relies on a particular pattern of disparity across a region (as given by the disparity-change/width constraint, since it is particularly tight) as well as analysis of match scores in the intervening region (observation of the poor match constraint, owing to its simple, yet proven to be effective nature).

2.4 Bayesian formulation

To deal rigorously with uncertainties in inference, the disparity-change/width (3) and poor match (4) constraints are cast in a Bayesian framework. Let $\mathcal{R}$ be a region of interest,

moving left-to-right along a scan-line, defined within the interval $[x_1, x_2]$. Let $\mathbf{D} = (D_\Delta, D_\rho)$ be a data vector, comprised of two terms, where D_Δ provides data relevant to disparity-change/width cue (3) and D_ρ provides data relevant to the poor match cue (4).

Let x_{1-} and x_{2+} stand for points immediately to the left and right (resp.) of $\mathcal{R}$ and d_{1-}, d_{2+} be the associated disparities, then D_Δ is given as

$$D_\Delta = (d_{2+} - d_{1-})/(x_{2+} - x_{1-}), \tag{5}$$

c.f., the disparity-change/width constraint (3). D_ρ is given as the set of match scores for each point in the region of concern

$$D_\rho = \{\rho(x) : x_1 \leq x \leq x_2\}. \tag{6}$$

The posterior probability, P, of some region $\mathcal{R}$ being half-occluded, $\mathcal{R} = occl$, given data, $\mathbf{D}$, is expressed via Bayes rule [26] as

$$P(\mathcal{R} = occl \mid \mathbf{D}) = \frac{P(\mathbf{D} \mid \mathcal{R} = occl)\, P(\mathcal{R} = occl)}{P(\mathbf{D})} \tag{7}$$

with evidence calculated according to

$$P(\mathbf{D}) = P(\mathbf{D} \mid \mathcal{R} = occl)\, P(\mathcal{R} = occl) + P(\mathbf{D} \mid \mathcal{R} = vis)\, P(\mathcal{R} = vis), \tag{8}$$

where $\mathcal{R} = vis$ stands for the region $\mathcal{R}$ being visible to both views (as opposed to half-occluded).

The likelihood $P(\mathbf{D} \mid \mathcal{R} = occl)$ depends on both terms of the data vector $\mathbf{D}$. For present purposes, these two terms are taken as independent. This assumption is justified by the observation that spatial change in disparity and match score can vary independently of one another. Empirically, the assumption is supported by the test data used in Section 3: Correlation coefficients between disparity gradient and match score goodness data have values of -0.09 and -0.04 for occluded and visible regions, respectively. Under the independence assumption, the likelihood of a region being occluded is

$$P(\mathbf{D} \mid \mathcal{R} = occl) = P(D_\Delta \mid \mathcal{R}) = occl)\, P(D_\rho \mid \mathcal{R} = occl) \tag{9}$$

Similarly, the likelihood of a region being visible is

$$P(\mathbf{D} \mid \mathcal{R} = vis) = P(D_\Delta \mid \mathcal{R} = vis)\, P(D_\rho \mid \mathcal{R} = vis) \tag{10}$$

The Bayesian formulation has led to the introduction of various distributions that must be instantiated for inference to ensue in practice. Based on the disparity-change/width constraint, (3), the expected value of the ratio of disparity change to width across which change is calculated is -1 for regions of half-occlusion. In contrast, if surfaces are assumed to be locally fronto-parallel (an implicit assumption of standard area-based stereo matchers), then the locally expected value for a binocularly visible region is 0. In either case, noise in disparity estimation will lead to values both larger and smaller than the expected values, with the possibility that values further away from those expected are less likely to be encountered. Based on these observations, the likelihood of a region being half-occluded based on the disparity-change/width constraint, $P(D_\Delta|\mathcal{R} = occl)$, has been modeled as a normal distribution [17], with a mean of -1 and variance estimated from a subset of data used in the empirical validation of the overall approach (see Section 3). Similarly, $P(D_\Delta|\mathcal{R} = vis)$, also is modeled as a normal distribution, but with a mean of 0 and variance estimated from a subset of data used in the empirical validation of the overall approach.

The validity of the normal distribution model fits to $P(D_\Delta|\mathcal{R} = occl)$ and $P(D_\Delta|\mathcal{R} = vis)$ was evaluated via the χ^2 goodness-of-fit test [17], which suggests that the observed data sets

do not differ significantly from the normal distribution fits (probability > 0.95 that rejection of the null hypothesis would be in error, with the null hypothesis being that the data are normally distributed).

Based on the poor match constraint (4), two additional distributions must be instantiated to realize the Bayes formulation, the probability of match score given that a region is (half) occluded, $P\left(D_\rho|\mathcal{R}=occl\right)$, and the probability of match score given that a region is (binocularly) visible, $P\left(D_\rho|\mathcal{R}=vis\right)$. To evaluate these distributions, the individual match scores, $\rho(x)$, comprising D_ρ, (6), are taken as independent, based on the local winner-take-all nature of the matching technique used in empirical evaluation (Section 3), which does not explicitly enforce relationships (e.g., smoothness) between nearby matches. Under independence alone, the likelihood of a region being half-occluded based on match scores would be the product of likelihoods of each pixel in the region being half-occluded. However, some weighting of the product is appropriate in the present context as different sized regions will contribute sets of match scores, D_ρ, of different cardinality. Without weighting, larger regions would have effectively more match score likelihood terms, which would lower the relative weight of the disparity-change/width cue. In this respect, the product of w likelihood terms can be weighted by taking its w-root to yield the geometric mean of the D_ρ components, leading to

$$P\left(D_\rho|\mathcal{R}=occl\right) = \sqrt[w]{\prod_{x\in\mathcal{R}} P\left(\rho(x)|x=occl\right)}, \tag{11}$$

with w the width of region $\mathcal{R}$. Similar reasoning yields an analogous definition for $P\left(D_\rho|\mathcal{R}=vis\right)$.

To instantiate the distributions of $P\left(\rho(x)|x=occl\right)$ and $P\left(\rho(x)|x=vis\right)$ the particular metric used to produce the match scores, $\rho(x)$, must be considered. For empirical evaluation (Section 3), an SAD match metric [40, 4] is used,

$$match_{SAD}(d;x) = \Sigma_{x\in W}|I_r(x) - I_l(x+d)|, \tag{12}$$

which evaluates th cost of associating disparity, d, with location, x, in the right image, I_r, with location $x+d$ in the left image, I_l, and W the match window support. For simple cases where surface patches in the world yield the same image intensity patterns to two views (excepting shift by disparity, d), the expected SAD match score is 0, with a systematic bias (e.g., photometric difference between the images) yielding some off-set. In either case, noise in image formation and/or the matching process will perturb match scores from the expected value, with larger perturbations less likely than smaller. So, ignoring the effects of the absolute value, one might consider a normal distribution to model match scores. Applying the absolute value remaps differences yielding negative values by reflecting them about zero, back to positive values, i.e., letting I denote images, $\forall m = I_r(x) - I_l(x+d) < 0, m \mapsto -m$. Correspondingly, the distribution for probability of match score (SAD), given that a pixel is (binocularly) visible, $P\left(D_\rho|\mathcal{R}=vis\right)$, or half-occluded, $P\left(D_\rho|\mathcal{R}=occl\right)$, is modeled as a "mirrored-about-zero-normal" distribution, i.e., as normal, but with negative valued observations reflected back along the positive axis

$$P(\rho) = \begin{cases} \left(\sqrt{2\pi}\sigma\right)^{-1}\left(e^{-\frac{1}{2}\left(\frac{\rho-\mu}{\sigma}\right)^2} + e^{-\frac{1}{2}\left(\frac{-\rho-\mu}{\sigma}\right)^2}\right); \rho > 0 \\ \left(\sqrt{2\pi}\sigma\right)^{-1} e^{-\frac{1}{2}\left(\frac{-\mu}{\sigma}\right)^2}; \rho = 0 \\ 0; \rho < 0 \end{cases} \tag{13}$$

with μ and σ the mean and standard deviation of the underlying normal distribution.

Separate parameters sets (μ, σ) for fits to (13) for $P\left(\rho(x)|x=occl\right)$ and $P\left(\rho(x)|x=vis\right)$

were estimated from a subset of test data of Section 3. As before, the validity of these model fits was evaluated through the χ^2 goodness-of-fit test, which suggests that the observed data sets do not differ significantly from the mirrored-normal distribution fits (probability > 0.95 that rejection of the null hypothesis would be in error, with the null hypothesis now being that the data are mirror-normally distributed). [2]

Finally, it is difficult to analytically define priors for a region being half-occluded, $P(\mathcal{R} = occl)$, and a region being binocularly visible, $P(\mathcal{R} = vis)$. For present purposes, these have been defined empirically, based on the ground-truth associated with a subset of the test data used in Section 3, to yield $P(\mathcal{R} = occl) = 0.08$ and $P(\mathcal{R} = vis) = 0.92$ with standard error of 0.03 between different ground-truth images.

2.5 Estimation

Having observed data $\mathbf{D} = (D_\Delta, D_\rho)$ for a region $\mathcal{R}$ it is straightforward to evaluate the probability that the region is half-occluded, $P(\mathcal{R} = occl \mid \mathbf{D})$, by solving the Bayes relation (7) with appropriate substitutions for evidence (8) and combined likelihoods (9), (10) and making use of the instantiated model distributions and priors given in Sec. 2.4.

Input to the estimation process is taken to be a disparity map and an associated map of match scores, which for convenience are assumed to be rectified. Along each horizontal scan-line, the probability of half-occlusion, $P(\mathcal{R} = occl \mid \mathbf{D})$, given by (7) is calculated for each region, $\mathcal{R}$, with widths ranging from 1 pixel to the largest magnitude disparity observed along the line. For each width of interest, the data vector, $\mathbf{D}$ is calculated by applying the formulae for its components, (5) and (6), to the disparity and match score maps, which provides the necessary fodder for calculation of $P(\mathcal{R} = occl \mid \mathbf{D})$. Any given pixel along a scan-line can be associated with a number of regions of varying widths. To make a final probability of half-occlusion assignment to any given pixel, it is associated with the largest posterior probability of being half-occluded out of all regions that could possibly cover it. Completion of this calculation for all pixels along a scan-line assigns a probability of occlusion to each pixel. Iteration across scan-lines yields a map of half-occlusion probability in spatial registration with the input disparity and match score maps.

3 Empirical evaluation

3.1 Methodology

To evaluate the efficacy of the combined constraints for half-occlusion detection, two considerations are critical. (i) It is desirable to document the utility of the constraints independent of other aspects of stereo processing. (ii) Comparison to previous empirical evaluation of half-occlusion detection methods must be supported, e.g., [12].[3] For both reasons the current

[2] In formulating the match score distributions it has been necessary to make a commitment to a particular matching metric, SAD. Other match metrics that avoid negative values, e.g., Sum of Squared Differences, might yield to a similar analysis; although, formulations that deal with symmetric distributions would be more appropriate in other situations, e.g., correlation-type metrics.

[3] Results reported in conjunction with the Middlebury test suite [29, 36] are not directly comparable as they do not show results for half-occlusion detection per se; they focus on disparity.

approach is evaluated in terms of post-processing on disparity and match-score maps recovered via a simple area-based stereo matcher. Use of a simple matcher forces half-occlusion detection to deal with a range of difficult situations (e.g., foreground fattening, poorly resolved thin structures, mismatches in low texture, etc). Post-processing allows for clear separation of the effects of the half-occlusion constraints from other aspects of stereo processing. Moreover, this methodology matches that of [12], which makes direct comparison possible. While operation in conjunction with a more sophisticated matcher or as a more integral component of a matcher has potential to yield better results (suggesting interesting future directions), such options are less well suited to the goal of understanding the current approach in comparison to others and with independence from other aspects of stereo processing.

Two classes of test data have been used to evaluate an algorithmic instantiation of the half-occlusion probability estimation procedure described in Section 2.5. First, a set of laboratory images with associated ground truth is used to allow for quantitative evaluation. Second, a set of natural, outdoor images is used to allow for evaluation of performance in more realistic scenes, albeit without quantitative analysis as ground truth is not available. The laboratory images are selected from the Middlebury test suite [29]. For these cases, ground truth half-occlusion was constructed by labeling any unmatched pixel according to the ground truth disparity as a half-occlusion. The particular data sets used were *Tsukuba*, *Map*, *Venus* and *Sawtooth* (as they were used in previous empirical comparisons, e.g., [36, 12]) and *Cones* (as it provides greater scene complexity). Natural outdoor images selected were *Pentagon* [8] and *Flower Garden* [5] (as they are standard and provide very different viewing scenarios, aerial and terrestrial, respectively.)

For the majority of test cases, tests are performed with input disparity and match-score maps recovered by an area-based matching algorithm [40] using the SAD match metric (12), fixed squared aggregation windows (7×7), winner-take-all and operating in a hierarchical, coarse-to-fine framework [32] using Gaussian pyramids [20]. During half-occlusion evaluation, one set of matching parameters was used across the entire test suite (hand selected to yield good overall performance), except for search range, which was set to the maximum disparity present for each test case. The exception to these specifications is *Tsukuba*: To facilitate direct comparison with previous results [12], the correlation and plane-plus-parallax-based disparity and match score maps (and ground truth half-occlusion) used in the previous study were employed (courtesy of G. Egnal). In all cases, the recovered disparity and match score maps are input to the half-occlusion detection procedure, as described in Section 2.5. Currently, algorithmic instantiation in software produces a half-occlusion map for 384×288 images in approximately 1 second, running in unoptimized C under Linux on a Xeon 3.06GHz processor with 1MB cache. Only right-based results are reported for reasons of space; left-based results are similar.

3.2 Results

For lab scenes (Figures 2 and 4), the algorithm generally gives consistently good qualitative results. Both wide and narrow half-occlusion regions are detected; although, localization can be imprecise due to disparity errors near object boundaries – a known weakness of area-based matchers [4, 36].

Quantitative analysis derives from comparison of the obtained occlusion maps to ground truth. Results are presented as ROC plots, percentage of hits and false positives (relative to the entire image) along the ordinate and abscissa, resp. It is found that the vast majority of true half-occlusions can be detected, while keeping the false positive rate in the $10-20\%$ range. While current state-of-the-art stereo matchers can yield impressive performance in the vicinity of 3D boundaries [36], these methods make use of constraints and heuristics beyond those embodied in the matcher used in the current study. With respect to a comparable matcher [12], all half-occlusion detection methods considered yielded a higher rate of false positives for similar hit rates. Also of note, performance is reasonably stable to variation in algorithm parameters, e.g., even extremes of changing half-occlusion and visible priors to equiprobable and changing match window size by a factor of 2 yield less than 2% change in the areas under the ROCs.

The relative contributions of the disparity-change/width and poor match score constraints are shown in Figure 3 using *Tsukuba*. While both constraints are seen to be preferentially indicative of half-occlusion, individually they yield lower hit-to-false-positive ratios than their combination, with poor match score alone being the worst in this regard. Inspection of the probability maps shows that the false signals from one constraint tend not to correspond to those in the other; whereas, hits do tend to correspond. Hence, the constraints are complementary; their combination yields the best result.

For *Tsukuba*, direct comparison can be made to five standard approaches [12]. Here, it is found that for both correlation-based (Figure 2) and plane-plus-parallax-based matching (Figure 4) the current approach outperforms all of the five standard approaches, as its ROC tends to the upper left of the ROCs for the alternative ones. In certain cases, the difference is dramatic, e.g., in the low false positive regions under correlation, the current method can improve on hits by as much as approximately 75% over the best previous results. The closest rival appears to be the occlusion constraint under plane-plus-parallax matching; although even here the current methods wins out beyond approximately 10% false positives, ultimately improving on hits by approximately 15% for a given false positive rate.

Outdoor scene testing also shows strong performance in the test cases considered (Figure 5). For *Pentagon*, the major half-occlusion regions on the right side of the building's inner and outer sides are correctly located, as well as many of the interior corridors; whereas, the number of false positives is small. For *Flower Garden*, the right side of the tree and the narrow inclined bar in the background are correctly detected as half-occluded. The algorithm also gives strong response, albeit poorly localized, in the vicinity of clustered tree branches and associated thin half-occlusions. There are few false positives in the flower field itself, with greatest spurious signal tending to coincide with sharp disparity gradient associated with poor match score.

Overall, the developed algorithm performs well in detecting half-occlusions across a wide variety of scenes while keeping false positives relatively low. The apparent weakness arises in fine localization of half-occluded regions: The offset from ground truth is noticeable and the width of detection regions is inexact. Also, false positives are often detected near non-half-occlusion 3D boundaries. Both of these phenomena can be attributed to the fact that the area-based matcher performs poorly near 3D boundaries, resulting in poor match scores and spurious disparity gradient patterns, which mimic the utilized constraints.

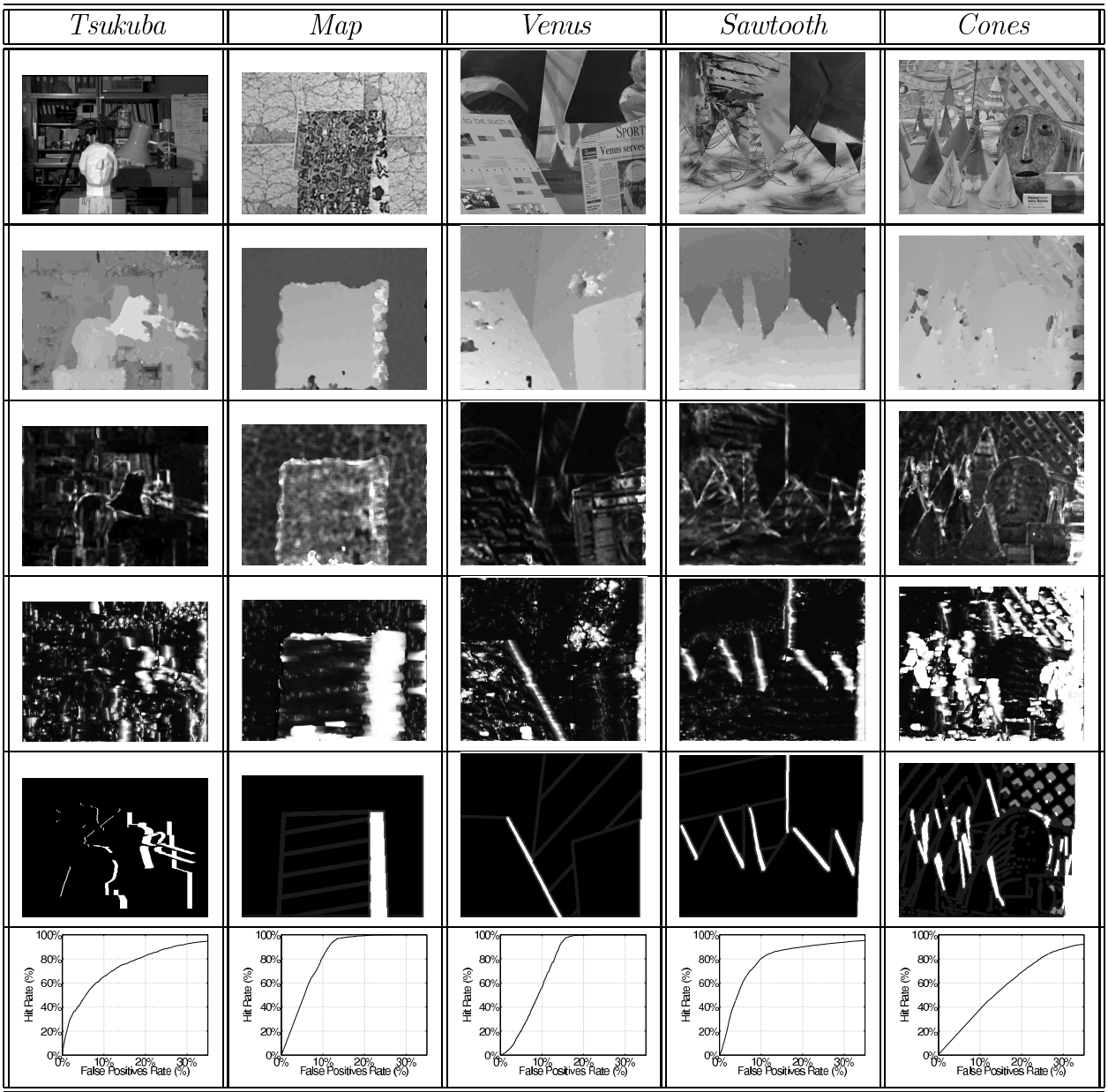

Figure 2: Half-Occlusion Results for Laboratory Imagery with Ground Truth. For five stereo pairs (left-to-right) shown are left image, estimated disparity (brighter is closer), match scores (brighter is poorer), probability of half-occlusion (brighter is higher), ground-truth half-occluded regions (white), and detection ROC (top-to-bottom).

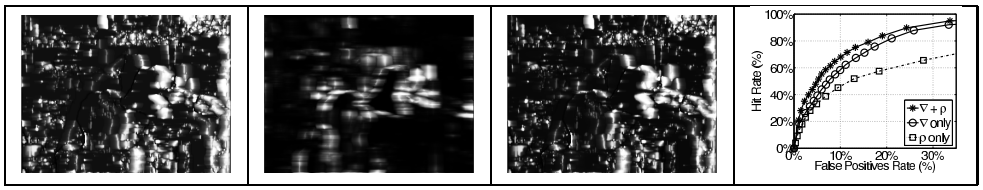

Figure 3: Half-Occlusion Constraint Comparison for *Tsukuba*. Shown are disparity-change/width ($\bigtriangledown$), poor match score (ρ), combination ($\bigtriangledown + \rho$, repeated from Figure 2) probability of half-occlusion and detection ROCs (left-to-right). Input disparity and match score maps are as in Figure 2.

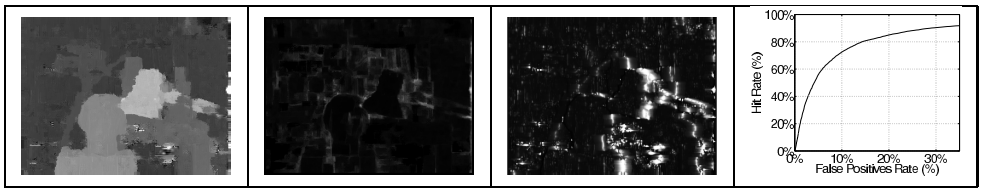

Figure 4: Half-Occlusion Results for *Tsukuba* under Plane-Plus-Parallax. Shown are disparity, match scores, probability of half-occlusion and detection ROC (left-to-right). ROCs for five previous methods under Plane-Plus-Parallax can be found in [12].

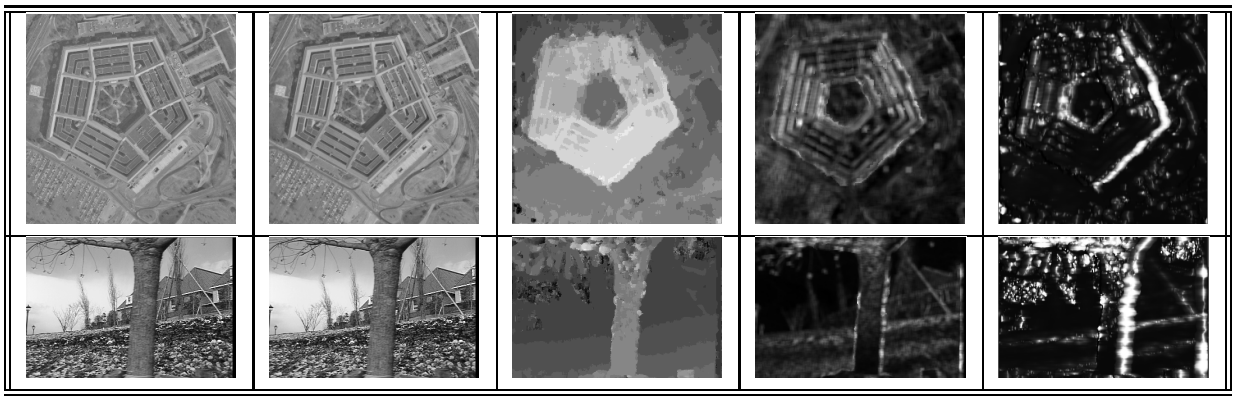

Figure 5: Half-Occlusion Results for Outdoor Imagery. For two standard stereo pairs, *Pentagon* and *Flower Garden* (top-to-bottom), shown are left and right images, estimated disparity, associated match scores, and probability of half-occlusion (left-to-right).

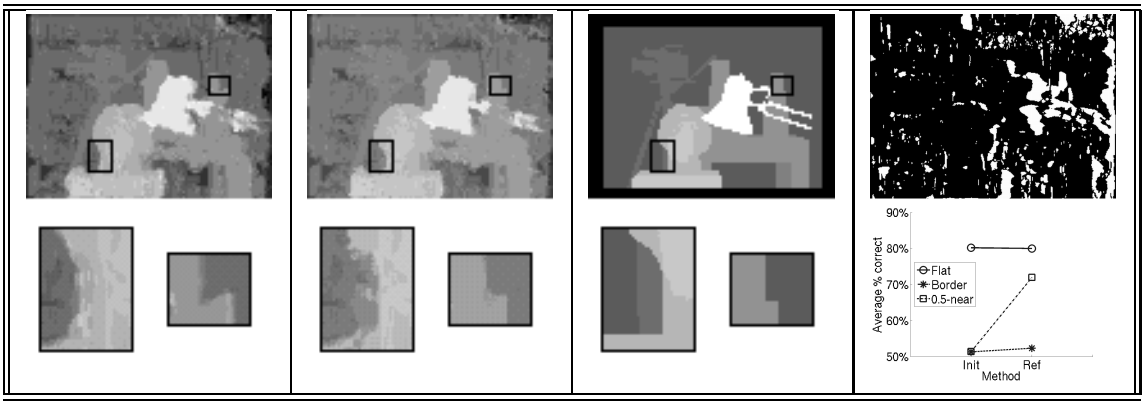

Figure 6: Application of Detected Half-Occlusions. Left to right: *Tsukuba* initial, refined and ground truth disparity maps; thresholded half-occlusion probability, $P \geq 0.5$; comparative plot of good-pixels in textureless regions (Flat), near general 3D boundaries, i.e., discontinuities not including half-occlusions (Border), and near half-occlusions (0.5-near) for initial (Init) and refined (Ref) disparity maps. Percentage correct given according to standard criteria of estimated disparity being within 1 of ground truth disparity taken as a good pixel [36] and percentage relative to an entire image. Averages taken across all test images used in Figure 2. Insets highlight a few selected improvements.

3.3 Application

In practice, the proposed method for half-occlusion detection could be used as post-processing for a fast area-based method of disparity estimation, or incorporated directly into either local or global methods to find disparity and occlusion simultaneously. Here, option one is demonstrated: Improvement of the initial disparity produced by the SAD-based matcher in the vicinity of 3D boundaries, especially half-occlusions. The originally recovered disparity map has been refined in regions that are marked as half-occluded by repeating the matching process while making use of adaptively defined windows that conform in spatial support to avoid half-occlusions; non-half-occluded regions are not altered; results are shown in Figure 6. Statistics are given across all test pairs shown in Figure 2, with *Tsukuba* disparity maps given as a visual example. It is seen that disparity estimates in the vicinity of 3D boundaries, especially half-occlusion, are improved without damaging results in other difficult areas (e.g., textureless), as adaptive matching also operates at false positives. Improvements are most noticeable in reducing fattening of foreground objects and increased precision of boundary details. Significantly, while the utility of adaptive windows has been documented previously [4, 36], here improvement is had without the need to apply adaptive processing everywhere as effort is concentrated in regions of half-occlusion.

4 Summary and conclusions

A novel approach to binocular half-occlusion detection has been presented. The approach is based on two complementary constraints derived from binocular matching data (disparities and match scores) that arise in the vicinity of half-occluded regions. The first constraint con-

cerns disparity change across a region of half-occlusion. In particular, the ratio of disparity change to width is equal to -1 in the ideal case. The second constraint concerns the expected behavior of a binocular matcher operating within a region of half-occlusion. In particular, since there is no physically defined match in such regions, any recovered match is expected to have a poor match score (e.g., large SAD). To allow for uncertainty in inference, the constraints have been combined in a Bayesian formulation. To understand applicability of the approach, a corresponding algorithm for detecting half-occlusions provided input maps of disparity and match score has been developed. To isolate the effects of half-occlusion constraints from other aspects of stereo proceessing and to facilitate comparison to previous empirical studies of half-occlusion analysis, the presented instantiation of the approach works as post-processing on precomputed disparities and match scores. Empirically, the algorithm yields strong performance on standard test data; indeed, performance is superior to all approaches evaluated in the most comparable study to date [12].

The computational analysis presented in this chapter can be applied in a number of ways. Application to correcting inappropriate matches near three-dimensional boundaries via disparity post-processing has been illustrated in this chapter. Alternatively, the approach could be embedded within a matching procedure, be it local or global [4], to inhibit bad correspondences as they emerge. Finally, the developed computational framework can be used to guide efforts aimed at modeling the recovery of half-occlusions and otherwise processing disparity information in biological systems [42].

Acknowledgments

This reseach was supported in part by a grant jointly funded by CRESTech, the Ontario Centres for Excellence and Macdonald-Detweiller Associates, Space Missions.

References

[1] Barnard S, Fischler M (1982) Computational stereo. ACM Computer Surveys (CSUR) 14: 553-572

[2] Belhumeur P (1996) A Bayesian approach to binocular stereopsis. International Journal of Computer Vision 19: 237-260

[3] Bolles R, Baker H, Hannah M (1993) The JISCT stereo evaluation. In April DARPA Image Understanding Workshop 263-274

[4] Brown M, Burschka D, Hager G (2003) Advances in computational stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence 25: 993-1008

[5] Brown University, *Image Sequences*, http://www.cs.brown.edu/people/black/images.html, current 2004

[6] Burt P, Julesz B (1980) A disparity gradient limit for binocular fusion. Nature 208: 615-617

[7] Burton H (1945) The optics of Euclid. Journal of Optical Society of America 35:357-372

[8] Carnegie Mellon University, *VASC Image Database*, http://www.vasc.ri.cmu.edu/idb, current 2004

[9] Criminisi A, Shotton J, Blake A, Torr P (2003) Gaze manipulation for one-to-one teleconferencing. Procedings of the Ninth IEEE International Conference on Computer Vision – vol 2:191

[10] Dempster A, Laird N, Rubin D (1977) Maximum likelihood from incomplete data. Journal of the Royal Statistical Society B 39: 1-38

[11] Deng Y, Yang Q, Lin X, Tang X (2005) A symmetric patch-based correspondence model for occlusion handling. Proceedings of the International Conference on Computer Vision

[12] Egnal G, Wildes R (2002) Detecting binocular half-occlusions: empirical comparisons of five approaches. IEEE Transactions on Pattern Analysis and Machine Intelligence 24: 1127-1133

[13] Fusiello A, Roberto V (1997) Efficient stereo with multiple windowing. Proceedings of the 1997 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 97) 885-863

[14] Geiger D, Ladendorf B, Yuille A (1995) Occlusion and binocular stereo. International Journal of Computer Vision 14: 211-226

[15] Grammalidis N, Strintzis M (1998) Disparity and occlusion estimation and their coding for the communication of multiview image sequences. IEEE Transactions on Circuits and Systems for Video Technology 8: 328-344

[16] Hirschmuller H, Innocent P, Garibaldi J (2002) Real-time correlation-based stereo vision with reduced border errors. International Journal of Computer Vision 47: 229-246

[17] Hoel P (1984)Introduction to Mathematical Statistics (Fifth Edition). John Wiley & Sons. Toronto, Ontario

[18] Howard I, Rogers B (2002) Seeing in Depth. I. Porteous. Thornhill, Ontario

[19] Ishikawa H, Geiger D (1998) Occlusions, discontinuities and epipolar lines in stereo. Lecture Notes in Computer Science 1406:232-248. Proceedings of the European Conference on Computer Vision 1-14

[20] Jahne B (1993) Digital Image Processing: Concepts, Algorithms and Scientific Applications. Springer-Verlag. Berlin

[21] Kanade T, Okutomi M (1994) A stereo matching algorithm with an adaptive window: theory and experiment. IEEE Transactions on Pattern Analysis and Machine Intelligence 16: 920-932

[22] Kang S, Szeliski R, Chai J (2001) Handling occlusions in dense multi-view stereo. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 103-110

[23] Krol JD, van der Grind WA (1980) The double nail illusion. Perception 9:651-659.

[24] Krol JD, van der Grind WA (1982) Rehabilitation of a classical notion of Panum's fusional area. Perception 11: 615-619

[25] Lawson R, Mount D (1967) Minimum condition for stereopsis and anomalous contour. Science 158:804-806

[26] Leonard T, Hsu J (1999) Bayesian Methods. Cambridge University Press. Cambridge, UK.

[27] Lin M, Tomasi C (2004) Surfaces with occlusion from layered stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence 26: 1073-1078

[28] Marr D, Poggio T (1976) Cooperative computation of stereo disparity. Science 194:283-287

[29] Middlebury College, *Stereo Vision Research Page*, http://www.middlebury.edu/stereo, current 2005

[30] Okutomi M, Katayama Y, Oka S (2002) A simple stereo algorithm to recover precise object boundaries and smooth surfaces. International Journal of Compuer Vision 47: 261-273

[31] Pollard S, Mayhew J, Frisby J (1985) PMF: a stereo correspondence algorithm using a disparity gradient limit. Perception 14(4): 449-470

[32] Quam L (1984) Hierarchical warp stereo. DARPA Image Understanding Workshop 149-155

[33] Read J (2002) A Bayesian approach to the stereo correspondence problem. Neural Computations 14(6):1371-1392.

[34] Richter J (1977) (Ed) Selections from the Notebooks of Leonardo da Vinci. Oxford, U.K. Oxford University Press

[35] Sara R (2002) Finding the largest unambiguous component of stereo matching. Proceedings of teh 7th European Conference on Computer Vision Part III 900-914

[36] Scharstein D, Szeliski R (2002) A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision 41(1):7-42

[37] Strecha C, Fransens R, van Gool L (2004) Wide-baseline stereo from multiple views: A probabilistic account. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04) v1:718-725

[38] Sun J, Li Y, Kang S, Shum H (2005) Symmetric stereo matching for occlusion handling. Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) v2:399-406

[39] Szeliski R (1990) Bayesian modeling of uncertainty in low-level vision. International Journal of Computer Vision 5(3):271-301

[40] Trucco E, Verri A (1998) Introductory Techniques for 3-D Computer Vision. Prentice-Hall. Upper Saddle River, NJ

[41] Veksler O (2003) Fast variable window for stereo correspondence using integral images. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04) v1:556-561

[42] Wilcox L, Wildes R, Lakra D, Spengler R (2005) The contribution of binocular and monocular texture elements to depth ordering. Vision Sciences Society.

[43] Yuille A, Poggio T (1984) A Generalized Ordering Constraint for Stereo. AI Lab Memo 777, MIT. Cambridge, MA

[44] Zitnick L, Kanade T (2000) A cooperative algorithm for stereo matching with occlusion detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 22:675-684