

# THE STEREO MATCHING PROBLEM FROM THE TOPOLOGICAL VIEWPOINT\*

A. Peter Blicher

Artificial Intelligence Laboratory, Computer Science Department, Stanford University  
Stanford, California 94305, USA, and  
Mathematics Department, University of California, Berkeley

## ABSTRACT

A unifying abstract mathematical structure is presented for a number of vision problems, notably stereo, motion stereo, optic flow, and matching. Ideas from modern differential topology are presented and applied to the general matching problem, a common approach to stereo matching, defined as follows. Given 2 picture functions  $F_1, F_2 : M^2 \rightarrow \mathbb{R}^n$ , one finds regions  $K_1, K_2 \subset M^2$  and a 1-1 matching function  $g_\pi : K_1 \rightarrow K_2$  such that  $F_1 = F_2 \circ g_\pi$ . It is shown that *generically* for monochrome pictures ( $n = 1$ ) there is a large infinity of solutions, but for 2 or more colors ( $n \geq 2$ ) the solution is unique.

The paper is offered partially in the hope of introducing vision workers to this type of mathematics and persuading them of its utility.

## 1 INTRODUCTION

Analogously to the Erlanger Programm, the task of computer vision can be viewed as finding invariants of irradiance functions under the rigid motion group of  $\mathbb{R}^3$ . This paper describes this structure in the language of modern abstract mathematics, providing a framework for understanding and the possibility of applying powerful methods to resolve fundamental questions. We prove a theorem which says that occlusion-free stereo matching requires at least 2 color dimensions or some knowledge of the imaging geometry.

For reasons of space, the treatment here is abridged and terse. The interested reader can find a more complete exposition, including mathematical details, definitions, and wider discussion in [Blicher 1983].

## II THE MATHEMATICAL STRUCTURE

The structure is depicted by the commutative diagram Fig. (\*'). The object surface  $\Sigma$  is embedded in  $\mathbb{R}^3$  via  $i$ .  $M^3$  is a fixed 3 dimensional subset of  $\mathbb{R}^3$ , and is the domain of definition for the imaging projection  $\pi$ , which maps it to  $M^2$ , the 2 dimensional image space.  $F_1$  is the observed image intensity on some closed set  $K_1$  of the image plane.  $S_1$  and  $K_1$  are corresponding visible regions of  $\Sigma$  and the image space  $M^2$ , resp.

We assume that the surface  $\Sigma$  admits a function  $F : \Sigma \rightarrow \mathbb{R}^n$  which describes intrinsic surface features. E.g., for  $F : \Sigma \rightarrow \mathbb{R}^1$  (i.e.  $n = 1$ ),  $F$  represents an *intrinsic* surface brightness or luminance. Thus we ignore the effect of viewpoint on image irradiance, or equivalently, we take the reflectance function to be constant. To the extent that we deal only with small changes in viewpoint, that will usually be a good approximation.

$F$  can be thought of as the intrinsic surface property *albedo*; then our analysis deals with quantities that depend only on albedo, to good approximation. For the case  $n \geq 2$ , we have in mind color images: normal human cone vision incorporates a function  $F_1 : K_1 \rightarrow \mathbb{R}^3$  ( $n = 3$ ). Note that we also subsume cases for a smaller (i.e.  $n = 2$ ) or larger ( $4 \leq n < \infty$ ) number of passbands, or in fact any surface attribute, such as a multi-dimensional texture measure, which can be thought of as taking pointwise values in some finite dimensional space.

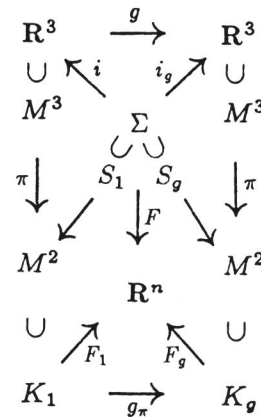


Fig. (\*')

The map  $g : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  is a rigid motion of  $\mathbb{R}^3$ . For many surfaces  $\Sigma$ , different viewpoints  $g$  have different domains of visibility of  $\Sigma$ . So, in general,  $g_\pi$  as pictured may not be well defined, since we cannot be sure that for  $K_1, K_g, S_1, S_g$  as we have defined them, that  $S_1 \subset S_g$ , or equivalently that  $\pi \circ i_g : \Sigma \rightarrow M^2$  is 1-1 on  $S_1$ . I.e., part of what we see in the picture  $F_1$  might be hidden from view when we look after doing  $g$ . Hence the regions  $K_1, K_g$  must be chosen in such a way that  $g_\pi$  is well-defined. With these restrictions,  $g_\pi$  is a diffeomorphism  $K_1 \rightarrow K_g$  with the property that  $F_1(p) = F_g(g_\pi(p))$ , which is the same as saying that  $g_\pi$  is a deformation of the picture  $F_1$  into the picture  $F_g$ . Note that this observation is also equivalent to asserting that the diagram is commutative (for the  $F_1, g_\pi, F_g$  loop).

We have sidestepped the issues of occlusion, shadowing and photometry. Nevertheless, major parts of the following problems are subsumed in the structure we have presented: area correlation stereo, general matching, motion stereo and optical flow, feature based stereo, singularity tracking.

\*This work was supported in part by ARPA contracts MDA903-80-C-0102 and N00039-82-C-0250.

### III STEREO BY GENERAL MATCHING

As a straightforward (but not trivial!) application of the abstract viewpoint we are proposing, we show that for monochrome images, the matching problem is insoluble, and study the conditions which allow unique solution.

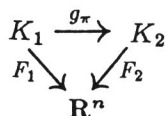


Fig. (GM)

A common approach to stereo matching is via general matching. I.e., given 2 picture functions  $F_1, F_2 : M^2 \rightarrow \mathbf{R}^n$ , one finds regions  $K_1, K_2 \subset M^2$  and a 1-1 matching function  $g_\pi : K_1 \rightarrow K_2$  such that the diagram Fig (GM) commutes. Only after the matching function is found is the surface embedding computed by associating relative depth with relative disparity at each point of (say)  $K_1$ . We assume that for the matching phase no information about imaging geometry is used, in particular that there is no assumption of rectification, and no knowledge of epipolar geometry is used, so that arbitrary (but sufficiently differentiable) distortions are possible. Since we are concerned with existence, this is an idealization of what happens in practice, where usually there is at least implicit use of some geometric constraints. We show, in fact, that such use is *necessary*.

The following question then arises:

**Problem** (Uniqueness of General Matching). If we seek an arbitrary (piecewise)  $C^1$  diffeomorphism  $g_\pi$  to make Fig.(GM) commute, when are we guaranteed a unique solution to the matching problem?

E.g., if  $F_1, F_2$  are both constant functions, i.e., we have uniformly gray pictures, the problem is completely degenerate, and any diffeomorphism  $g_\pi$  is a solution.

We do not consider problems related to occlusion, and instead assume that in fact it is possible to find regions  $K_1, K_2$  and a map  $g_\pi$  which fulfill the smoothness criteria and make Fig. (GM) commute. Our concern is whether  $g_\pi$  is then unique.

**Theorem** (2 color theorem). Stereo requires at least 2 color dimensions or 3 space dimensions. I.e., for a monochrome picture, general matching has infinitely many solutions, but for 2 or more color dimensions, it is generally unique. Hence the monochrome case requires knowledge of the imaging situation to constrain the problem. (In particular, this applies to gray-level correlation.)

More precisely, consider the commutative diagram Fig. (GM) where  $g_\pi$  is a  $C^1$  diffeomorphism,  $F_1, F_2$  are  $C^1$ , and  $K_1, K_2$  are differentiable submanifolds of  $\mathbf{R}^2$ .

If  $n = 1$  (i.e. the picture is monochrome), then  $\exists$  an infinite dimensional family of  $C^1$  diffeomorphisms  $\{h_\varphi\}$  such that replacing  $g_\pi$  by  $h_\varphi$  also results in a commutative diagram (i.e. is a solution). The family  $h_\varphi$  is parametrized by (at least) the continuous functions  $K_1 \rightarrow \mathbf{R}$ , and contains an isomorph of a neighborhood of the identity.

If  $n = 2$  (i.e. the picture has 2 color dimensions), then generically there will be a finite number of  $g_\pi$  which make the

diagram commute (note we have assumed that such a  $g_\pi$  exists). If we take  $K_1, K_2$  to be rectangles or discs (as in a usual picture) then generically there is a unique  $g_\pi$ .

If  $n \geq 3$  (i.e. the picture has at least 3 color dimensions), then generically there will be a unique  $g_\pi$  which makes the diagram commute.

**Proof** (case  $n = 1$ : monochrome pictures). The idea of the proof is very simple; the difficulty lies in establishing when it is valid. The idea is this. Observe that if Fig. (GM) is a commutative diagram then  $g_\pi$  takes contour lines to contour lines. Conversely, *any* diffeomorphism  $h : K_1 \rightarrow K_2$  which takes contour lines of  $F_1$  to contour lines of  $F_2$  satisfies the conditions for  $g_\pi$ . Thus any  $g_\pi$  taking contour lines to contour lines will solve our local matching problem. But how many such  $g_\pi$ 's can there be? Assume for the moment that a typical contour map contains a diffeomorphic image of the fragment represented by the solid lines in Fig. (frag).

If  $\psi : K_1 \rightarrow K_1$  is a diffeomorphism leaving contours of  $F_1$  invariant, then if  $g_\pi$  is a matching function so is  $g_\pi \circ \psi$ . Define  $\psi$  as follows. As you go along the dotted line  $t \mapsto \gamma(t)$  in Fig (frag), slide each contour along itself by an angle  $\theta(t)$ . As long as  $\theta : I \rightarrow \mathbf{R}$  is a diffeomorphism onto its image, the map  $\psi$  will be a diffeomorphism in a neighborhood of the dotted line. To the extent that this picture is valid, there will be as many matchings  $g_\pi \circ \psi$  as there are such maps  $\theta$ .

Actually, we are going to use a slightly more general (and technical) method to construct a family of diffeomorphisms  $\psi_\alpha$  roughly in 1-1 correspondence with the set of all  $C^r$  functions  $K_1 \rightarrow \mathbf{R}$ . For this we will use a canonical vector field defined along the contour lines of  $F_1$ , which will tell us how much to slide each contour line. Define a new vector field on  $K_1$  by rotating each of the local vectors of  $\nabla F_1$  by  $+90^\circ$ . The orbits of  $Z$  are exactly the contour lines of  $F_1$ , and the critical points are exactly the critical points of  $\nabla F_1$ . We now consider the flow generated by the vector field  $Z$ .

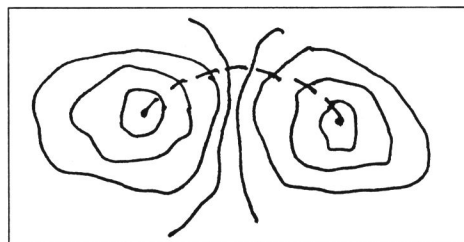


Fig. (frag)

Near the boundary of  $K_1$ , the time-one map of this flow may not be defined if a contour line has a boundary. However, this is easily overcome by using a "bump" function [Abraham 1978]  $\beta : K_1 \rightarrow \mathbf{R}$  to get a vector field  $\beta \cdot Z$  on  $K_1$  which smoothly goes to zero very close to the boundary, and hence has a flow  $\varphi_t$  which never leaves  $K_1$ .

Thus for each  $t$ ,  $\varphi_t : K_1 \rightarrow K_1$  is a diffeomorphism on  $K_1$  leaving contour lines invariant. This family of diffeomorphisms can be enlarged even more. Notice that multiplying the vector field  $Z$  by a scalar  $C^r$  function  $\rho : K_1 \rightarrow \mathbf{R}$  does not alter orbits. Therefore we can enlarge the class of diffeomorphisms  $\varphi_t$  by taking all diffeomorphisms  $\varphi_{t,\rho}$  given by the flows of  $\rho \cdot \beta \cdot Z$  on  $K_1$ . Observe that for any constant  $\alpha$ ,  $\varphi_{\alpha t, \rho} = \varphi_{t, \alpha \rho}$ , so if  $\rho$  is a constant function,  $\varphi_{t, \rho} = \varphi_{\rho t, 1} = \varphi_{1, \rho t}$ . Thus  $\{\varphi_{t, \rho}\} = \{\varphi_{1, \rho}\}$ , so by abuse of notation we will write  $\varphi_\rho$  for  $\varphi_{1, \rho}$ . QED ( $n = 1$ ).

### A. Discussion of what we have shown so far

In the monochrome matching of 2 regions free of occlusions, the match is far from unique. In fact there are essentially as many matches as  $C^r$  functions from such a region to the reals. This stems directly from the fact that the iso-brightness loci constitute connected differentiable 1-dimensional objects. That in turn is a consequence of the fact that the picture is a map from a 2-dimensional object to a 1-dimensional object.

Away from critical points, the matching diffeomorphisms can differ greatly: contour lines can be slid along themselves arbitrarily large amounts. From a practical point of view, given 2 pictures and a matching function, it is a simple matter to choose a  $\rho$  and compute  $\varphi_\rho$ , the time-one map of  $\rho \cdot \beta \cdot Z$ , giving a new match. A matching strategy based on this analysis would first match critical points (generically a discrete combinatorial problem), and then contour lines intersecting the gradients through the critical points. Of course, an actual program would also have to deal with noise, digitization, occlusion, and variation of image irradiance with viewing position; and it would have additional constraints available.

### B. The validity of the intuitive picture

Although the theorem is proved for  $n = 1$ , it's not clear how the intuitive idea of the proof is related to the technical method we actually used. We now turn to the validity of the picture (Fig. (frag)). This is interesting beyond the confines of our present problem; e.g. it casts light on the structure of zero crossings.

First, the proof given above fits into the intuitive scheme presented earlier for using Fig. (frag), since  $\rho\beta|Z|$  is essentially the rotation function  $\theta$  we discussed earlier. But is Fig. (frag) a reasonable picture for the contour lines of a picture function? The following propositions are consequences of the implicit function theorem, Milnor's theorem on 1-manifolds, Sard's theorem, and the genericity of Morse functions (see [Blicher 1983] for details).

- 1) Almost every level set of a picture is a circle or a line
- 2) These 1-manifolds account for almost all of the brightness values; the rest are extrema or saddles (critical points).
- 3) Typically, pictures have isolated critical points (i.e. the critical points do not form blobs, lines, or accumulations).

Now, here's what all this means in terms of Fig. (frag). Choose a picture at random. (Say the picture is bounded by a rectangle  $R$  with interior  $V$ .) If it has no critical points, then all the level sets are diffeomorphic to (disjoint unions of) line segments (and not circles, which are the only other possibility by Milnor's result, cited above).

Suppose the picture does have critical points. Then "generically" the critical points are isolated.

First let's see what happens near such a critical point. By Morse's Lemma [Guillemin 1974, Hirsch 1976] we know that there is a coordinate system  $(u, v)$  in a neighborhood of the critical point  $p$  such that  $f = f(p) \pm u^2 \pm v^2$ . The possible signs correspond to a maximum ( $--$ ), a minimum ( $++$ ), or a saddle ( $+-$  or  $-+$ ). So for an extremum, it's easy to see that the level sets are just a point surrounded by circles. For a saddle, the level sets are the sets  $u^2 - v^2 = \text{const}$ , shown in Fig. (saddle).

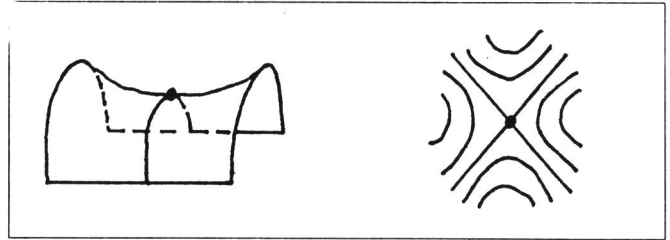


Fig. (saddle)

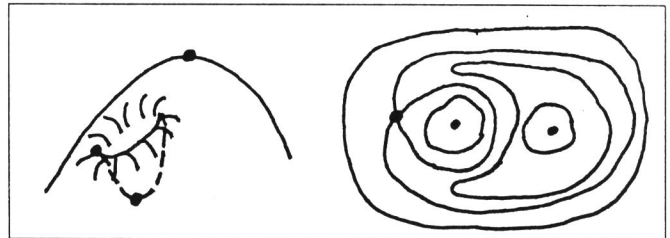


Fig. (dimple)

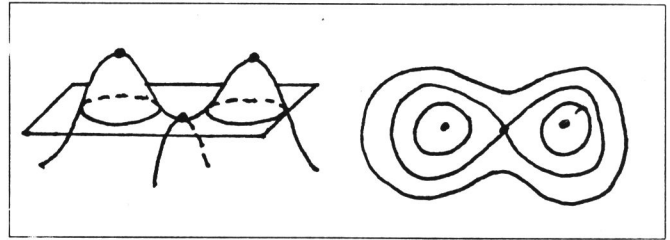


Fig. (pass)

The Morse inequalities tell us that the Euler characteristic is related to the number and type of critical points. In our case, if we assume that the whole region of interest lies within a single circular level set, this means that the number of extrema must be 1 more than the number of saddles. In Fig. (frag), for the rotation directions of the level sets to be consistent with the way we proved the first part of the theorem, we must assume that one of the critical points is a maximum and the other a minimum. But from the Morse inequalities, there must be a saddle somewhere, too. In fact, the larger picture looks like Fig. (dimple), and when there are two maxima (or minima, in Australia), like Fig. (pass).

### C. Open dense, usually, generically, almost all, typically

A crucial result used above is that the Morse functions are open dense. This allows us to restrict our attention only to pictures whose critical points are isolated and thus to avoid considering pathological behavior. A property shared by all members of an open dense subset (or a countable intersection of such subsets) is called *generic*, which can be thought of as "most" (see [Blicher 1983, Hirsch 1976, Nitecki 1971, Golubitsky 1973, Guillemin 1974]). Then the (countable) conjunction of generic properties is generic. "Generic" is a key idea in modern differential topology.

### D. The cases $n \geq 2$

Let  $f: M^m \rightarrow M^n$ , be  $C^r$  and regular at  $p$ . The analysis is based on the fact that at a regular point, if there is enough room in the range space,  $f$  is a diffeomorphism from a neighborhood  $U$  of  $p$  to  $f(U)$ . This is yet another version of the implicit function theorem. The idea of enough room can be made precise simply by requiring the Jacobian to be 1-1. This is the case for a

regular point if the dimension of the range space is at least that of the domain space, i.e. if  $m \leq n$ , which is the situation for us if there are at least 2 color dimensions.

As before, the possible maps  $g_\pi$  which solve the matching problem are exactly those which take level sets to level sets. Since the  $g_\pi$  are diffeomorphisms, we can just study the maps of the level sets of, say,  $F_1$ , since they are equivalent by a given  $g_\pi$  to the set of all  $g_\pi$ . (To see this, consider Fig. (equiv). Let  $h$  be a diffeomorphism which takes level sets to level sets, i.e. which makes the diagram commutative, and define  $g'_\pi = g_\pi \circ h$ , so that any  $h$  gives us a  $g'_\pi$ . Likewise given such a  $g'_\pi$ , define  $h = g_\pi^{-1} \circ g'_\pi$ .)

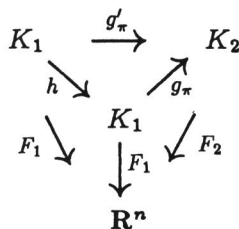


Fig. (equiv)

First, let's look at how many points can be in  $F_1^{-1}(p)$ . By the implicit function theorem, since the dimension of the range (i.e. the color space) is at least that of the domain, the level set of a regular value is at most a discrete set of points. Since we are restricting ourselves to compact pictures, the level set must be a finite set (to avoid an accumulation point). Hence on a level set,  $g_\pi$  is constrained to be one of a finite number of permutations of the finite level set. Furthermore, since  $F_1$  is a local diffeomorphism at a regular value, the permutation cannot jump around wildly among neighboring points, so that in fact  $g_\pi$  is a permutation of "sheets."

As it turns out, the higher dimensions are easier to deal with in our context, so we will start with them.

#### E. Regular points when $n \geq 3$

**Theorem.** Let  $M, N$  be embedded submanifolds of  $\mathbb{R}^n$ . Then generically,  $\dim M + \dim N - n = \dim M \cap N$ , where a negative dimension means the intersection is empty.

Locally, the regular sets are embedded submanifolds (by the inverse function theorem), so we can use the preceding to study the inverse images of regular values. In particular,  $F_1$  will fail to be 1-1 at places where the embedded regular sets intersect. We are interested in the case that  $\dim M = \dim N = 2$ , so we see that the intersection is generically of dimension 2, 1, 0, and empty for  $n = 2, 3, 4, 5$  resp. Thus if  $n \geq 3$  there is no diffeomorphism  $h$  other than the identity which makes Fig. (equiv) commute, i.e. such that  $F_1 = F_1 \circ h$ . So for the regular points, we have disposed of all the cases of 3 or more color dimensions. Now we look at the singular points, and their dimension.

The genericity of Morse functions can be generalized as follows.

**Theorem**(Critical set dimension). For an open dense subset of  $C^\infty(M^m, N^n)$ , the set of critical points of  $f$  where the Jacobian is of rank  $r$

- 1) comprise a submanifold of  $M^m$
- 2)  $= \emptyset$  if  $(m-r)(n-r) > m$
- 3) is of codimension  $(m-r)(n-r)$  in  $M^m$  if  $(m-r)(n-r) \leq m$

In our case,  $m = 2$ , so what the theorem tells us is that the dimension of the critical set is respectively 1, 0, for  $n = 2, 3$ , and it is empty for  $n \geq 4$ .

By reasoning as we did for multiple points of the regular set,  $h$  has unique continuation to the critical set for  $n \geq 3$ , yielding the conclusion that  $h$  is generically unique when  $n \geq 3$  (for  $n = 2$  the 1-1 set need not be dense, so the conclusion wouldn't follow).

To summarize, we have thus far shown that  $h$  must be the identity for  $n \geq 3$ , and is at worst one of a discrete set of sheet permutations for  $n = 2$ . Now we will pursue the case  $n = 2$  a bit further.

#### F. More about $n = 2$

If we allow the support of a picture to be all of  $\mathbb{R}^2$  or  $S^2$ , that is all we can say. But a real picture must be finite in extent, so if we are considering subsets of the plane, a rectangle (i.e. a disc) is an appropriate domain to consider. In that case,  $h$ , being a homeomorphism, must map the boundary of the disc (a circle  $S$ ) to itself. If  $f$  is 1-1 then  $h$  must be the identity. If not, then consider what must happen on this circle.  $h$  must be continuable along  $S$ , so for  $p \in S$ ,  $f^{-1}(p)$  must contain a constant number of points. This excludes the possibility of transverse crossings of  $f(S)$ . But transverse crossings for such a map are generic, so  $h$  must therefore generically be the identity. **QED**

## REFERENCES

- [Abraham 1978] Abraham, R. and J.E. Marsden, **Foundations of Mechanics**, 2nd ed., Benjamin/Cummings, Reading, Mass., 1978. [QA805.A2 1977, ISBN 0-8053-0102-X].
- [Blicher 1983] Blicher, A.P., "Computer Vision from the Topological Viewpoint," Stanford Univ., AI Memo and CS report, spring/summer 1983. (number unavailable at press time.)
- [Golubitsky 1973] Golubitsky, M. and V. Guillemin, **Stable Mappings and their Singularities**, (Graduate Texts in Mathematics 14). Springer, New York, 1973. [QA613.64.G64, 516.36, ISBN 0-387-90073-X (soft), ISBN 0-387-90072-1 (hard), ISBN 3-540-90073-X (soft)]
- [Guillemin 1974] Guillemin, V. and A. Pollack, **Differential Topology**, Prentice-Hall, Englewood Cliffs, N.J., 1974. [QA613.5.G84, 514'.7, ISBN 0-13-212605-2].
- [Hirsch 1976] Hirsch, M.W., **Differential Topology**, Graduate Texts in Mathematics, vol. 33, Springer, New York, 1976.
- [Nitecki 1971] Nitecki, Z., **Differentiable Dynamics**, MIT Press, Cambridge, 1971. [ISBN 0-262-14009-8 (hard), ISBN 0-262-64011-2 (paper)].