Generation of Dense Range Maps by Data Fusion from Active and Passive Colour Stereo Vision

Alois Knoll, Ralf Schröder, and André Wolfram

University of Bielefeld, Faculty of Technology,
Department of Computer Science,
Postfach 10 01 31, D-33501 Bielefeld, Germany

Abstract

The problem of correspondence search in a stereo system may be alleviated if an active light source projects colour codes onto the scene. To reduce the amount of ambiguous or false matches, additional information drawn from feature-based passive stereo techniques can be used. In this paper we discuss the application of fuzzy data fusion techniques to a dual-mode stereo system. Information from passive stereo is used whenever the active procedure cannot find an unambiguous match. The algorithm presented in the paper uses a fuzzy rule-based approach to integrate crisp as well as fuzzy input data. It is shown that the combination of data obtained in both modes of operation may lead to a significant increase in the number of matches meeting a certain credibility criterion. We present some of our experimental results based on real stereo image data and conclude the paper with a discussion of perspectives for possible future research.

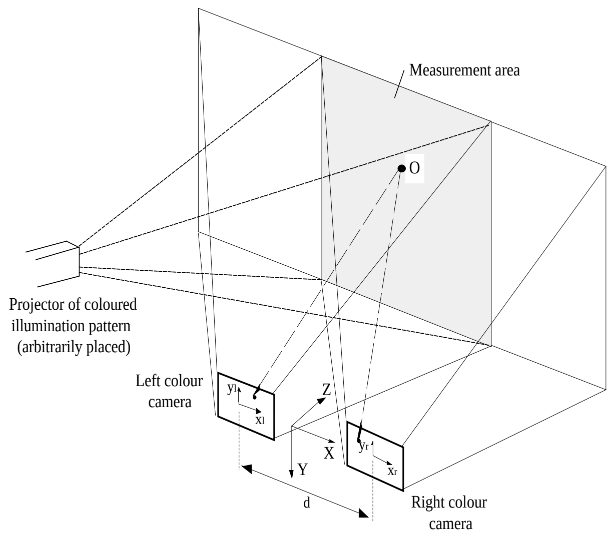

Figure 1: The basic experimental set-up

1 Introduction

In general, passive stereo algorithms are more reliable in matching pixels in the vicinity of features such as edges than matching pixels belonging to large and unstructured regions. By contrast, active procedures as developed in [1, 2, 3] use light codes projected onto the objects in the scene to identify corresponding pixels even in completely featureless areas (with different resolutions). In our system the code pattern consists of light with colour hues changing continuously along the horizontal direction of the stereo images, while the hue of the pattern remains constant in the vertical direction. Searching for corresponding pixels is therefore reduced to locating identical colour hues on corresponding epipolar lines determined by the configuration of the stereo cameras. As opposed to the passive approach the active method suffers from deficiencies pertaining to the identification of correspondences in the neighbourhood of sharp edges, e.g. that produce occluding contours not visible in the passive case. It is therefore desirable to combine the advantages of both approaches, thereby eliminating these disadvantages. The experimental setup is shown in Fig. 1. Further advantages of this particular setup are discussed in [3] and references are given.

Since both the active and the passive mode of operation require two stereo cameras, the stereo rig of Fig. 1 is used in both modes. The rig records the images from the scene while it is illuminated by colour codes as well as the colour images resulting from ambient lighting. To produce the colour codes, the setup includes a projector (in our case a standard slide projector). Since the active system requires less computational effort for calculating object point distances it is used as long as colour codes can be found in the scene. Once the measurements become erratic or ambiguous the algorithms for the evaluation of the data of the passive system are invoked and combined with the ambiguous active results to yield improved estimates of the true locations of correspondences.

Our experiments show that when an ambiguous match occurs, the passive stereo is very frequently capable of resolving the ambiguity, in particular if the pixels under consideration are located in the vicinity of characteristic features such as edges. Fusion between data of active and passive stereo algorithms may result in a more dense range map than could be achieved by each of the individual procedures.

2 Processing Steps

The active system matches colour hues to find corresponding object points. The passive technique extracts "straight edges" from the images. Information on edge attributes such as length, angle, etc. is exploited to find corresponding edges. Both the active and the passive stereo procedure comprise the following processing steps:

- **Acquisition of image data.**
  Two colour cameras are used for recording the images of illuminated objects (in the active case) or objects reflecting only the ambient light (in the passive case).
- **Low-level processing.**
  Determination of the colour code for all pixels (active case); estimation of gradients in the neighbourhood of each pixel and determination of the angular deviation among these gradients. Based on this information identification of points in the vicinity of edges (passive case).
- **Extraction of line segments.**
  Computation of epipolar lines that contain the corresponding pixels (active case); extraction of straight edge segments composed of the edge points mapped in the previous stage (passive case).
- **Correspondence search.**
  In the active procedure the colour codes of pixels along corresponding epipolar lines are compared, while the passive procedure compares the attributes of extracted straight edges.
- **Generation of range maps.**
  For related pixels, i.e. those marked as corresponding in the previous step range values may be calculated and visualized in a range map.

3 Fusion Objectives

The fusion of data acquired by the active and passive stereo algorithms may be established between various corresponding lower or higher processing levels. Depending on the way the data integration is organised, several different tasks may be performed:

- **Filtering of the input data.**
  For the purpose of generating range values for contour edges of an object, the input data of the active procedure may be restricted, for example, to those pixels labelled by the passive procedure as being located in the vicinity of an edge (the active procedure supports the passive procedure in order to examine *boundaries* more closely). Alternatively, the active procedure is applied only to those regions that are limited by continuous lines (*regions* are examined more closely).
- **Simplification of the search process.**
  As the illumination of objects is completely different in both cases, the information sets gathered from both procedures are in some sense "orthogonal", i.e. may complement each other. The feature-based passive procedure may profit from additional colour information at object lines, whereas the intensity-based active procedure may benefit from information on edge locations during the colour matching process. Decisions made while the passive algorithm progresses (e.g. an image point is close to an edge, for which possibly another correspondence candidate has already been found) can be utilized in an iterative process to resolve ambiguities of the active procedure. Conversely, if corresponding pixels were found by the active procedure lying on (straight) edges, these edges are likely to be corresponding edges.
- **Improving the credibility of the input data.**
  Since there are a number of images of the same scene available and the effects of image noise are different for each image, the signal-to-noise ratio may be improved. Furthermore, edges created by shadows or edges of regions with a different hue may or may not be visible under various lighting conditions.
- **Improving the resulting output data.**
  The accuracy of the correspondence/range map may be improved if both methods are carried out independently and a certainty measure is computed for every pixel/region/object. Based on this measure better estimates of the true location may be obtained through fusion. A special case are data gaps: if in the results of one procedure data are missing, these gaps may be eliminated from the correspondence map by substituting the results produced by the other method. It is obviously desirable to mutually check the results of the output of both procedures and to register distance values in the correspondence map only if the distance between the results remains below a certain threshold or if the result of the fusion is above a certain credibility/plausibility.

Fusion methods aiming at improving the credibility of input data or at restricting the working area of the active procedure are not considered further. In the sequel we describe a fusion algorithm that utilizes two evaluations of images of the same scene with the first evaluation carried out by applying the active stereo algorithm, the second evaluation carried out (on the same images) by using algorithms known from passive stereo.

4 Fusion Algorithm

We now describe the algorithm for correspondence search using the active method (section 4.1). We then go on to detail the fusion of theses results with passive data (section 4.2).

4.1 Area-Based Active Stereo: Matching of Colour-Codes

In a first step the colour code of the pixel block centered around a *reference pixel* (whose associated distance is to be determined) is compared with the colour codes of all blocks along the epipolar line in the other image. The search space is restricted by employing a local disparity limit. To determine the similarity Ξ^* between two blocks H_F^{l} and H_F^{r} of adjacent colour pixels of intensity $I \in [0, 255]$ and centered around the pixels to be compared (l/r denoting the left/the right image and H_F/H_f a colour block/an intensity block with $f \in F = \{R, G, B\}$), the commonly used mean square error (MSE) or modifications of it based on a suitable normalisation may be used ($\forall f$):

$$\Xi_1^*(H_F^{\mathrm{l}}, H_F^{\mathrm{r}}) = 1 - \left[1 + \exp\left(-3 \, \frac{\sqrt{\zeta(H_F^{\mathrm{l}}, H_F^{\mathrm{r}})} - 100}{100} \right) \right]^{-1} \quad (1)$$

$$\text{with} \quad \zeta(H_F^{\mathrm{l}}, H_F^{\mathrm{r}}) = \frac{1}{3\,(2p+1)} \sum_f \sum_{j=-p}^{p} \left(I_{f_j}^{\mathrm{l}} - I_{f_j}^{\mathrm{r}} \right)^2$$

Alternatively, more complex measures have been studied in detail [4], e.g. a similarity measure based on fuzzy integration of a suitable intensity-based similarity measure Ξ (such as the normalised MSE above) over $F' = \{(H_R^{\mathrm{l}}, H_R^{\mathrm{r}}), (H_G^{\mathrm{l}}, H_G^{\mathrm{r}}), (H_B^{\mathrm{l}}, H_B^{\mathrm{r}})\}$ with respect to a g_λ-fuzzy measure (which models the credibility of the individual colour channels) ($f' \in F'$) :

$$\begin{aligned} \Xi_2^*(H_F^{\mathrm{l}}, H_F^{\mathrm{r}}) &= \oint_{F'} \Xi(f') \circ g_\lambda(\cdot) \quad (2) \\ &= \sup_{A \in \mathcal{P}(F')} \min[\min_{f' \in A}(\Xi(f')), g_\lambda(A)] \end{aligned}$$

As long as this procedure yields only unambiguous matches, they are marked accordingly and used to calculate range values. Ambiguous matches may occur during the search for corresponding pixels due to various reasons:

- The fuzzy integral can propagate identical similarity measure values for slightly different colour codes of the compared line blocks. This effect may be reduced, if definitions other than the classical form of the fuzzy integral (2) are used. Examples are other t-norms, e.g. the algebraic product instead of the minimum-operator.
- Significant changes of depth in the stereo scene may cause identical colour codes to appear multiply along a single epipolar line (even if the coloured light pattern itself is non-repetitive).
- The colour of objects in the scene may change the projected light pattern by absorption in such a way that several identical colour codes are generated.

In the case of ambiguous matches the following processing steps are carried out for the pixels under consideration: For the point in the left picture, firstly a list of possible matches i (with m list elements) along the corresponding epipolar line in the right picture is generated. The elements are selected according to the local maxima of the pixel colour similarity measure.

These possible matches are then examined further, first by considering a measure for the smoothness of range values: Assuming that the reference point in the left picture and every list entry i representing a point in the right picture are actual corresponding points, the range d_i of the object point related to the pair of image points is calculated by triangulation. Furthermore, an average range value $\underline{d}$ at the position of the considered reference point in the range map is determined using entries of the range map calculated previously within a certain neighbourhood and the difference

$$\Delta\sigma = |\underline{d} - d_i| \quad (3)$$

is computed to evaluate the consistency.

Defining a *measure for smoothness* from the range data using fuzzy sets is useful because there are no crisp criteria for up to which limit a difference in range should be regarded as smooth. Moreover, the average range may grossly differ from the true range at a specific point and therefore could indicate a possible match creating a smooth surface. Therefore the range values $\underline{d}$ and d_i are fuzzified resulting in two fuzzy values $\widetilde{\underline{D}}$, $\widetilde{D_i}$ with membership functions $\mu_{\widetilde{\underline{D}}}(d)$, $\mu_{\widetilde{D_i}}(d)$ and rated by the measure

$$\sigma(\widetilde{\underline{D}}, \widetilde{D_i}) = \max_d \min[\mu_{\widetilde{\underline{D}}}(d), \mu_{\widetilde{D_i}}(d)] \quad (4)$$

with d denoting range. Furthermore, gradient information is taken into account, which is obtained as described in the next section.

4.2 Fusion with Information from the Feature-Based Passive Stereo

Two additional measures are computed based on information drawn from low-level processing of the feature-based passive method.

4.2.1 Measuring the angular deviation of gradients

A measure is calculated for the *angular deviation* of the gradients in the neighbourhood of the reference pixel as well as all possible matches. This measure has been derived from a method presented in [5]. For each pixel in the neighbourhood of the point of interest, gradient vectors are estimated using Sobel operator masks. Summation of all these gradients $\mathcal{G}$ and their absolute values $|\mathcal{G}|$ leads to the angular deviation δ indicating the presence of an edge in the vicinity of the considered point if the angular deviation is low with respect to the gradients evaluated (for reasons of simplicity the colour channel index f is omitted):

$$\delta(\mathcal{G}) = 1 - \frac{|\sum \mathcal{G}|}{\sum |\mathcal{G}|} \tag{5}$$

Considering the angular deviation for both pixels of a possible match three slightly different cases may occur:

- **Grossly different values for $\delta(\mathcal{G}^l)$ and $\delta(\mathcal{G}^r)$.**
  With one of the related pixels being located in the vicinity of an edge while the other is not, the pixels are unlikely to be corresponding pixels.
- **High values for $\delta(\mathcal{G}^l)$ and $\delta(\mathcal{G}^r)$.**
  Both pixels are likely to be located in image regions without features suitable for the verification of possible correspondences. There are no edges to compare and the directions of the gradients vary randomly.
- **Very low values for $\delta(\mathcal{G}^l)$ and $\delta(\mathcal{G}^r)$.**
  Both pixels are likely to be located in the vicinity of an edge and thus are probably actual corresponding pixels.

A measure α which indicates whether a possible match belongs to a pair of corresponding edge points or at least edge region points may be defined as:

$$\alpha\left(\delta(\mathcal{G}^l), \delta(\mathcal{G}^r)\right) = 1 - \max\left[\delta(\mathcal{G}^l), \delta(\mathcal{G}^r)\right] \tag{6}$$

4.2.2 Measuring the similarity of gradients

Gradients in the neighbourhood of the reference pixel in the left image and of the possibly corresponding pixel are compared to evaluate a further similarity measure γ, which is constructed from intersections of the fuzzified components of all gradient vectors concerned (denoted as $\widetilde{\mathcal{G}}^l$ and $\widetilde{\mathcal{G}}^r$, resp.). This is done separately for the X- and Y-components in order to consider the direction of the gradients as well as their magnitude and their normalized average is computed (this will be denoted as $\underline{\gamma}$). Particularly in case the angular deviation indicates the presence of edges in the vicinity of the possibly corresponding pixels, the *measure for the similarity of gradients* can be interpreted as a measure for the similarity of the edges close to the regarded pixels. If, for example, the pixels of a possible match can be identified as being close to edges with significantly differing directions, these pixels are not likely to represent an identical object point.

4.2.3 Determination of the best match by fusion of the similarity measure values

In this step the credibility of all possible matches i, $i = 1, \ldots, m$ is assessed by two different methods:

Method 1: Simple aggregation of all similiarity values separately for each of the three colour channels $f \in F$ by the minimum-operator as proposed in [6]:

$$\begin{aligned} \epsilon_{f_i} \quad = \quad & \min\Big[\Xi_2^*(H_F^{\mathrm{l}}, H_{F_i}^{\mathrm{r}}),\ \sigma(\underline{\widetilde{D}}, \widetilde{D_i}), \\ & \alpha(\delta(\mathcal{G}_f^l),\ \delta(\mathcal{G}_{f_i}^r)),\ \gamma(\widetilde{\mathcal{G}}_f^l, \widetilde{\mathcal{G}}_{f_i}^r)\Big], \quad \forall i, f \end{aligned} \tag{7}$$

This "and-ing" of the individual ratings implies a very pessimistic (or safe) decision for correspondences. The disadvantage of this method is its inflexibility: no specific scene-related a priori knowledge can be taken into account.

Method 2: A rule-based approach was utilised to overcome this inflexibility. The following rules were used:

R1: "IF the difference between angular deviations is HIGH THEN set output to LOW."

If one pixel is positioned in the vicinity of an edge (indicated by a low angular deviation), whereas a possible match is not (i.e. angular deviation is high), both pixels are unlikely to represent an identical world point. Therefore the overall rating should be reduced. (More commonly the angular deviation may also be interpreted as a measure for the local signal-to-noise ratio).

R2: "IF the difference between angular deviations is MEDIUM THEN set output to MEDIUM."

Like with rule R1 the similarity between angular deviations is evaluated but possible matches with these characteristics are at least more likely to be an actual correspondence. Propagating a medium output is therefore justified.

R3: "IF the difference of ranges is LOW AND both deviations are LOW THEN set output to HIGH."

If the range calculated for a possible match is only little different from a local average range value, fitting this range value into the map may provide a smooth surface. The smoothness is further used to determine, whether or not the matched pixels are likely to be positioned near the same edge (low deviations for both pixels just indicate positions near any edge). The overall rating should be rather high, if both matched pixel are positioned near edges and also contribute to a smooth surface in the range map.

R4: "IF the difference of ranges is HIGH OR at least one deviation is HIGH THEN set output to LOW."

This rule reduces the overall rating, if one of the constraints given in rule R3 is grossly violated. The motivation for this reaction is therefore the same as given for rule R3.

R5: "IF the intersection of gradient-trapezoids is HIGH THEN set output to HIGH."

A local similarity of gradients at the positions of possible matches may not be sufficient to identify the correct correspondence without further knowledge but this rather unspecific information may be useful to prevent confusing light/dark-transitions with dark/light-transitions.

The models for the input/output fuzzy sets used in the experiments are shown in Fig. 2 ("difference of ranges" (a), "intersection of gradient-trapezoids" resp. "difference between angular deviations" with $\Delta\delta = |\delta(\mathcal{G}^l) - \delta(\mathcal{G}^r)|$ (b) and "output" (c)). The logical operators "AND" and "OR" in rules R3 and R4 were modelled as min and max. The ratings ϵ_{f_i} are determined through defuzzification (center of gravity) of the resulting fuzzy set [7].

In both cases the highest ratings

$$\hat{\epsilon}_f = \max_i \epsilon_{f_i}, \quad \forall f \tag{8}$$

indicate the possible matches which are most likely to be the actual correspondence considering gradient information of just a single colour channel f.

For the final decision we select the match with the overall maximum

$$\epsilon_F = \max_f \hat{\epsilon}_f \tag{9}$$

because either a channel detecting only image noise instead of an edge will produce a very low rating, or some edges may only be visible in a single colour channel due

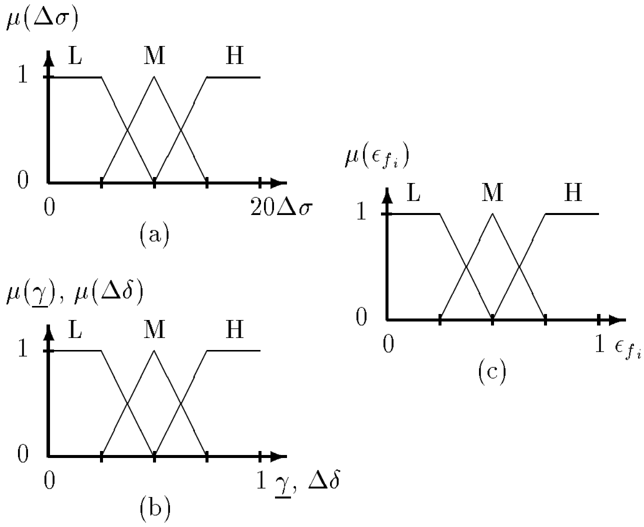

Figure 2: Input/output fuzzy sets

to different absoption of the light pattern projected onto the scene.

That match is then regarded as the new reference pixel in the right picture. The whole procedure of finding the match with the greatest similarity is repeated, comparing this new reference with possible matches on an epipolar line in the left picture. If the pixel with the highest rating, determined after completion of the second run, coincides with the original reference pixel (within a small tolerance), these pixels are most likely to be actually corresponding, and therefore this match is marked as unambiguous. Otherwise the status of the regarded pixel remains ambiguous.

5 Experimental Results

We evaluated stereo images of a size of 325×355 pixels (a white styrofoam part used in packaging, see Fig. 3). Using the mean square error for comparing the pixel colour codes of line blocks of width 17 and the similarity measure Ξ_1^* of (1), the results shown in Tab. 1 were obtained. Out of the total of 115375 pixels for 38793 no matches were found. They were not considered further. The colour-coded active stereo procedure produced 35937 ambiguous and 40645 unambiguous matches, i.e. 76582 or 66% of all pixels were matched. Note that this

Table 1: Experimental results

Procedure:	Active stereo	FRB-Fusion applied
Pixels overall:	115375	
No match:	38793	
Ambiguous:	35937	19124
Unambiguous:	40645	57458

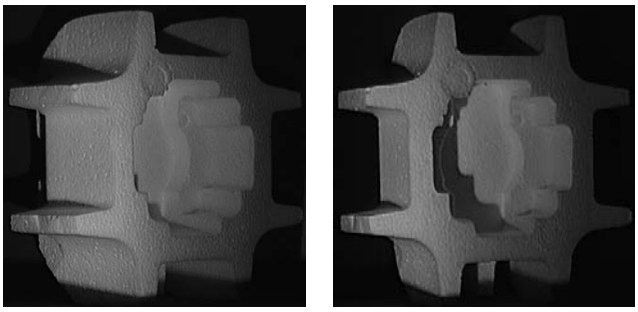

Figure 3: Test scene, left and right image

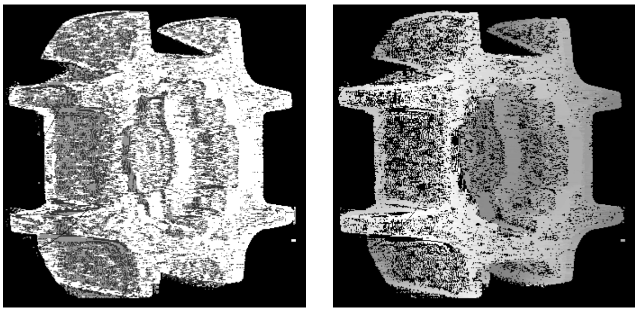

Figure 4: Correspondence (left) and range map (right)

refers to the whole image but that for the black background areas included in the images in Fig. 3 no sensible matches can be found. The additional use of the fuzzy rule-based fusion approach (FRB) generates 16813 more unambiguous matches: out of the 76582 matches now 57458 are unambiguous while only a remaining 19124 are ambiguous. In Fig. 4 (left) unambiguous matches appear as white, ambiguous matches as light grey, those produced as unambiguous by fusion are marked as dark grey. Black areas contain pixels for which no matches were found. Fig. 4 (right) shows the corresponding range map (brighter grey levels represent shorter distances between the object and the stereo camera system).

6 Conclusions

The results presented in this paper indicate a clear improvement over the pure active stereo procedure by reducing ambiguities in the correspondence search. However, fusion has only been applied to the problem of reducing ambiguity on a pixel level. It will be investigated in the future how information on object points or feature locations can be fused after independently processing the input data of both methods.

The amount of information drawn from the passive method is currently rather small. It remains to be seen how the evaluation of additional processing results from passive methods can be incorporated to improve the results further. To this end some of the methods as described in [8, 9, 10, 11] are currently being implemented and will be tested on scenes similar to that shown in Fig. 3. Moreover, the rules described in section 4.2.3 will be refined and heuristics for the selection of thresholds will be developed.

References

[1] Boyer, K.L. and Kak, A.C., "Color-encoded structured light for rapid active ranging", *IEEE Trans. on PAMI*, Vol. 9, No. 1, pp. 14-28, 1987.

[2] Vuylsteke, P., Price, C.B., and Oosterlinck, A., "Three dimensional image acquisition", in Henderson, T.C. (ed), *Trad. and Non-Trad. Rob. Sens.*, NATO ASI, Vol. F63, Springer, pp. 187-210, 1990.

[3] Knoll, A. and Sasse, R., "An active stereometric triangulation technique using a continous colour pattern", in Straßer, W. and Wahl, F.M. (eds.), *Graphics and Robotics*, Springer, pp. 191-206, 1995.

[4] Knoll, A. and Wolfram, A., "Searching correspondences in colour stereo images - an approach using the fuzzy integral", *Proc. FUZZ-IEEE/IFES'95*, Yokohama, Japan, pp. 2033-2038, 1995.

[5] Gregson, P.H., "Using angular dispersion of gradient direction for detecting edge ribbons", *IEEE Trans. on PAMI*, Vol. 15, No. 7, pp. 682-696, 1993.

[6] Knoll, A., Schröder, R., and Wolfram, A., "Fusion of data from fuzzy integral-based active and passive colour stereo vision systems for correspondence identification", *Proc. EUSIPCO-96*, Trieste, Italy, pp. 1733-1736, 1996.

[7] Klir, G.J. and Yuan, B., *Fuzzy sets and fuzzy logic, theory and applications*, Prentice Hall, 1995.

[8] Bolles, R.C., Baker, H.H., and Hannah, M.J., "The JISCT stereo evaluation", in *Proc. 1993 Image Understanding Workshop*, DARPA, University of Maryland Research Projects, pp. 263-274, 1993.

[9] Kanade, T. and Okutomi, M., "A stereo matching algorithm with an adaptive window: theory and experiment", in *Proc. 1991 IEEE Int. Conf. on Robotics and Automation*, Sacramento, California, pp. 1088-1095, 1991.

[10] Kanade, T. "Development of a video-rate stereo machine", in *Proc. 1994 Image Understanding Workshop*, ARPA, Vol. I, Monterey, California, pp. 549-557, 1994.

[11] Matthies, L., "Stereo vision for planetary rovers: stochastic modeling to near real-time implementation", *Int. J. of Computer Vision*, Vol. 8, No. 1, pp. 71-91, 1992.