The Non-Parametric Parzen's Window in Stereo Vision Matching

Gonzalo Pajares and Jesús M. de la Cruz

Abstract—This paper presents an approach to the local stereovision matching problem using edge segments as features with four attributes. From these attributes we compute a matching probability between pairs of features of the stereo images. A correspondence is said true when such a probability is maximum. We introduce a nonparametric strategy based on Parzen's window to estimate a probability density function (PDF) which is used to obtain the matching probability. This is the main finding of the paper. A comparative analysis of other recent matching methods is included to show that this finding can be justified theoretically. A generalization of the proposed method is made in order to give guidelines about its use with the similarity constraint and also in different environments where other features and attributes are more suitable.

Index Terms—Bayes, matching, Parzen's window, stereovision.

I. INTRODUCTION

A significant amount of research in the computer vision community has been aimed at the study of the three-dimensional (3-D) structure of objects using machine analysis of images [1]. Analysis of video images in stereo has become an important passive method for extracting the 3-D structure of a scene.

The key step in stereovision is that of image matching, namely, the process of identifying the corresponding points in two images that are generated by the same physical point in space. This paper is devoted solely to this problem. The stereo correspondence problem can be defined in terms of finding pairs of true matches that satisfy three competing constraints: similarity, smoothness and uniqueness [2]. The similarity constraint is associated with a local matching process where a minimum difference attribute (properties of features) criterion is applied. The results computed in the local process are later used by a global matching process where other constraints such as smoothness [2], minimum differential disparity [3] and figural continuity [4] are imposed. A good choice of local matching strategy is the key to good results in the global matching process.

This paper presents an approach to the local stereopsis correspondence problem by estimating a probability density function (PDF) via the nonparametric Parzen's window, with which a matching probability is computed. This probability is used as the criterion to classify the matches as true or false correspondences. This is the main finding reported in this paper. In order to match pairs of features, we estimated a PDF via the parametric Bayesian approach through a likelihood method as in [5]. This Bayesian PDF was later used in a global matching strategy implemented by the Hopfield neural network of [6]. This network maps the similarity, smoothness and uniqueness constraints in an energy function, which is then minimized to find the optimal correspondence. The similarity constraint is mapped using the Bayesian PDF given in [5]. When we map the similarity constraint in [6] using the PDF estimated from Parzen's window (instead of the

Bayesian PDF estimated in [5]) the results are improved. The mapping of the other two constraints in [6] is left unchanged. Also, an improvement in the results is found when the Parzen approach is compared to other recent matching methods using only the similarity constraint [5], [7], [8]. This justifies the choice of Parzen's window as a good method to measure the similarity between features.

Two types of techniques have been broadly used for stereovision matching, namely the correlation-based and the feature-based methods. In the correlation-based method, the elements to be matched are image windows of fixed size and the similarity criterion is a measure of the correlation between windows in the two images. The window that maximizes the similarity criterion within a search region gives the corresponding element. The number of pairs of features to be considered becomes high, because all pixels in the left image must be matched with all pixels in the right one. The feature-based methods use sets of pixels with similar attributes, usually either pixels belonging to edges or the corresponding edges themselves. Instead of image windows they use numerical and symbolic properties of features. The feature-based methods lead only to a sparse depth map, leaving the rest of the object surface to be reconstructed by interpolation. They are faster than area-based methods because there are fewer points (features) to be considered. See [9] and associated references for both types of techniques.

The matching is made difficult in part by changes in the images of the corresponding points due to different viewpoints and to the different physical cameras. Hence, the corresponding attributes in the two images may display different values. This may lead to incorrect matches. Thus, it is very important to find features in both images that are independent of possible variations in the images. Our experiment has been carried out in an indoor space where edge segments are abundant and are therefore the most appropriate features [10].

Moreover, they have been used in previous stereovision matching works [1], [3], [5]–[8], and in [9] we can find additional references where edge segments are used. Hence, our choice of features is justified, although such features are produced by intensity changes. Four attribute values (modulus and direction vector of the gradient, Laplacian and variance) are computed for each edge segment. This is explained in Section II-A.

This paper is organized as follows. In Section II the stereo matching system is considered, which comprises three stages:

- 1) extraction of features and attributes;
- 2) PDF estimation through Parzen's Window;
- 3) matching for the current stereo-pairs.

To show the effectiveness of the proposed method, in Section III a test strategy is designed and a comparative analysis among other existing strategies is carried out. A generalization of the method is proposed in Section IV. Finally, in Section V the conclusion is presented.

II. PARAMETRIC PARZEN'S WINDOW IN STEREOVISION MATCHING

Our local stereo matching system is designed with a parallel optical axis geometry and works in three stages:

- 1) extracting information (features and attributes) from the images;
- 2) performing an estimation (training) process with the samples (true matches) to estimate a PDF through Parzen's window. We call this the OFF-LINE process;
- 3) performing a matching process for the current pairs of features. We call this the ON-LINE process.

The first stage is common for both the OFF-LINE and the ON-LINE processes.

Manuscript received December 8, 1999; revised March 26, 2001, June 24, 2001, and November 6, 2001. This work was supported in part by Project CICYT TAP94-0832-C02-01. This paper was recommended by Associate Editor P. K. Willett.

The authors are with the Departamento Arquitectura de Computadores y Automática, Facultad de Ciencias Físicas, Universidad Complutense, 28040 Madrid, Spain (e-mail: pajares@dacya.ucm.es; jmcruz@dacya.ucm.es).

Publisher Item Identifier S 1083-4419(02)00704-5.

A. Feature and Attribute Extraction

The contour edge pixels in both images are extracted using the Laplacian of the Gaussian filter in accordance with the zero-crossing criterion, [11]. At each zero-crossing in a given image we compute the magnitude and the direction of the gradient vector as in [12], the Laplacian as in [13] and the variance as in [14]. These four attributes are computed from the gray levels of a central pixel and its eight immediate neighbors. To find the gradient magnitude of the central pixel, we compare the gray level differences from the four pairs of opposite pixels in the eight-neighborhood; the largest difference is taken as the gradient magnitude. The gradient direction of the central pixel is the direction out of the eight principal directions whose opposite pixels yield the largest gray level difference, and also points in the direction in which the pixel gray level is increasing. The normalization of the gradient direction is achieved by assigning a digit from 0 to 7 to each principal direction. In order to avoid noise effects during edge-detection that can lead to later mismatches in realistic images, the following two globally consistent methods are used: 1) the edges are obtained by joining adjacent zero-crossings following the algorithm in [15], in which a margin of deviation of $\pm 20\%$ and $\pm 45^\circ$ is tolerated in magnitude and direction respectively; 2) then each detected contour is approximated by a series of line segments as in [16]; finally, for each segment an average value for the four attributes is obtained from all computed values of its zero-crossings. All average attribute values are scaled, so that they fall within the same range. Each segment is identified by its initial and final pixel coordinates, its length and its label.

Therefore, each stereo-pair of edge-segments has two associated four-dimensional (4-D) vectors $\mathbf{x}_l$ and $\mathbf{x}_r$, where the components are the attribute values and the subindexes l and r which denote features belonging to the left and right images respectively. A 4-D difference vector of the attributes $\mathbf{x} = \{x_m, x_d, x_p, x_v\}$ is obtained from $\mathbf{x}_l$ and $\mathbf{x}_r$, whose components are the corresponding differences for the modulus of the gradient vector, the direction of the gradient vector, the Laplacian and the variance respectively.

B. Parzen's Window Density Estimator

Here, the goal is to estimate a PDF associated to true matches as a function of theoretical patterns $\mathbf{x}$, where $\mathbf{x}$ is the difference vector defined above. We use the Parzen's window density estimator taking into account that, due to the nature of the matching problem, the samples come from the class of true matches, w_T . The pattern classifier assigns the input vector $\mathbf{x}$ to w_T and according to the Bayes classifier [17], [18] the objective is to maximize the function $f(\mathbf{x}, w_T) = p(\mathbf{x}/w_T)p(w_T)$ where $p(\mathbf{x}/w_T)$ is the conditional probability of observing $\mathbf{x}$ if the class is w_T , and $p(w_T)$ is the *a priori* probability of w_T . One of the ways of estimating this quantity is via the nonparametric Parzen technique [17]–[19]. We proceed by placing a region R of volume V , centered at pattern $\mathbf{x}$. The points $\mathbf{x}$ in the space where the density is to be estimated are called theoretical patterns. The estimate of the density function at $\mathbf{x}$ is defined as

$$\hat{p}(\mathbf{x}/w_T) = \frac{k}{n_T V} \quad (1)$$

where k is the number of training patterns falling in the region R and n_T is the total number of training patterns, both related to w_T . The volume V is previously fixed, and the value of k is then obtained by counting the number of training patterns falling in the volume.

Let $X = \{\mathbf{x}_1, \dots, \mathbf{x}_{n_T}\}$ be the set of n_T 4-D training patterns belonging to w_T . The Parzen estimate of the density at $\mathbf{x}$ is given by [20]

$$\hat{p}(\mathbf{x}/w_T) = \frac{1}{n_T} \sum_{j=1}^{n_T} K_h(\mathbf{x} - \mathbf{x}_j) \quad (2)$$

where $K_h(\mathbf{x}) = (K(\mathbf{x}/h)/V_h)$, $K(\cdot)$ is a window or kernel function with $\int_{-\infty}^{\infty} K(t) dt = 1$ (to ensure the area under the probability density function is 1) and h is the window-width or smoothing parameter, determining the spread of the kernel function, ($h > 0$) [21]. Typically, the kernel is a monotonically decreasing function g of the distance of its argument from 0. That is

$$K(\mathbf{x} - \mathbf{x}_j) = g(d(\mathbf{x}, \mathbf{x}_j)) \quad (3)$$

where $d(\mathbf{x}, \mathbf{x}_j)$ is a metric distance. The statistical properties of this estimator can be found in [18].

Equation (2) shows that the contribution of each training pattern $\mathbf{x}_j$, toward the density estimate at a point $\mathbf{x}$ is determined by the width and the shape of the kernel. For example, if the window is Gaussian then patterns that fall close to $\mathbf{x}$ contribute more toward the estimate of the density at $\mathbf{x}$ than the patterns that are far away from $\mathbf{x}$. Such an equation can be interpreted as centering the kernel K over each training pattern and summing the overlapping functions at each point $\mathbf{x}$ in the space. Therefore, this process of centering kernels over training patterns and summing is a convolution [17]. It happens that the expected value $\hat{p}(\mathbf{x}/w_T)$ is the convolution of $p(\mathbf{x}/w_T)$ and K . To compute $f(\mathbf{x}, w_T)$ we also need to estimate $p(w_T)$. According to [18] a reasonable estimate is the fraction of training patterns in class w_T ,

$$\hat{p}(w_T) = \frac{n_T}{N} \quad (4)$$

where N is the number of pairs of features extracted by the system according to Section II-A and n_T is the number of true matches classified by the system after each ON-LINE process. With this, we get the following estimate of $f(\mathbf{x}, w_T)$:

$$\hat{f}(\mathbf{x}, w_T) = \hat{p}(\mathbf{x}/w_T)\hat{p}(w_T) = \frac{1}{N} \sum_{j=1}^{n_T} g(d(\mathbf{x}, \mathbf{x}_j)). \quad (5)$$

The shape of the kernel is not as important as its width in density estimation. There are different types of kernel functions [22]. In this paper we have chosen the commonly used Gaussian kernel for the following two reasons. The first is because with such a kernel, we have information about the cross-correlation between the four attributes through the embedded covariance matrix C and, as a consequence, about the discriminatory properties of such attributes. The second reason is its convenient analytical properties following (2). So, we get the following density estimate:

$$f(\mathbf{x}, w_T) = \frac{1}{N} \sum_{j=1}^{n_T} \left\{ \frac{1}{(2\pi)^{d/2}} \frac{1}{h_T^{d/2} |C|^{1/2}} \cdot \exp \left[\frac{-1}{2h_T^2} (\mathbf{x} - \mathbf{x}_j)^t C^{-1} (\mathbf{x} - \mathbf{x}_j) \right] \right\} \quad (6)$$

where d is the dimension of the pattern space (which in this paper is 4 as $\mathbf{x}$ is a 4-D vector); t denotes transpose. Each kernel function considers a possible cross-correlation between attributes, through C , and also that the samples are around a theoretical pattern $\mathbf{x}$ where the average vector for each kernel is $\mathbf{x}$ itself; $f(\mathbf{x}, w_T)$ is estimated for each $\mathbf{x}$ in the space and finally C is the covariance matrix of the Gaussian window. The smoothing parameter is often expressed as a function of the number of patterns

$$h_T = h n_T^{(-k)/d} \quad \text{for } 0 \leq k \leq 0.5. \quad (7)$$

The choice of the bandwidth h is very critical in Parzen density estimation [23]. An overly small h gives a spiky or noisy estimate of $p(\mathbf{x}/w_T)$ according to (2), with each spike corresponding to the kernel itself at the training patterns. When h is very large, each training pattern

provides essentially the same contribution toward density estimation at every point $\mathbf{x}$ and the result is an oversmoothed estimate of $p(\mathbf{x}/w_T)$. As shown in [23], the window-width parameter appears to be fixed according to the set of data from the stereo images. A more in-depth discussion of window width may be found in [18], [20], [22]. Indeed, in [18], [22] we can find a study in the unidimensional case of Parzen's window estimates of a Gaussian kernel for both n_T and h values, where good results are obtained with n_T set to 256 and h set to 4, although the best results are obtained for $n_T \rightarrow \infty$ for any h . From this study we have established the optimal h_T value through equation (7) with $n_T = 256$, $h = 4$ and for each k in steps of 0.1 ranging in the interval $[0, 0.5]$. The constraint for k is required in (7). So, since during the OFF-LINE process n_T is known, this allows us to select any k in $[0, 0.5]$ to obtain h from equation (7). This is done as the h value is required later. Therefore, the only parameter to be fixed in our approach is k . We have verified experimentally that a good choice is 0.4. With 0.2, 0.3, and 0.5 the results are also acceptable. The worst results are obtained with 0 and 0.1.

As mentioned before, $f(\mathbf{x}, w_T)$ is estimated for each $\mathbf{x}$ in the space. Therefore, we still require the grid of theoretical patterns $\mathbf{x}$ to be fixed, as well as the number of sample patterns belonging to the hyper-sphere centered at each $\mathbf{x}$ and finally the covariance matrix C for the samples in the corresponding hyper-sphere. The method is summarized as follows [22].

- 1) As we are only concerned with true matches, it is assumed that the difference vector of attributes $\mathbf{x}$ for any ideal true match is the null vector. Therefore, a good choice for the theoretical patterns should be a d -dimensional grid of points $\mathbf{x}$ in the space around the d -dimensional null vector. Nevertheless, in [22] it is proposed that the theoretical patterns could be the sample patterns. We choose this approach because it is simpler and also because we deal with the real samples directly.
- 2) Fix the d attribute values for the fixed theoretical patterns, i.e., the components of the difference vector $\mathbf{x}$, namely x_m, x_d, x_p, x_v .
- 3) Determine the number of training patterns belonging to the hyper-sphere centered at each theoretical pattern $\mathbf{x}$. Once again, following [22], a training pattern represented by its attribute difference vector $\mathbf{y}$ belongs to the hyper-sphere centered at $\mathbf{x}$ if $|(x_j - y_j)/h| \leq 1$ for all $j = m, d, l, v$, the attribute components. Therefore, the pattern $\mathbf{y}$ belongs to the hyper-sphere centered at $\mathbf{x}$ if none of its attribute values differ from the corresponding attribute values of $\mathbf{x}$ by a quantity greater than the parameter h .
- 4) With the samples $\mathbf{y}$ in the hyper-sphere of center $\mathbf{x}$ (i.e., mean $\mathbf{x}$) we compute the covariance matrix according to [5], [18], [22]

$$C = \frac{1}{N-1} \sum_{j=1}^N (\mathbf{y}_j - \mathbf{x})(\mathbf{y}_j - \mathbf{x})^t \quad (8)$$

where N is the number of training patterns $\mathbf{y}_j$ belonging to the hyper-sphere centered at $\mathbf{x}$, t denotes the transpose and $\mathbf{x}$ is the mean vector.

C. Current Stereo Matching Process

This is an ON-LINE or decision process in which two new features are to be matched. This is carried out by obtaining the 4-D difference vector of the attributes $\mathbf{x}$. With this $\mathbf{x}$ we compute the matching probability through the PDF estimated according to (6).

During the decision process there are unambiguous and ambiguous pairs of features, depending on whether a given left image segment corresponds to one and only one, or several right image segments, respectively. In any case, the decision about the correct match is

made by choosing the pair with the greater probability value (in the unambiguous case, there is only one) provided that it surpasses the threshold of 0.50. This value is the intermediate probability value in the interval where the probability ranges. This is the *uniqueness* stereovision matching constraint [2]. When the probability value for a pair of features (edge-segments) does not exceed 0.5, it is considered a false match.

III. COMPARATIVE ANALYSIS AND PERFORMANCE EVALUATION

To assess the validity and performance of our method, we designed a test strategy with four goals.

- 1) To verify that the matching probability increases as the number of training patterns increases.
- 2) To verify the improvement in the results for the global matching strategy in [6] when the PDF estimated from the nonparametric Parzen's window is used to map the similarity constraint instead of the PDF estimated from the parametric Bayesian method. This is explained in the Introduction.
- 3) To compare the performance of this nonparametric Parzen's window estimation to the parametric Bayesian approach [5], and also to the improvement in the decision process introduced in [5] and taken from [27].
- 4) To compare the performance of this local matching approach to other recent local learning matching strategies [7], [8].

A. Design of a Test Strategy

The objective is to test the method by varying indoor environmental conditions in two ways: by using new images with different features (different objects) and by changing the illumination. With this aim in mind a set SP0 of 12 pairs of stereo images captured with natural illumination was used to extract initial training patterns. Figs. 1–4 show four representative left images of this set. Five additional sets of stereo-images, SP1, SP2, SP3, SP4, and SP5, all different to each other and to SP0, were used for the test. They were composed of 10, 10, 15, 12, and 11 stereo images respectively. The sets SP1 and SP4 were captured with natural illumination, as was the initial set SP0, and the sets SP2, SP3, and SP5 with artificial illumination. Two representative stereo image pairs are shown for sets SP2 and SP3 in Figs. 5(a) and (b) and 6(a) and (b). The remaining stereo image pairs belong to the same indoor environment and it is irrelevant to show representative pairs. The total number of pairs of edge-segments extracted from all stereo images is 4638 (i.e., this is the number of potential true matches used by our system). This number of pairs of edge-segments is completely representative of the environment where our mobile robot, equipped with our stereovision system, navigated. From this number, during all the ON-LINE processes, 2925 pairs of edge-segments were classified as true matches. Therefore, this is the total number of training patterns n_T used during all the OFF-LINE processes.

The process can be summarized as follows.

Step 0: ON-LINE $\rightarrow$ Classify the pairs of features in the set SP0 as true or false matches by using the unsupervised learning strategy described in [5].

OFF-LINE $\rightarrow$ Estimate $f(\mathbf{x}, w_T)$ through (6) with the matches classified as true during the previous ON-LINE process.

For $n = 1$ to $n = 5$ do

STEP n: ON-LINE $\rightarrow$ Classify the pairs of features as true or false matches for the set SPn with the $f(\mathbf{x}, w_T)$ estimated during **STEP n - 1**

OFF-LINE $\rightarrow$ Estimate $f(\mathbf{x}, w_T)$ through (6) with the matches classified as true during all the ON-LINE processes in **STEPs 0 to n**.

Fig. 1. Left original training image (blocks).

Fig. 2. Left original training image (furnitures).

Fig. 3. Left original training image (computers I).

Fig. 4. Left original training image (computers II).

As explained below and only for testing purposes, the training patterns and the parameters are stored after each ON-LINE process, so that the estimated function in (7) can be recovered at each STEP.

The true matches required during STEP 0 can be supplied to the system by using any unsupervised stereovision matching method. We have used the method of [5], as it has already been tested. When this is not possible, the use of a minimum distance criterion is appropriate and the Euclidean distance is sufficient. Also, a human expert could provide such true matches, although this implies that the system loses its automatic capability at this stage.

We can point out that, as the matching probability for a true match increases, a better decision can be taken for its classification as a true match. Hence, to verify the behavior of the matching probability with the number of training patterns, we have used an additional set SP6.

Fig. 5. (a) SP2: original left stereo image, (b) SP2: original right stereo image, (c) SP2: labeled segments left image, and (d) SP2: labeled segments right image.

Fig. 6. (a) SP3: original left stereo image, (b) SP3: original right stereo image, (c) SP3: labeled segments left image, and (d) SP3: labeled segments right image.

This set comes from the same indoor environment, with only two additional stereo images, one with natural illumination and the other with artificial illumination, both different from all the above SP0 to SP5 sets. With the PDF estimated during the OFF-LINE process in STEP 5, the system classifies the true matches from SP6. Then, for testing purposes only, each one of these true matches is used as a theoretical pattern for computing its matching probability through (7). This is done for each STEP from 1 to 5 and carried out by recovering the training patterns and the parameters previously stored after each ON-LINE process.

For reasons of legibility, instead of using all matching probabilities, Fig. 7 shows the average matching probability at each STEP for all theoretical patterns from SP6 against the training patterns used at each STEP.

From Fig. 7, it is clear that the slope of the average matching probability in STEP's 4 and 5 is smooth compared with the corresponding slope for STEP's 1 to 3. This means that the training patterns used in

Fig. 7. Average matching probability for the theoretical patterns from SP6 against the number of training patterns.

TABLE I

RESULTS FOR THE THREE SETS OF STEREO PAIRS SP1, SP2, AND SP3. GP AND GB: GLOBAL METHODS IMPLEMENTED THROUGH THE HOPFIELD NEURAL NETWORK IN [6] WHERE THE SIMILARITY CONSTRAINT IS MAPPED TO THE CORRESPONDING ENERGY FUNCTION BY USING THE PDF COMPUTED VIA THE NON-PARAMETRIC PARZEN'S WINDOW AND THE PARAMETRIC BAYESIAN RESPECTIVELY. LP, LB, LV, LH: LOCAL MATCHING STRATEGIES USING ONLY THE SIMILARITY CONSTRAINT, CORRESPONDING TO PARZEN (P), AND BAYESIAN (B) ESTIMATORS, LEARNING VECTOR QUANTIZATION (V) AND HEBBIAN LEARNING (H), RESPECTIVELY

% successes	SP1	SP2	SP3
GP	96.1	96.2	97.8
GB	95.1	95.3	95.6
LP	79.3	88.4	94.0
LB	78.2	83.3	86.7
LV	79.1	86.0	92.2
LH	80.0	88.1	93.9

the STEP's 4 and 5 do not considerably increase the matching probability and the training patterns used in STEP's 1 to 3 are sufficient to estimate an appropriate PDF via the Parzen's window approach with our experiments.

B. Comparative Analysis

In this section we have analyzed, purely for comparative purposes, the percentage of successes for each method. These percentages are summarized in Table I for STEP's 1 to 3, as these STEP's are the most relevant from the results in Fig. 7.

As mentioned during the introduction, we select the global matching strategy in [6] and map the similarity constraint with the PDF estimated from the Parzen approach instead of the original PDF estimated by the Bayesian method of [5]. The results are given in Table I corresponding to GP and GB respectively, where G means global and P and B mean Parzen and Bayesian respectively.

We call local (L) matching strategies those which use only the similarity constraint. We have compared the nonparametric Parzen's window (LP) to the following local strategies: the parametric Bayesian (LB) [5], the learning vector quantization (LV) [7] and the Hebbian learning (LH) [8] respectively.

From the results of Table I, the following conclusions may be inferred.

- 1) With the PDF estimated via the non-parametric Parzen's window the results are better than with the PDF estimated via the parametric Bayesian approach, in both global and local methods.
- 2) As expected, global approaches provide better results than local ones. This is the consequence of using a global relaxation ap-

proach with more stereovision matching constraints than in the local methods.

- 3) The results of the approach proposed in this paper are very similar to those obtained by the Hebbian learning strategy and improve the results of learning vector quantization. Since these two methods compare favorably with other existing methods, we can conclude that the nonparametric Parzen's window estimator appears to be a valid method for local stereovision matching.

In order to further validate this approach, we took as a reference the Bayesian approach in [5]. We have improved the maximum Bayesian decision rule in [5] by considering the combination classifier defined in [24]. The results obtained are very similar to those obtained in [5]. Hence, Parzen's window still achieves better results than those obtained with this improved Bayesian decision approach.

IV. GENERALIZATION

In this paper, we have used edge segments as features and the similarity constraint has been mapped through a PDF estimated by a non-parametric approach. The proposed method can be generalized and applied as follows.

- 1) For any type of features, i.e., curved segments, regions, edge points, corners, etc. This is because our approach is based on the estimation of a PDF where the training and theoretical patterns are the attribute values for the edge segments. Hence, for any given feature the only problem is to compute its attributes, which is beyond the scope of this paper.
- 2) For a multiresolution scheme where the similarity matching constraint is used. Three types of multiresolution are commonly used:
 - 2.1) classical pyramidal structures using Laplacian of the Gaussian filters extract information at different resolution levels based on the standard deviation parameter [11]. The matching process at a given resolution level can be carried out through (6) and is driven by disparity values produced at a coarser level;
 - 2.2) multiscale decomposition [25] including wavelets [26]. At each level an activity-level measurement is required. In stereovision matching this is performed through a similarity measurement [26] which can be achieved by using the PDF estimated using (6);
 - 2.3) a two level scheme, where the first level matches intensity edges, i.e., a feature-based matching, which drives the disparity at a second level based on area-based matching [27]. As before, the similarity measurement can be obtained through (6).

V. CONCLUSIONS

The Parzen's window stereo-matching estimation approach improves results when it is used to measure the similarity constraint in both local and global strategies. Some of these local methods involve learning. An interesting conclusion is that the results obtained with a nonparametric PDF estimation are better than those obtained with a parametric PDF estimation.

The number of stereo images, features and experiments was adequate for our stereo-matching approach. This number of pairs of edge-segments was completely representative of the environment where our mobile robot, equipped with our stereovision system, navigated.

The performance of the Parzen approach improves as the number of training patterns increases. This behavior is affected neither by the nature of the different objects nor by the illumination conditions.

We have made a generalization of the proposed method and given guidelines for its extension and application to other matching strategies where the similarity constraint is used. It has also been generalized to other types of environments, such as outdoor scenes or aerial images, where edge segments are probably inappropriate features and therefore different features and attributes would be more suitable.

ACKNOWLEDGMENT

The authors would like to thank the reviewers and Associate Editor for constructive recommendations.

REFERENCES

- [1] S. H. Lee and J. J. Leou, "A dynamic programming approach to line segment matching in stereo vision," *Pattern Recognit.*, vol. 27, pp. 961–986, 1994.
- [2] D. Marr and T. Poggio, "A computational theory of human stereovision," *Proc. R. Soc. London B*, vol. 207, pp. 301–328, 1979.
- [3] G. Medioni and R. Nevatia, "Segment based stereo matching," *Comput. Vis., Graph., Image Process.*, vol. 31, pp. 2–18, 1985.
- [4] S. B. Pollard, J. E. W. Mayhew, and J. P. Frisby, "PMF: A stereo correspondence algorithm using a disparity gradient limit," *Perception*, vol. 14, pp. 449–470, 1981.
- [5] J. M. Cruz, G. Pajares, and J. Aranda, "A neural network approach to the stereovision correspondence problem by unsupervised learning," *Neural Networks*, vol. 8, no. 5, pp. 805–813, 1995.
- [6] G. Pajares, J. M. Cruz, and J. Aranda, "Relaxation by Hopfield network in stereo image matching," *Pattern Recognit.*, vol. 31, no. 5, pp. 561–574, 1998.
- [7] G. Pajares, J. M. Cruz, and J. A. López-Orozco, "Improving stereovision matching through supervised learning," *Pattern Anal. Applicat.*, vol. 1, pp. 105–120, 1998.
- [8] —, "Stereo matching using Hebbian learning," *IEEE Trans. Syst., Man, Cybern. B*, vol. 29, pp. 553–559, Aug. 1999.
- [9] G. Q. Wei, W. Brauer, and G. Hirzinger, "Intensity- and gradient-based stereo matching using hierarchical Gaussian basis functions," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, pp. 1143–1160, Nov. 1998.
- [10] E. Trucco and A. Verri, *Introductory Techniques for 3-D Computer Vision*. Upper Saddle River, NJ: Prentice-Hall, 1998.
- [11] A. Huertas and G. Medioni, "Detection of intensity changes with sub-pixel accuracy using Laplacian-Gaussian masks," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, no. 5, pp. 651–664, 1986.
- [12] J. G. Leu and H. L. Yau, "Detecting the dislocations in metal crystals from microscopic images," *Pattern Recognit.*, vol. 24, no. 1, pp. 41–56, 1991.
- [13] M. S. Lew, T. S. Huang, and K. Wong, "Learning and feature selection in stereo matching," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, pp. 869–881, Sept. 1994.
- [14] E. P. Krotkov, *Active Computer Vision by Cooperative Focus and Stereo*. Berlin, Germany: Springer-Verlag, 1989.
- [15] S. Tanaka and A. C. Kak, "A rule-based approach to binocular stereopsis," in *Analysis and Interpretation of Range Images*, R. C. Jain and A. K. Jain, Eds. Berlin, Germany: Springer-Verlag, 1990.
- [16] R. Nevatia and K. R. Babu, "Linear feature extraction and description," *Comput. Vis., Graph., Image Process.*, vol. 13, pp. 257–269, 1980.
- [17] R. L. Coultrip and R. H. Granger, "Sparse random networks with LTP learning rules approximate Bayes classifiers via Parzen's method," *Neural Networks*, vol. 7, no. 3, pp. 463–476, 1994.
- [18] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.
- [19] E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Statist.*, vol. 33, pp. 1065–1076, 1962.
- [20] A. K. Jain and M. D. Ramaswami, "Classifier design with Parzen windows," in *Pattern Recognition and Artificial Intelligence Toward an Integration*, E. S. Gelsema and L. N. Kanal, Eds. Amsterdam, The Netherlands: Elsevier, 1988, pp. 211–218.
- [21] K. Fukunaga and D. M. Hummels, "Bayes error estimation using Parzen and K-NN procedures," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, no. 5, pp. 634–643, 1987.
- [22] L. F. Escudero, *Reconocimiento de Patrones*. Madrid, Spain: Paraninfo, 1977.
- [23] R. P. W. Duin, "On the choice of smoothing parameters for Parzen estimators of probability density functions," *IEEE Trans. Comput.*, vol. C-25, pp. 1175–1179, 1976.
- [24] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, pp. 226–239, Mar. 1998.
- [25] Z. Zhang and R. S. Blum, "A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application," *Proc. IEEE*, vol. 87, pp. 1315–1325, Aug. 1999.
- [26] Y. S. Kim, J. J. Lee, and Y. H. Ha, "Stereo matching algorithm based on modified wavelet decomposition process," *Pattern Recognit.*, vol. 30, no. 6, pp. 929–952, 1997.
- [27] C. Baillard and O. Dissard, "A stereo matching algorithm for urban digital elevation models," *Photogramm. Eng. Remote Sens.*, vol. 66, no. 9, pp. 1119–1128, 2000.

Histogram-Based Fuzzy Filter for Image Restoration

Jung-Hua Wang, Wen-Jeng Liu, and Lian-Da Lin

Abstract—In this paper, we present a novel approach to the restoration of noise-corrupted image, which is particularly effective at removing highly impulsive noise while preserving image details. This is accomplished through a fuzzy smoothing filter constructed from a set of fuzzy membership functions for which the initial parameters are derived in accordance with input histogram. A principle of conservation in histogram potential is incorporated with input statistics to adjust the initial parameters so as to minimize the discrepancy between a reference intensity and the output of defuzzification process. Similar to median filters (MF), the proposed filter has the benefits that it is simple and it assumes no *a priori* knowledge of specific input image, yet it shows superior performance over conventional filters (including MF) for the full range of impulsive noise probability. Unlike in many neuro-fuzzy or fuzzy-neuro filters where random strategy is employed to choose initial membership functions for subsequent lengthy training, the proposed filter can achieve satisfactory performance without any training.

Index Terms—Fuzzy filter, histogram, image restoration, impulsive noise, median filter.

I. INTRODUCTION

Median filtering (MF) is a nonlinear technique that is well known for its effectiveness in removing impulsive noise. The 1-D MF is realized by passing a window over the input data, and taking the median value of the data inside the window as the output associated with the center of the window. In image processing applications two-dimensional median filters have been used with some success, and various methods [1]–[3] have been proposed to extend the median operation to two dimensions. The simplest way [3] is to pass a 2-D window, such as a square mask, over the 2-D input image. As with the 1-D MF, the pixels inside the window are ranked according to their gray intensity values and the median value taken as the output. By associating the nonlinear operation of the median filter with a linear cost function, Qiu [4] has shown that me-

Manuscript received April 5, 1998; revised October 29, 1999 and November 7, 2001. This work was supported by the National Science Council of Taiwan, R.O.C., under Grants NSC 86-2611-E-019-024 and NSC 89-2611-E-019-054. This paper was recommended by Associate Editors M. Kam and K. Pattipati.

J.-H. Wang and L.-D. Lin are with the Department of Electrical Engineering, National Taiwan Ocean University, Keelung, Taiwan, R.O.C. (e-mail: jhwang@celab1.ee.ntou.edu.tw).

W.-J. Liu is with the Department of Electronics Engineering, Nan Kai College, Nantou, Taiwan, R.O.C..

Publisher Item Identifier S 1083-4419(02)00705-7.