

# Multiple Correspondences in Stereo Vision Under a Genetic Algorithm Approach

JOSÉ DEMISIO SIMÕES DA SILVA<sup>1,2</sup>

PAULO OUVERA SIMONI<sup>2,3</sup>

KAMAL KANT BHARADWAJ<sup>4</sup>

<sup>1</sup>Computing and Applied Mathematics Laboratory (LAC), National Institute for Space Research (INPE)  
Av. dos Astronautas, 1758, Sao Jose dos Campos, 12227010, SP, Brazil  
demisio@lac.inpe.br

<sup>2</sup>Braz Cubas University (UBC) Av. Francisco Rodrigues Filho, 1233, Mogi das Cruzes, SP, Brazil

<sup>3</sup>Guarulhos University (UnG) Praça Tereza Cristina, 1, Guarulhos, SP, Brazil  
paulosimoni@ig.com.br

<sup>4</sup>Systems Sciences, JNU – University, New Delhi, India  
drkkb@jnuniv.ernet.in

**Abstract.** This paper presents a multiple point hierarchical approach to the stereo correspondence problem in computer vision. The low-level processing employs an area-and-token hybrid method to obtain, for distinctive points in one image, a set of points in the other image that are candidates for correspondence. The refinement of the set of low-level correspondences obtained is performed by a high-level N-point simultaneous correspondence process, a new constraint introduced for this problem. The high-level processing uses a genetic algorithm approach for searching the solution space. Experimental results show the effectiveness of the method on real world scenes.

## 1 Introduction

Stereopsis is the process that computes depth from two or more images of the same scene obtained by spatially separated cameras (Bertero et al. [3]). This process is a preprocessing phase for computer vision tasks such as, stereo vision (3D information) and motion analysis using 2D data (Anadan [1], Jones [9]). The main problem in stereopsis is the correspondence process that aims at establishing relations among images of the same scene in order to guarantee that the images belong to the same scene. The work presented in this paper is concerned with the recovery of 3D depth information using two images, which can be estimated by calculating the disparity between the image points, as it is shown in Figure 1. The depth information  $D$  for a point can be derived by the expression

$$D(x_L, x_R) = \frac{fb}{x_L + x_R} \quad (1)$$

where,  $x_L + x_R$  is the disparity between the corresponding left and right points;  $f$  is the focal distance, which is a camera characteristic; and  $b$  is the baseline, defined as the distance between the two optical axis resulting from the cameras arrangement (see Figure 1).

There has been a lot of research work on the correspondence problem, but no generalized approach has been achieved yet. Existing algorithms are based on different paradigms such as neural networks,

regularization, learning strategies, and optimization techniques. New research works have been stimulated by the availability of new agile stereo heads with a large number of degrees of freedom, permitting concentration on temporal integration of salient image information under the active vision paradigm (Jones [9]).

The different approaches to the correspondence problem are either area- or token-based, and use similarity metrics, such as cross-correlation, sum of squared differences, Euclidean distance, etc. to achieve the correspondence. They normally try to satisfy the constraints stated by Marr [10]: point compatibility, uniqueness of correspondence, and map continuity.

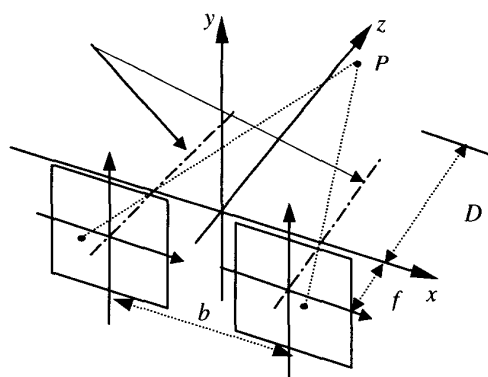


Figure 1 – The geometry of stereoscopy.

The solution to the correspondence problem requires in general the choice of a reference feature in one of the images, to be used in the correspondence process. Although it is not unusual to find several similar features in the other image, the constraints of the problem impose the existence of a unique correspondence.

Area-based approaches are sensitive to illumination changes, to occlusion (partial blocking of objects with different depths), and to the size and shape of the window used. Token-based approaches are less sensitive to occlusion, but require detection and description phases, which may be time consuming.

This paper proposes a point-wise, hybrid (area- and token-based) approach to the correspondence problem, using contextual (area) and structural (token) features of a point and its neighborhood. The similarity of candidate corresponding points to a reference point is computed using difference, correlation, and distance metrics.

In order to cope with the problem of choosing among a very large number of correspondences that can be found, it is also proposed in this paper that a new constraint should be added to the problem besides Marr's constraints: the simultaneous correspondence of many ( $N$ ) reference points. In these regards, polygonal regions must be matched, besides single points, as in traditional methods. Thus, the correspondence takes place if a set of candidate points (polygonal region) satisfies the constraints stated in Marr [10] and the new constraint (structural coherence constraint). The polygonal regions may not represent real objects in the scene.

The polygonal region geometric features may vary from one image to the other, due to occlusion and possible excessive disparity between the images, thus increasing the computational complexity of the problem. Moreover, each individual reference vertex (point) may have a large number of corresponding candidate points, which means there may be a very large number of possible polygonal regions that may correspond to the reference polygonal region. However, the bigger the number of simultaneous points, the more complex is the polygonal region. Token complexity for the correspondence problem has already been observed by Jones [9]. More complex geometric features can then considerably reduce the number of correspondence possibilities. Nevertheless, the number of corresponding candidate polygonal regions may still be very large, requiring a search over a huge space, for which a genetic algorithm is proposed.

GAs have already been used in some image processing, computer vision, and pattern recognition problems (Bala and Wechsler [2], Tsang [15], Singh et al. [14], Mirmehdi et al. [12]). For the stereo correspondence problem, Saito and Mori [13] have developed a method employing a GA to determine the disparity map

optimizing both the compatibility between corresponding points and the map continuity. Chai and Ma [5] have presented an evolutionary framework for stereo correspondence for non-calibrated images.

It is proposed in this paper to use genetic algorithms to the correspondence problem under a quite different and more complex approach, in which low-level and high-level processing are conducted.

## 2 The Proposed Model

### 2.1 Contextual and Structural Features for the low-level correspondence process

The considered contextual features of a point are the micro area (within a predefined window size), the macro area (within a window  $n$  times the micro area), and the gray level of the point. Such features provide local and global contextual information about the point and its neighbors and are computed for both images.

The structural features are defined as those to which the considered point belongs. The following 8 structural features have been proposed to be used: the vertical and horizontal lines; the principal and secondary diagonals; the bottom right, top right, top left, and bottom left corners, defined within a predefined window. These features are binary vectors obtained from differences among the gray level of a point and its neighbors, computed through Perceptron neural networks (Haykin [8]), with weight insertion for each different structure. As an example, the construction of the vertical line structure within a 3 x 3 window is shown in Figure 2. All the other binary features are extracted using similar networks. Bigger windows require redefinition of the neural networks for extracting such structures.

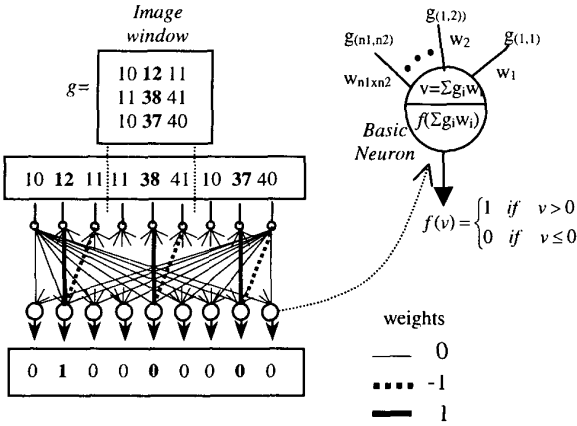
In addition to the 8 structural features mentioned before, two additional structures are also considered: the pattern of differences among the central pixel and its connected neighbors, and the predominant structure. For the example in Figure 2, the pattern of differences is computed as:

$$D = \text{sign}(12-38, 11-38, 41-38, 40-38, 37-38, 10-38, 11-38, 10-38) \\ = (0, 0, 1, 1, 0, 0, 0, 0)$$

where the *sign* function is applied to each array element.

The second additional structural feature is computed by applying the convolution operator to a group of basic morphological structures and to an image window. The mask with highest computed convolution value is called the predominant structure. Such process can be thought of as an energy analysis given by the convolution. For the

example in Figure 2 the predominant structure is the highlighted one in Figure 3. Following Jones [9], the bigger the window size and the larger the number of structural features, the more complex the binary structures are, thus increasing the likelihood of the uniqueness constraint to be satisfied.



**Figure 2** - The vertical line structure in a 3 x 3 window. Weights are inserted as shown.

0 0 0	0 1 0	0 1 0	0 0 0	0 0 0	0 1 0	1 0 0	0 0 1	1 0 0	0 0 1
1 1 1	0 1 0	0 1 1	0 1 1	1 1 0	1 1 0	0 1 0	0 1 0	0 1 0	0 1 0
0 0 0	0 1 0	0 0 0	0 1 0	0 1 0	0 0 0	0 0 1	1 0 0	1 0 0	0 0 1
1 0 1	0 0 0	0 1 0	0 1 0	0 0 1	1 0 0	0 0 0	0 0 1	0 0 0	1 0 0
0 1 0	0 1 0	0 1 0	0 1 0	0 1 0	0 1 0	1 1 0	1 1 0	0 1 1	0 1 1
0 0 0	1 0 1	1 0 0	0 0 1	0 1 0	0 1 0	0 0 1	0 0 0	1 0 0	0 0 0

**Figure 3** - Basic binary morphological structures within a 3 x 3 window.

## 2.2 Metrics for similarity

Considering the constraint of map continuity, that is, disparity smoothness over the image, it is assumed that the correspondence will take place if both the reference point and a candidate point lie within similar macro and micro contexts and belong to similar structures. For that purpose, point correspondences are performed under 14 similarity criteria:

- Non-normalized Euclidean distance between the micro areas ( $ED$ ).
- Correlation index between the micro ( $C_{micro}$ ) areas.
- Correlation index between the macro ( $C_{macro}$ ) areas.
- Hamming distance between the 10 different binary structures ( $H_1 H_2 H_3 H_4 H_5 H_6 H_7 H_8 H_9 H_{10}$ ).
- Absolute difference between the points' gray levels ( $G$ ).

The non-normalized Euclidean distance between the micro areas captures illumination differences from one image to the other, within the micro context.

Following the feature extraction phase of the process, each candidate point  $i$  is assigned a vector of 14 similarity measurements for the different matching criteria, together with the point image coordinates ( $l_i, p_i$ ).

$$Q_i = [l_i p_i ED C_{macro} H_1 H_2 H_3 H_4 H_5 H_6 H_7 H_8 H_9 H_{10} G C_{micro}]$$

A point  $i$  with a feature vector  $Q_i$  will be chosen as the corresponding one to a certain reference point, if its similarity metrics show the maximum match possible.

## 2.3 Geometric Features for the high-level correspondence

Three geometric features are employed to cope with the structural coherence constraint between the polygonal regions ( $N$ -point simultaneous correspondence):

- Sum of the angles between each side of the polygonal region and a horizontal line, in the image plane;
- Area of the polygonal region;
- Sum of the Euclidean distances of the incenter of the polygonal region to its vertices.

The geometric features are computed for both the reference and candidate polygonal regions and they are compared through absolute differences.

The use of contextual and structural features of each point (vertex), as well as the geometric properties of the polygonal regions characterizes a hierarchical approach to the correspondence problem, in which point correspondence is conducted at a low level by analyzing local properties of a point, whereas high level correspondence occurs globally across the image, while matching the polygonal regions. Thus, for a reference polygonal region, the correspondence is the process of finding a polygonal region in the other image, among a large number of candidate polygonal regions (or search space), that may match the reference. Although the geometric features are not invariant under projection, they are used as a heuristic metric for the similarity between the geometric shapes of the polygons, under different views.

## 2.4 Genetic Algorithms (GAs)

GAs were introduced by John Holland in 1975, based on a method for studying natural adaptive systems and designing artificial adaptive systems, with roots in

Darwinian natural selection and Mendelian genetics. GAs search a problem representation space of artificial adaptive systems, eliminating weak elements by favoring retention of optimal and near optimal individuals (survival of the fittest), and recombining features of good individuals to perhaps make better individuals.

The elements of the search space represent possible solutions to the problem and are coded as strings (chromosomes) derived from an alphabet (Goldberg [7]). They work with a coding of the problem rather than the problem itself. Such characteristic has made them to be known as robust optimization methods (Mirmehdi et al. [12], Michalewicz [11]).

In order to use a GA approach, it is necessary to first derive an initial population of chromosomes (potential solutions) and set a cost function to measure each chromosome's fitness. The best solution is then searched in the solution space. The optimization is performed by manipulating the population of chromosomes, during a number of generations, in each of which the GA creates a set of new individual by crossover and mutation operations, similar to natural reproduction processes. The crossover operation takes two parent chromosomes and mates them to produce two child chromosomes. The mutation operation is used for exploring new regions in the search space, thus maintaining population diversity.

GAs are particularly suitable for applications that require adaptive problem-solving strategies.

Following is the structure of a simple Genetic Algorithm:

```

Procedure Genetic_Algorithm:
begin
   $t=0$  ; initialize  $P(t)$ ;
  evaluate  $P(t)$ ;
  while (NOT termination-condition) do
    begin
       $t=t+1$ ;
      select  $P(t)$  from  $P(t-1)$ ;
      recombine  $P(t)$ ;
      evaluate  $P(t)$ ;
    end;
  end;

```

## 2.5 The proposed GA approach to the correspondence problem

Consider  $P_1, P_2, \dots, P_b, \dots, P_N$  as  $N$  vertices of a reference polygonal region in the left image ( $L$ ), and that for each vertex  $P_b$ ,  $S_b = \{Q_b^j : j = 1, 2, \dots, k_b\}$  is the set of candidate corresponding points to  $P_b$  in the right image ( $R$ ).  $k_b = |S_b|$

is the number of elements in the set  $S_b$ . Each set,  $S_b$ , is constructed under a rule tuned to avoid irrelevant points, by analyzing the correlation index between the micro areas of the candidate point and its reference, and the correlation index of their macro areas. Both indices are compared to a threshold  $T$  ( $T=0.75$  in the experiments).

**IF**  $C_{micro} \geq T$  &  $C_{macro} \geq T$  **THEN** choose point

A solution (a genotype) to the problem is coded as an ordered sequence of randomly chosen vertices, forming a candidate polygonal region,

$$Q_1 Q_2 \dots Q_i Q_{i+1} Q_{i+2} \dots Q_N \quad \text{where } Q_i \in S_i$$

The genotype is a string in which each element is the aggregation of 16 floating point numbers representing pixel coordinates and the 14 similarity metrics (section 2.2).

The use of  $N$  simultaneous points implies the existence of  $N$  sets of candidate points that may result in a large number of possible solutions (a search space with  $k_1 \times k_2 \times \dots \times k_N$  possible candidate polygonal regions may result in the right image). If 4 vertices are considered, each with 1000 candidate points, then there may be  $10^{12}$  solutions to the problem. The proposed method avoids dealing with a huge population, acting on a randomly chosen sub set (population) of candidate points. Diversity in the working population (new candidate points) is brought in by the mutation operation. Then, the initial population is a pool of ordered sequences representing random candidate polygonal regions as potential solutions to the correspondence problem.

### Crossover

The crossover operator is used to exchange genetic material among highly fitted individuals in the population. In the proposed model, a uniform crossover operator exchanges genes at randomly chosen positions across the parent chromosomes. The operation consists in the generation of a binary template (bit string) of the same length as the parents, and the offspring chromosomes are generated by swapping the genes between the parents at the positions corresponding to 0's in the binary template, while those at positions 1's remain unaltered. In the proposed approach, the crossover operator is designed to fit the needs of the non-binary representation used. Uniform crossover is illustrated in Figure 4-c. In addition to the uniform crossover the usual one point and two point crossovers (Figures 4-a and 4-b) are also used.



the current population as a basis for the next generation, and guarantees that the best answer found throughout the generations is present in the last one.

### 3 Results

The proposed GA approach was applied to pairs of real world images with 120 x 160 pixels and 240 x 320 pixels, each one with 64 gray levels. Such images were acquired from indoor scenes, with an assembly of two parallel cameras under the same illumination conditions.

The experiments aimed at using the proposed approach to establish the correspondence of  $N=6$  simultaneous points. Besides the previously referred images, the method was also tested on several different images.

The general process starts by the choice of the 6 reference points in the left image ( $L$ ) (Figure 6), based on gradient information within a 5 x 5 window and a contrast analysis of regions of the image. For reference point, several candidate corresponding points are then chosen by applying the rule of section 2.5.

The contextual and structural features are extracted for the reference and candidate points. The similarities among the features are computed by the similarity criteria mentioned in section 2.2. Such metrics are computed for each candidate point in relation to its reference point. Each candidate point is assigned a vector of similarity metrics.

The initial population is created with  $Npop$  ( $Npop=100$ , in the experiments) individuals, for each of which the geometric features, described in section 2.3, are computed. Similarity metrics are applied to compute the differences relative to the reference polygonal region.

In the experiments, the real corresponding points were picked from the right image. Such information was used as a reference to allow a numerical analysis of the answer provided by the proposed approach (see table 1).

The genetic algorithm was executed to evolve a population using 85% crossover probability and 50% mutation probability. These probabilities resulted from experiments developed to test the influence of different crossover and mutation rates. Higher mutations could have been used to introduce more diversity in the process.

Table 1 presents the coordinates of the reference and corresponding points, as well as, the coordinates of the correct correspondence in the right image (Figure 6). The computed squared errors show the effectiveness of the proposed approach, which can be seen by the mean squared error between the set of references and the set of corresponding points found (less than 1 squared pixel).

Table 1 - Computed squared errors for the correspondences in Figure 6.

	Reference Points		Corresp. Points found		Real Corresp. Points		Squared Errors	
	Line	Pixel	Line	Pixel	Line	Pixel	Line	Pixel
1	33	109	32	94	32	94	0	0
2	45	101	47	69	47	69	0	0
3	40	100	43	70	43	70	0	0
4	18	98	18	83	18	82	0	1
5	19	107	19	92	19	92	0	0
6	25	108	25	93	25	93	0	0
Mean squared error:							0	0.166

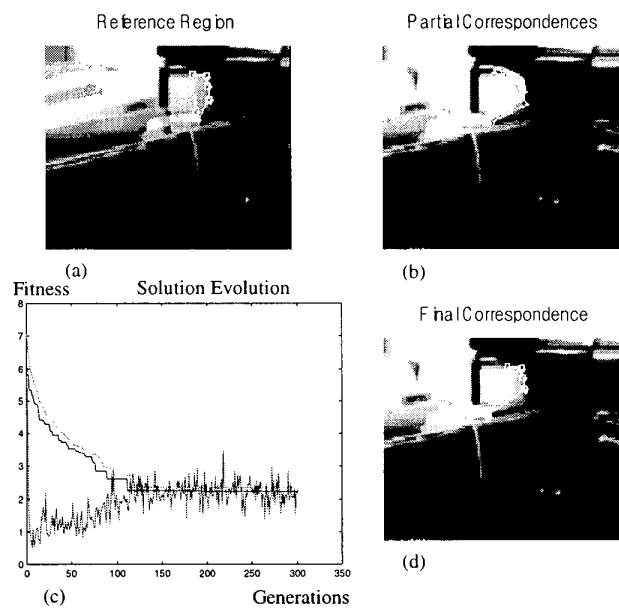
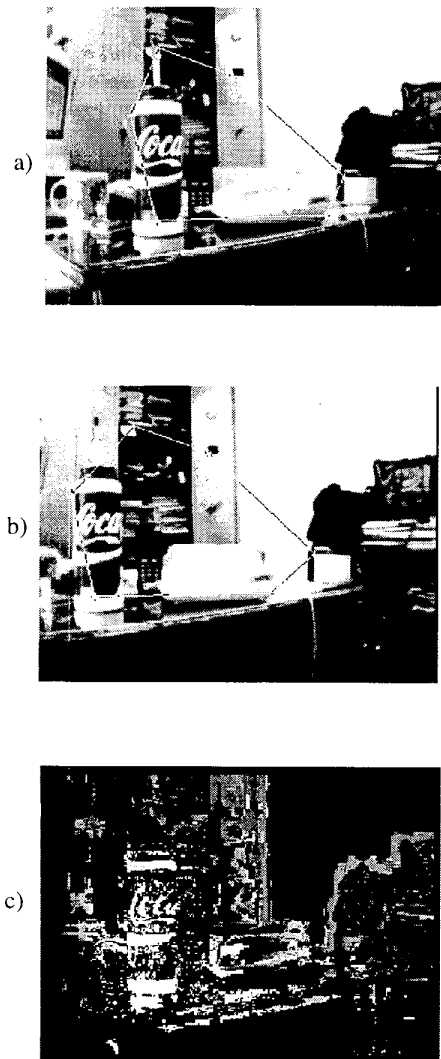


Figure 6 - (a) reference polygonal region with vertices at (33,109), (45,101), (40,100), (18,98), (19,107), (25,108); (b) partial solutions; (c) solution evolution; (d) final corresponding polygonal region with vertices at (32,94), (47,69), (48,70), (18,83), (19,92), (25,93).

Figure 7 shows the results for another image. Figure 7-a) displays the image of an office room in which 6 reference points were chosen through a global analysis of predominant structural features. The proposed method was applied using the same parameters mentioned before for the genetic algorithm. The corresponding points are displayed in Figure 7-b). Figure 7-c) shows the disparity map obtained by applying equation (1). It is to be noted

that the closer a point is to the camera the larger will the disparity (bright pixels) be.



**Figure 7** – a) Reference Image with 6 control points; b) Corresponding points; c) Disparity map generated from the 6 control points.

The disparity map was constructed based only on the correspondence of the 6 reference points. An algorithm was developed to compute the disparity map considering the 6 correspondences found as control points, thus avoiding exhaustive calculations to establish many correspondences before computing the disparity map. The

computations performed verify the correspondence between a point in the reference image, and a point in the other image, by minimizing the sum of differences among the similarity measures of their structural and contextual features, and the differences among their spatial localization in the images.

#### 4 Conclusions

This paper presents an approach based on genetic algorithms to multiple-point correspondence of a pair of stereo images.

The process is modeled under a point-wise fashion, using point contextual and structural features, and the geometric features of the polygonal region defined by these points.

Structural features extraction is performed by a group of neural networks, thus not requiring preprocessing for gradient extraction and token establishment prior to the correspondence process. The structures are restricted to local operators in the neighborhood of interesting points (see section 3), picked up as those belonging to edges in the image.

The structural coherence of the polygonal regions allows the simultaneous correspondence of multiple points, thus accelerating the correspondence process. The combination of multiple points with contextual and structural features makes the proposed approach different from those in Saito and Mori [13] and Chai and Ma [5].

The results presented show the correspondence of control points in the image, before recovering a complete disparity map. The number of multiple points can be large, thus implying more complex geometric features of the polygonal region and a larger number of correspondences established simultaneously. However, an increase in the number of simultaneous points means an increase in the processing time.

The effectiveness of the use of a genetic search in pursuing the best correspondence of multiple reference points within a huge search space, has been shown with the experiments conducted.

An algorithm was developed to calculate the disparity map shown in Figure 7-c). The result may be enhanced if the number of simultaneous points is larger, because more control points can be used to guide the calculations.

The occlusion problem is implicitly treated here by performing a number of iterations in the correspondence process. Along the iterations the reference image is alternated. The idea consists in fixing a set of initial reference points in one image (say the left one), pursuing their corresponding ones in the other image (the right image). The correspondences found become the references

for the next iteration whose corresponding points are pursued back in the left image, and so on. The algorithm stops when convergence to a set of repeated points in both images is reached.

boundaries, *Pattern Recognition Letters* 18 (1997), pp. 631-639.

## 5 References

- [1] Anadan, P.A., Review of Motion and Stereopsis Research. COINS Technical Report 85-52, University of Massachusetts at Amherst, December 1985.
- [2] Bala, J., H. Wechsler, Shape analysis using genetic algorithm, *Pattern Recognition Letters* 14(12), 1993, pp. 965-973.
- [3] Bertero, M., T. A. Poggio, T.A., V. Torre, III-Posed Problems in Early Vision. Proceedings of the IEEE 76(8), August 1988, pp. 869-889.
- [4] Brooks, R. R., S.S. Iyengar, and J. Chen, Automatic correlation and calibration of noisy sensor readings using elite genetic algorithms, *Artificial Intelligence*, 84, 1996, pp. 339-354.
- [5] Chai, J., S. D. Ma, Evolutionary Framework Stereo Correspondence. Int. Conf. Pattern Recognition, 1998, pp. 841-844.
- [6] Gen, M., Y. Tsujimura, and E. Kubota, Solving Job-Shop Scheduling Problems by Genetic Algorithm. IEEE Conf. System, Man and Cybernetics, 1994. pp. 1580-1579.
- [7] Goldberg, D. E. Genetic algorithms in Search, Optimization and Machine Learning, Reading, Addison Wesley, 1989.
- [8] Haykin, S. Neural Networks: A comprehensive foundation. Macmillan, New York, 1994.
- [9] Jones, G. A., Constraint, Optimization, and Hierarchy: Reviewing Stereoscopic Correspondence of Complex Features. *Computer Vision and Image Understanding* 65(1), 1997, pp. 57-78.
- [10] Marr, D., Vision. Freeman. San Francisco, CA. 1982.
- [11] Michalewicz, Z. Genetic Algorithms + Data structures = Evolution Program, Springer-verlag, New York, 1996.
- [12] Mirmehdi, M., P. L. Palmer, and J. Kittler, Genetic optimization of the image feature extraction process, *Pattern Recognition Letters* 18, 1997, pp. 355-365.
- [13] Saito, H. and M. Mori, Application of genetic algorithms to stereo matching of images, *Pattern Recognition Letters* 16, 1995, pp. 815-821.
- [14] Singh, M., A. Chatterjee, and S. Chaudhary, Matching structural shape descriptions using genetic algorithms, *Pattern Recognition* 30 (9), (1997), pp. 1451-1462.
- [15] Tsang, P. W. M., A genetic algorithm for affine invariant recognition of object shapes from broken