A Corresponding Points Classification Approach by Delaunay Triangulation

Elizabeth Vargas*, Maria Trujillo†

*Multimedia and Vision Lab, Universidad del Valle, Cali, Colombia, elizabeth.vargas@correounivalle.edu.co
† Multimedia and Vision Lab, Universidad del Valle, Cali, Colombia, maria.trujillo@correounivalle.edu.co

Keywords: Stereo vision; correspondence problem; Delaunay triangulation.

Abstract

Stereo vision has become in one of the most important research topics, given the wide field of applications. The challenge of stereo vision is reconstruct a 3D scene, by estimating corresponding points in two projections of the same points in the scene. The correspondence problem consists in finding for each point in the reference image, the corresponding point in the target image by measuring the similarity between points. However, a major problem is mismatching, due to phenomena such as ambiguity, occlusion and untextured regions. Misclassified corresponding points may badly impact the 3D reconstruction. In this paper, an approach for classifying corresponding points based on the Delaunay triangulation is proposed. The proposed classification consists in deciding whether or not a matching is reliable.Corresponding points are mapped into two graphs, points belonging to the right image into one graph and points belonging into the left image into the other. Thus, corresponding points are classified as "correctly estimated" if and only if their graphs are isomorphic.

1. Introduction

Stereo Vision refers to the ability to infer information on the 3D structure, such as depth of a scene, using two or more images taken from different viewpoints. The correspondence problem has to be solved from a stereo system in order to infer the 3D structure of a scene.
The correspondence problem consists in determining which point on the left image corresponds to which point on the right one [1]. Different techniques have been employed in order to find corresponding points given two images [6,7,8]. When the correspondence problem is addressed, corresponding points may be inaccurately estimated: falsely matched or badly located. Moreover, small inaccuracies in estimated corresponding points may have a large impact on the 3D reconstruction [1].
Around this problem, it has been developed a large number of applications in different sectors of the industry. In the aerospace business, for instance, it is known the use of exploring robots that are able to move (or navigate) autonomously on unknown surfaces [2]. Additionally, companies like Google are investing millions of dollars in researching on self-driving cars [3]. In the healthcare sector, image registration is everyday more frequent in the diagnose stage [4].
All these applications require finding reliable correspondences from stereo images. However, the stereo correspondence problem is not still solved. In fact, it is an inverse and ill-posed problem. For this reason, it is still open to proposals that improve the quality of the estimated correspondences by eliminating inaccuracies.
This paper presents an approach based on the Delaunay triangulation for classifying corresponding points according to the following criteria: a matching is considered correct if a set of corresponding points is mapped into an isomorphic graph.
The rest of the paper is organized as follows. Section II presents a summary of the previous work on stereo vision and Delaunay triangulation applied to this field. Section III introduces the proposed approach. Section IV shows the experimental evaluation. Section V includes final comments and conclusions.

2. Related Works

Several methods have been developed to find point correspondences. These methods have been classified based on different criteria, according to [5].

2.1 Type of Produced/Computed/EstimatedDisparity Map

Disparity is the shift between corresponding points. Dense disparity maps are the most required, because disparity values are determined for all or almost all pixels. On the contrary, sparse disparity maps have disparity values determined just for selected image points. In most situations, sparse disparity maps are calculated faster, but they have limited applications since there are missing values that, in several cases, have to be interpolated.

2.2 Format of the Taken Signal

There are methods where the taken signal comes directly from the intensity values, whereas others transform intensity into other domains or compute some characteristic features which are then used for matching [5].

2.3 Local and Global Methods

The first ones compute disparity values based only on the local information around certain position of pixels, while global methods use all cost values in the optimization process to determine disparity and occlusions.
One matching methods is Belief Propagation (BP). An implementation of BP is presented in [6], where stereo correspondence problem is formulated as a Markov network and it is solved using Bayesian belief propagation. Graph Cut is also a representative technique of global methods [5]; methods from this group formulate a solution to the stereo problem as energy functional implying a computation of a maximum flow in graphs. An application of this method can be found in [7]. Another approach is the use of Dynamic Programming, where the main idea lies in the division of the 2D search problem into a series of separate 1D search problems on each pair of epipolar lines, as illustrated in [8].
Delaunay triangulation has been widely used as a tool for feature correspondence, region corresponding and block matching. One of the most influential articles for this work is “Estimation of large-amplitude motion and disparity fields: Application to intermediate view reconstruction" [9]. In the paper, it is described a method for establishing dense correspondence maps between two images in a video sequence or in a stereo pair by combining feature matching with Delaunay triangulation. This approach has several advantages: it allows estimating large displacements and subsequently taking into account motion/disparity discontinuities. It is also, computationally efficient in regard to a typical multi-resolution block matching. The method produces high-quality intermediate views for natural (complex) stereoscopic image sequences. However, the results depend on the presence of texture in the images; the method works well in sequences with strong textures.
In [10] it is presented an improved SUSAN (Smallest Univalue Segment Assimilating Nucleus) corner detection algorithm and also an effective corners correspondence algorithm based on Delaunay Triangulation. The procedure is divided in three main stages: In the first stage, corners are extracted using an improved SUSAN detector, where an adaptive threshold strategy is adopted and the search space is limited. In the second stage, Delaunay triangulations are constructed among corners and triangle pairs with the similarity score above a threshold are obtained. At last, every edge of the matched triangle pairs is extended circularly until all the other matching corners are triangulated and mismatching corners are discarded. The proposed algorithm is robust and effective with scaling of images and rotation and translation with short baseline, given that the interior angles and local structural similarity are robust to noises and motion transform of images.
Another work based on Delaunay triangulation is presented in [11], where the image pairs are triangulated, and then triangles are classified into matched and unmatched triangles using a matching criterion: matched triangle and unmatched one, and the disparityvalues in two types of triangle region are solved respectively, finally the dense disparity map of the whole image can be obtained by integrating the disparity in all matched and unmatched triangle region. After that, disparities of all pixels within matched triangles, similar both in size and shape, are solved directly. For unmatched triangles, corner’s eigenvalue is introduced as a constraint to divide unmatched triangles into non-occluded, semi-occluded and occluded triangle according to the number of matching points, and disparities of pixels within these triangles are solved respectively. To conclude, a dense disparity map of the image is obtained by integrating the disparity in matched and unmatched triangle region. Standard images were used to evaluate the performance of the algorithm, showing that the average runtime is about 5/9 of a global matching algorithm and the average matching accuracy is improved by 4.03% comparing with local matching algorithm.
A generative probabilistic model to binocular stereo for fast matching is proposed in [12]. A Bayesian approach that uses a 2D mesh via Delaunay triangulation for stereo matching is presented, which is able to compute accurate disparity maps of high resolution images at frame rates close to real time without requiring global optimization. The result is an efficient algorithm that reduces the search space and it can be parallelised.
To conclude, in [13] it is presented a method for stabilising the computation of stereo correspondences using Delaunay triangulation. The input images are partitioned into small and localised regions. Adjacent triangles are merged into larger polygonal patches first and then, the planarity assumption was verified, in order to achieve the planarity test robustly. Point correspondences were established by planar homographies, and are used as control points in the final matching process using dynamic programming. The approach was tested on three different types of data, including medical data. The results showed that it can speed up the stereo matching process significantly.

3. Proposed Corresponding Points Classification

The approach proposed in this paper could be classified as a sparse, indirect and local method, according to the work of [5]. The approach consists of three main components: computation of a set of feature points, matching the initial collection of feature points and classification of correspondences based on the Delaunay triangulation. These components are described below.

3.1 Feature Points

The proposed approach uses feature matching. An initial set of features can be obtained using an algorithm from the state of the art. In particular, the SIFT (Scale Invariant Feature Transform) and FAST (Features from Accelerated Segment Test) corners are considered in this approach [14, 16].
The SIFT features are invariant to image scaling and rotation, and partially invariant to change in illumination and 3D camera viewpoint. They are well localized in both the spatial and the frequency domains, reducing the probability of disruption by occlusion, clutter, or noise [16].

FAST is a high quality corner detector implemented using machine learning. It is several orders faster than other existing corner detectors. The FAST detector also provides high levels of repeatability under large aspect changes and for different kind of features [14].

3.2 Initial Correspondences

The matching SIFT keypoints algorithm is provided in [15]. According to David Lowe in [16], the matching criteria for the best candidate match for each keypoint is found by identifying its nearest neighbour in the database of keypoints from training image. The nearest neighbour is defined as the keypoint with the minimum Euclidean distance for the invariant descriptor vector. The keypoint descriptors are highly distinctive, allowing a single feature to find its correct match with good probability in a large database of features. Nevertheless, many features from an image will not have any correct match in the reference image, because they arise from background clutter or were not detected in the reference image. The FAST corners are matched using SDD (Sum of squared differences) of the corners descriptor.
Fig. 1 illustrates the keypoints generated with SIFT and FAST algorithms for the right Teddy stereo image.

(a)SIFT (b) FAST

Figure 1. Illustration of keypoints calculated using (a) SIFT (b)FAST.

3.3 Corresponding Ponts Classification

The obtained correspondences in the matching stage are classified using the Delaunay triangulation. It is assumed that given two corresponding points, they are correctly estimated (i.e. they represent the same feature point in both images) if and only if they are mapped into an isomorphic graph the same triangle in both stereo images, implying that they have the same adjacent vertices in both images. Two vertices are considered the same, if they belong to the initial correspondence points set. The matches that do not satisfy this condition are labelled as "incorrectly estimated". The evaluation condition could be illustrated as:

If right.vertex *equal* left.vertex **then**
    mark as correct estimated
else
    mark as incorrect estimated

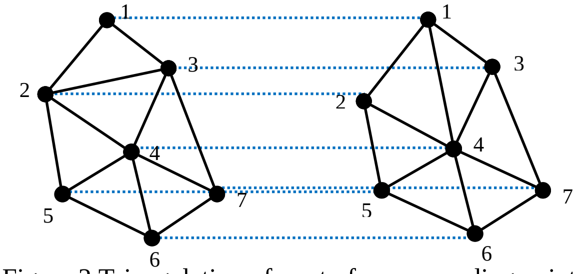

Figure 2.Triangulation of a set of corresponding points.

For the sake of completeness, the proposed approach is introduced using a set of calculated matching points. Fig.2 illustrates the graphs from left and right triangulations respectively. Vertices are labelled with the numbers 1 to 7 and the one with the same number represents a matching, which is also identified with a dashed line.
Initially, the vertex labelled as number 5 is considered and its adjacent vertexes 2, 4, and 6; as they are shown in Fig. 3 using green colour. According to the evaluation criteria, the matching point identified as numbers 5 is labelled as "correctly estimated ", given that the vertexes in both graphs have exactly the same adjacent vertexes on the left and right triangulations.

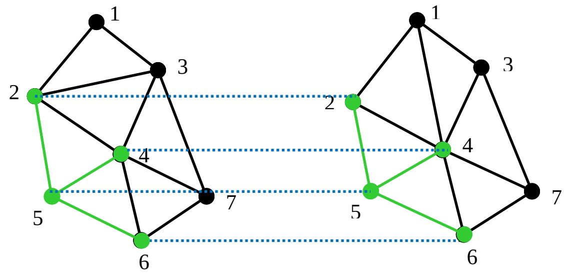

Figure 3.Proposed classification criteria illustrated using the vertex number 5.

Now, the vertex number 1 is considered and its adjacent vertices 2 and 3; as they are shown in Fig. 4 using green and red colours. The triangulation of the points from the right stereo image, the vertex 1 has three adjacent vertices: 2, 3 and 4. Even though, there are two common vertices (2 and 3), the matching point labelled as vertex 1 is classified as “incorrectly estimated” because the adjacent vertices are not exactly the same on both set of points, therefore they do not shape the same triangles.

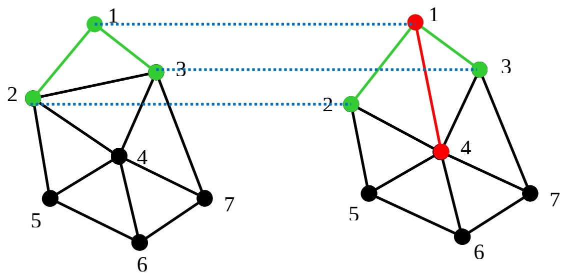

Figure 4. “Incorrectly estimated” classification criteria using the vertex number 1.

The last case is illustrated in Fig. 5 using another set of matching points. In this situation, the vertex number 7 is analysed and its adjacent vertexes are 2, 3, 4 and 6. In the right triangulation, the adjacent vertexes are 2, 6 and 7. However, in the left triangulation the vertex number 3 appears as an adjacent vertex instead of the vertex number 2. Therefore, the matching between the points identified as number 7 is labelled as "incorrectly estimated".

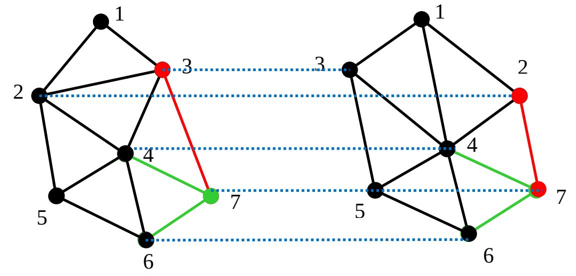

Figure 5.Classification criteria for the vertex number 7.

Formalising, given a set of corresponding points, they are mapped into an undirected graph, as it is illustrated in Fig. 6, and corresponding points are classified as "correctly estimated" if and only if their graphs are isomorphic. A graph isomorphism is defined as follows:

Let G and G' be a set of corresponding points, from the right and the left images respectively. A Delaunay triangulation produces a set of vertices and edges (V, A) and (V',A') respectively. A bijective function ϕ: V → V'is a graph isomorphism if:

$$w,v,z \in A \leftrightarrow \phi(v),\phi(w),\phi(z) \in A' \qquad (1)$$

That is, if ϕ preserves the adjacency between vertices [17]. In the proposed approach, the bijective function ϕ is represented by the initial map of correspondences and the set of adjacent vertices might have cardinality larger than 3.

4. Experimental Evaluation

It is necessary to compare the obtained results with a "gold standard" in order to evaluate its accuracy, therefore it is used the ground-truth of the image for this purpose. In combination with the stereo images, each dataset includes the true disparity map.The datasets used to test the propose approach were taken from Middlebury Stereo Datasets [18]. In total, 8 of them were used.The metrics employed to evaluate the algorithm performance are Sensitivity, Specificity, Positive Predictive Value (PPV) and Negative Predictive Value (NPV).

For validating an identified correct or incorrect corresponding point, the Bad Matching Pixel is used in order to calculate a confusion matrix. For each correspondence calculated in the initial matching, the value is compared with the ground-truth using the absolute error equation. The error at the *(i,j)* position is calculated using the following formula:

Figure 6. Delaunay Triangulation of a set of corresponding points using the Teddy stereo par.

$$error_{i,j} = \left| GT_{i,j} - \Delta x_{i,j} \right| \qquad (2)$$

$GT_{i,j}$represents the ground-truth disparity value at (*i*,*j*) and $\Delta x_{i,j}$is the estimated disparity value at (*i*,*j*). With an error value bigger than 1, the correspondence is labelled as "bad-match", otherwise is considered a "good- match". In this way, the confusion matrix is built using the "bad-match" and the "good match" as ground-truth classification.

SIFT keypoints were extracted with subpixel precision using Lowe's implementation provided in [15]. The algorithm generates large numbers of features that densely cover the image over the full range of scales and locations. A typical image of size 500x500 pixels will give rise to about 2000 stable features (however this number depends on both image content and choices for various parameters). Obtained corresponding points were classified as "correct" or "incorrect" corresponding point and compared with the ground-truth classification in order to calculate the confusion matrix in Table 1. The most remarkable result is the sensitivity of the dataset art because it is the highest one, meaning that most "bad-match" are recognized as "incorrect corresponding point". However, the score is just 80%, which is still low for the algorithm performance. In general, the unconstrained triangulation presents, in this set of matching, a poor result, because the values of sensitivity are not higher than 80%, which means, that there are many "bad matches" that are not considered "incorrect corresponding point". On the other hand, the specificity value, even though it is not high, presents an acceptable value, given that the existence of "good matches" that are consider "incorrect corresponding point" is not critical in this case.

Dataset	Sensitivity (%)	Specificity (%)	PPV(%)	NPV(%)
Art	81,82	51,16	24,32	93,62
Books	68,63	70,86	25,55	93,94
Cones	50,82	63,30	17,22	89,55
Dolls	62,61	54,26	26,09	84,91
Laundry	70,53	52,38	45,58	75,86
Moebius	51,79	54,96	19,73	84,21
Reindeer	64,86	41,55	16,55	86,87
Teddy	57,63	55,02	21,94	85,55

Table 1: Confusion Matrix calculaed using the SIFT algorithm.

The proposed approach tends to present a huge number of false positives. The reason is that one pair labelled as an "incorrect corresponding point" makes that the pairs around them (the adjacent vertices in the triangulation graph) are also labelled the same because of the missing edges. It means that there is an error propagation, which decreases the specificity. In order to improve this, it was proposed another evaluation method for both algorithms, which is more flexible.

A, the proposed evaluation was that, being n the cardinality of the set of adjacent vertices, evaluate that the target corresponding point has $(n - 1)$ equal vertices, if and only if $n > 3$. However, this evaluation decreased the sensitivity value for the Teddy dataset, where it was initially tested. Therefore, the constraint on the cardinality of the set of adjacent vertices, value of n, was relaxed until the sensitivity and specificity values were acceptable, it means the sensitivity was not lower than with the original condition. The final relaxed condition was established as:

```
if left.numberOfVertex > 3 then
        if right.vertex equal (left.vertex – 1) or
        right.vertex equalleft.vertex then
                mark ascorrect estimated
        else
                mark asincorrect estimated
else
        ifright.vertex equalleft.vertex then
                mark ascorrect estimated
        else
```

This was tested for all the datasets. These results are presented on the Table 2. In general, the specificity value increased, but the sensitivity value decreased considerably for some datasets like art.

Dataset	Sensitivity (%)	Specificity (%)	PPV(%)	NPV(%)
Art	63,64	55,81	21,65	88,89
Books	58,82	76,29	26,55	92,71
Cones	45,90	70,69	19,05	89,69
Dolls	57,39	58,74	26,40	84,24
Laundry	63,16	57,14	45,45	73,28
Moebius	50,00	60,31	21,21	84,95
Reindeer	59,46	47,83	16,92	86,84
Teddy	52,54	62,08	23,31	85,64

Table 2: Confusion Matrix using the SIFT algorithm with Relaxed Condition.

In order to detect corners, it was used the Computer Vision System Toolbox provided by MATLAB. The function *vision.CornerDetector* provides the performance for this task, allowing the change in some parameters. In our particular situation, the detection was implemented by using a method of local intensity comparison [14].

First of all, the feature representation from the corners is extracted by using the *extractFeatures* method. The resulting extracted points from both images are matched by using the function *matchFeatures*. The metric specified for this function was SDD (Sum of squared differences) and the match threshold was 0.2. Additionally, given that the input images are rectified, it was checked the condition that the correspondence points have the same x-coordinate. The number of matches obtained for each stereo par is variable, but it is around 150 and 300.

The results after the application of the triangulation to the points generated with the corners algorithm are presented on the Table 3. The values of sensitivity are higher than the ones for the SIFT points. However, the specificity is at the same time lower than with the other set of keypoints. For the datasets Books and Laundry is especially high the sensitivity value: 93,29% and 97,42% respectively. It could be inferred that the proposed approach has a better performance, when the initial set of correspondences has around 50% of "bad matches".

The new evaluation criterion, already explained, was also tested using the FAST corners. For this kind of features the results were better than for the SIFT feature points. For the half of the datasets the sensitivity keeps equal, meanwhile the specificity raised around two or three percent. Table 4 presents the results.

Dataset	Sensitivity (%)	Specificity (%)	PPV(%)	NPV(%)
Art	70,83	25,00	42,50	52,27
Books	93,29	23,38	54,09	78,26
Cones	70,51	46,32	35,03	79,28
Dolls	80,00	44,62	38,64	83,65
Laundry	97,42	25,25	71,86	83,33
Moebius	84,13	43,33	38,41	86,67
Reindeer	87,93	27,97	33,12	85,11
Teddy	78,86	34,20	43,30	71,74

Table 3: Confusion Matrix using the FAST corners.

Dataset	Sensitivity (%)	Specificity (%)	PPV(%)	NPV(%)
Art	69,44	28,26	43,10	54,17
Books	92,62	24,68	54,33	77,55
Cones	70,5	50,53	36,91	80,67
Dolls	77,65	47,18	39,05	82,88
Laundry	97,42	27,27	72,41	84,38
Moebius	82,54	46,00	39,10	86,25
Reindeer	87,93	31,47	34,23	86,54
Teddy	78,86	38,34	44,91	74,00

Table 4: Confusion Matrix using the FAST corners with Relaxed Condition.

5. Conclusions

In this paper, an approach for classifying estimated corresponding points based on the Delaunay triangulation is presented. The performance and accuracy were evaluated in order to determine its pertinence of use.

The classification was done by triangulating the estimated corresponding points of feature points obtained using the SIFT algorithm and the FAST corners.

The proposed approach presents better performance using the FAST corners than using the SIFT keypoints, in terms of sensitivity, which is the most representative measure of performance in this context.

The proposed approach produces error propagation where one bad match affects directly the neighbour's points in the

graph. Also, it exhibits the best performance with the FAST corners, where the percentage of initial bad matches is larger than approximately 50%. It means, that the proposed approach classifies correctly the initial bad matches – 93% and 97% of them – when the initial corresponding points have a large percentage of "bad matches".
The considered relaxed evaluation condition worked better using the FAST corners, because of the locations of the points. Using the SIFT algorithm, there are many internal points, for this reason, the new evaluation criteria only decrease the sensitivity. In the FAST corners situation, the criterion presents the same sensitivity and higher specificity for half of the datasets, while for the remaining the specificity increase, but the sensitivity decrease.

REFERENCES

[1] E. Trucco and A. Verri. "Stereopsis", in *Introductory Techniques for 3D computer vision*, 1th ed. Cambridge, MA: Prenticen Hall, 1998, ch. 7, pp. 139-151.

[2] Jet Propulsion Laboratory, California Institute of Technology (2012).*Mars exploration rovers*[Online]. Available: http://marsrovers.jpl.nasa.gov/home/index.html.

[3] Google (2012). *German scientists unveil self-driving car*[Online]. Available: http://www.google.com/hostednews/afp/article/ALeqM5izCRPtPhsc2k6kCY2_YOeUs7r_w?docId=CNG.c668be9320e3376a10d767deb0b0649f.a91.

[4] Koninklijke Philips Electronics N.V. (2012). *Syntegra*[Online]. Avaible: http://www.healthcare.philips.com/in_en/products/ros/products/syntegra.

[5] J. Paul Siebert, B. Cyganek. "Image Matching Algorithms", in*An Introduction to 3D Computer Vision Techniques andAlgorithms*, 2nd ed. West Sussex, UK: John Wiley & Sons, 2011, ch. 6. pp. 193-306.

[6] J. Sun, N.N. Zheng, and H.Y. Shum. *Stereo matching using belief propagation*. IEEE Trans-actions on Pattern Analysis and Machine Intelligence, pages 787-800, 2003.

[7] M. Bleyer and M. Gelautz. *Graph-cut-based stereo matching using image segmentation with symmetrical treatment of occlusions*. Signal Processing: Image Communication, 22(2):127-143, 2007.

[8] M.C. Sung, S.H. Lee, and N.I. Cho. *Stereo matching using multi-directional dynamic programming and edge orientations*. IEEE International Conference on Image Processing, ICIP, 2007, pp. I-233.

[9] M. Kardouchi, J. Konrad, and C. Vazquez. Estimation of large-amplitude motion and disparity Fields: Application to intermediate view reconstruction. Proc. *SPIE Visual Communications and Image Process*, 2001, pp. 340-351.

[10] D. Zhou, G. Li, Y. Liu, "Effective Corner Matching Based on Delaunay Triangulation", in *International Conference on Robotics & Automation*, New Orleans, LA, 2004, pp. 2730-2735.

[11] S. Zhang, X. Qu, S. Ma, Z. Yang and L. Kong, "A Dense Stereo Matching Algorithm Based on Triangulation", in *Journal of Computational Information Systems*, 2012, pp. 283-292.

[12] A. Geiger, M. Roser and R. Urtasun, "Efficient Large-Scale Stereo Matching", in *ACCV'10 Procedings of the 10th Asian Conference of Computer Vision*, 2010, pp. 25-38.

[13] C. Chen, D. Sargent, C. Tsai, Y. Wang, D Koppel, "Stabilizing Stereo Correspondence Computation Using Delaunay Triangulation and Planar Homography", *ISVC '08 Proceedings of the 4th International Symposium on Advances in Visual Computing*, 2008, pp. 836-845.

[14] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection", *European Conference on Computer Vision*, 2006, pp. 430-443.

[15] D. Lowe (2005). *Demo software: Sift keypoint detector* [Online]. Available: http://www.cs.ubc.ca/~lowe/keypoints/siftDemoV4.zip

[16] D.G. Lowe. *Distinctive image features from scale-invariant keypoints*. International journalof computer vision, 2004, pp. 91-110.

[17] P. Gallardo (2010). *Notas de matematica discreta* [Online]. Available: http://www.uam.es/personal_pdi/ciencias/gallardo/capitulo8a.pdf.

[18] D. Scharstein (2011). *Middlebury stereo datasets*[Online].Avaible: http://vision.middlebury.edu/stereo/data/.

Figure 7. Middlebury Stereo Datasets: (a) Art. (b) Books. (c) Cones. (d) Dolls. (e) Laundry. (f) Moebius. (g) Reindeer. (h) Teddy.