An Efficient Image Depth Extraction Method Based on SVM

Zhipeng Fan, Mingjun Li and Ying Lu

School of Computer and Information Engineering, Harbin University of Commerce, Harbin 150028, China

Abstract

Compared with the two-dimensional media, the image depth of three-dimensional media can offer more intuitive and real scenes to the audience for feeling. But with the development of 3D display machines, there is a serious contradiction between the rapid of the 3D display machines and the lack of resources for the machines. To solve the problem, proposed an efficient image depth extraction method which utilizes the support vector machine (SVM). The label is established from the true depth of different videos, and the vector feature, haze is utilized as the feature vector. The training set is divided into a number of small training sets, reducing the sample size of the training sets, in order to improve training speed. The experimental results show that the algorithm is effective.

Keywords: *feature vector, the depth map, SVM*

1. Introduction

Three-dimensional television (3D-TV) is nowadays often seen as the next major milestone in the ultimate visual experience of media. Although the concept of stereoscopy has existed for a long time, the breakthrough from conventional 2D broadcasting to real-time 3D broadcasting is still pending. However, in recent years, there has been rapid progress in the fields image capture, coding and display [1], which brings the realm of 3D closer to reality than ever before.

Three-dimensional video with a wide range of applications in the field of industrial simulation, architectural design, military simulation, medical and health, education, machinery manufacturing, pour observed. measurements, entertainment and advertising media.2D-3D video conversion technology, therefore, will promote the development of technology in these areas.

2. Depth of Image Extraction Method

Currently, most of the 2D-3D video conversion technology via dense depth maps for each frame of the sequence using depth-image-based rendering (DIBR). As shown in Figure 1, which involves the projection of a viewpoint into another view? This projection is actually based on warping [2]. Based on this theory, depth extraction is a key issue for 2D-3D video conversion, becoming an important direction in the 2D-3D video conversion technology.

Depth map estimation techniques generally fall into one of the following categories: manual, semi automatic and automatic. For the manual methods, an operator manually traces the outlines of objects that are associated with an artistically chosen depth value. As expected, these methods are extremely time consuming and expensive. For this reason, semi automatic and automatic techniques are preferred. These techniques are designed based on the visual depth perception mechanism. There are several factors (referred as monocular depth cues) such as light and shade, relative size, motion parallax, interposition (partial occlusion),

textural gradient and geometric perspective, which help the viewer to perceive the relative distance of objects within a scene. In fact, the depth map estimation techniques try to generate binocular parallax (disparity) using monocular depth cues. But each clue has its specific scope [3], relying on a single clue to extract depth information is limited by the specific scene .For example, Cheng, *et al.*, Proposed 2D video to 3D video conversion algorithm based on the macroblock's bilateral filtering [4], and Feng X., *et al*., proposed Optical flow method to obtain the motion vector field to determine the depth information technology [5]. Both methods are based on Based on vision geometry, objects moving horizontally at a similar speed but with a different distance to the camera will produce different results in a recorded sequence, such as the closer the object to the camera, the bigger the change in distance of two continuously frames. This algorithm ideal when the camera motion, scene still. Otherwise can not be satisfied with the estimated results.

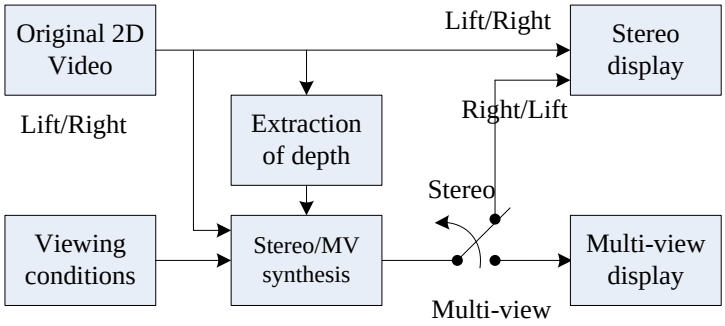


Figure 1. The 3D Image Generation System Block Diagram based on DIBR

Zhou, *et al*., proposed based on DFD（depth from defocus）Edge ambiguity model established by the Gaussian gradient method [6], improved in Edge treatment effect and Scene depth calculation accuracy by this method. But little defocus comparison image result is not satisfactory. In order to compensate for this deficiency, proposed a kind of one clue-based, other clues as a supplement method for depth extraction. Y. Feng, *et al*., proposed based on the defocus combine with optical flow depth estimation algorithm is proposed [7]. However, the way which multi clue combine together is unknown, the accuracy of the final result by the judge of a clue to get the wrong results will be affected. For instance, when the depth map is estimated by defocus, a large area of sky background will be misidentified as clear foreground, it will have a certain influence on the effect of the final depth map.

3. The Depth Extraction Methods based on Support Vector Machine

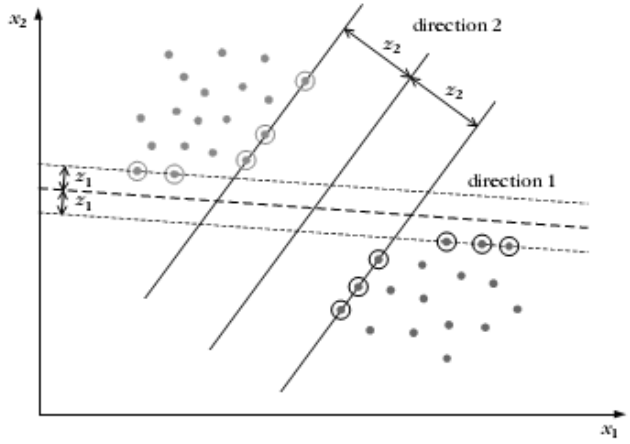
The human eye can obtain depth information from a two-dimensional video. Therefore, when viewing movies or photos. Observer can easily recognize depths of the relationship between the different feature. This is because the human brain can get different clues from the image, and combine these cues together to determine the depth information. But the way of multi clue combine together is not clear now. If the depth extraction algorithm also has such a multi clues learning ability, it will have a greater adaptability. SVM can obtain approximate exists but does not explicitly model through small sample learning. In this paper, therefore, the image depth extraction algorithm based on support vector machine is proposed. The support vector machine (SVM) training study was used for more clues combining ways to create a model of the image depth information. Using the model to extract the texture change,

texture gradient, and a haze clues of the image, combined them together for depth prediction. Therefore the accuracy of the extracted of the depth was increased.

3.1. Support Vector Machine

The original SVM algorithm was invented by Vladimir N. Vapnik and the current standard incarnation (soft margin) was proposed by Vapnik and Corinna Cortes in 1995 [8].

More formally, a support vector machine constructs a hyper-plane or set of hyper-planes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyper-plane that has the largest distance to the nearest training data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier [9].



Consider the problem of separating the set of training vectors belonging to two separate classes [10].

$$\{x_i, y_i\} \quad i = 1, \cdots, n$$

$$x \in R^d$$

$$y_i \in \{-1, +1\}$$

With a hyper-plane,

$$w^T x + b = 0$$

The set of vectors is said to be optimally separated by the hyper-plane if it is separated without error and the distance between the closest vector to the hyper-plane is maximal.

A separating hyper-plane in canonical form must satisfy the following constraints,

$$y_i(w \cdot x_i - b) \geq 1, 1 \leq i \leq n. \tag{3.1}$$

The optimal hyper-plane is given by maximising the margin,.

$$m\arg in = \min_{\{x_i | y_i = 1\}} \frac{w^T \square x_i + b}{\|w\|} - \max_{\{x_i | y_i = -1\}} \frac{w^T \square x_i + b}{\|w\|} = \frac{2}{\|w\|} \tag{3.2}$$

Hence the hyper-plane that optimally separates the data is the one that minimizes subject to (3.1)

$$\min_{w} \frac{\|w\|^2}{2} \qquad (3.3)$$
$$s.t. \quad y_i(w^T x_i + b) - 1 \geq 0, i = 1, 2 \cdots n$$

The solution to the optimization problem of Equation 3.3 under the constraints of Equation 3.1 is given by the saddle point of the Lagrange functional (Lagrangian) ,

$$L = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{l} \alpha_i y_i (x_i \cdot w + b) + \sum_{i=1}^{l} \alpha_i \qquad (3.4)$$

$$\frac{\partial L}{\partial b} = 0 \Leftrightarrow \sum_{i=1}^{n} \alpha_i y_i = 0 \qquad (3.5)$$

We can get the dual problem of (3.3),

$$\max_{\alpha} Q(\alpha) = L(w, b, \alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j \qquad (3.6)$$
$$s.t. \quad \sum_{i=1}^{n} \alpha_i y_i = 0, \alpha_i \geq 0$$

the optimal separating hyper-plane is given by,

$$w^* = \sum_{i=1}^{l} \alpha_i^* y_i x_i \qquad (3.7)$$

The hard classifier is then,

$$f(x) = Sgn\left\{ \sum_{i=1}^{l} y_i \alpha_i^* (x_i \cdot x) + b^* \right\} \qquad (3.8)$$

The discriminant function is used to solve the problem of binary classification. , Generally the multi-classification problem is separated severe binary support vector machine. By combining multiple binary support vector machine (BSVM) achieve multiple classification. Currently used methods: One-to-rest, One-to-one, Directed Acyclic Graph, Binary Tree, *et al*., In this paper, one-to-one way is used to achieve multi-class classification [11]. In this way, between any two types of samples, training a SVM. For K classes samples, the number of SVM which need to be constructed is K (K-1)/2. When predicting a sample, all of SVM classifier should be used for classification, cumulative forecast score of each category, Select the highest score as the category type of data for testing.

3.2 Depth extraction based on support vector machine

The structure which depth extraction method of image as show Figure 2. System processing is divided into data preparation and SVM learning which will be down by two step [12]. In the data preparation stage, Classification labels obtained by processing the depth map, depth characteristic is extracted to make feature vector in the original image. At the SVM training stage. Firstly, the kernel function and its parameters should be determined, then

according to the classification label obtained by the data preparation phase and feature vector training SVM to get classification model. Ultimately, the depth map is obtained by the model and feature vector.

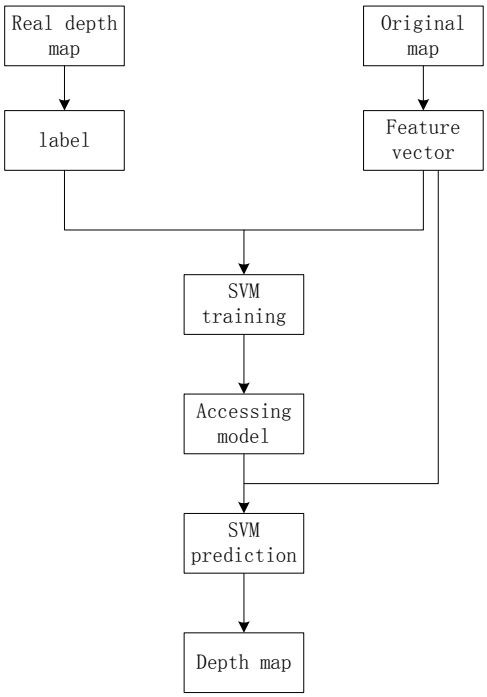


Figure 2. Block diagram of the image depth extraction method based on SVM

4. Image Depth Extraction Method

4.1. Classification and acquisition of Labeling

In this article depth forecast is based on macro-block, by prediction the depth of each macro-block, the depth map of each frame is achieved. For every macro-block, the mean value of the depth as the depth of the macro-block.

$$d_{patch(i)} = \sum_{(m,n)\in patch(i)} \frac{d(m,n)}{N^2}$$

In addition，when the two-dimensional image is observed by human eyes，the depth of different scene is classified into different depth value automatically. In this projected，the depth of each macroblock is classified into one of L levels. The number of level as the label in the SVM classifier is

$$label_{patch(i)} = floor\left(d_{patch(i)} / l\right) + 1$$

After considering the subjective feelings and deal with complexity，in this paper the scale of macro-block is 8*8, and the level is 16.

4.2. Extract Feature Vector

A large amount of depth information has been lost in the production process of the two-dimensional image，to restore depth information, Depth characteristics need to be extracted according to the residual in the image different cues，this is a very important prerequisite for inference depth. The depth of the feature extraction of the present study learn from [6].

The experience shows that depth information is difficult predicted using a single macro-block，for determined the depth, the adjacent or larger scene should be referenced. For an isolated blue macro-block，it could not be judged as the sky or a part of the near blue objects. Therefore, global characteristics of the features of depth should be taken account. The multi-scale Order Pro-domain system was established in the original picture (Figure 3). This can be introduced characterized with the directly adjacent macro blocks and macro-block farther away from it for the judgment of depth. In addition，Because of the scene in the image (such as buildings, trees, people, *etc*.) great majority are vertical，the column where the macro block is divided into equal four portions, forming four columns macro-block which can capture vertical attribute of the macro-block. Therefore, for a macro-block, the characteristics is represented by the 19 macro-block which have space constraint relation with it.

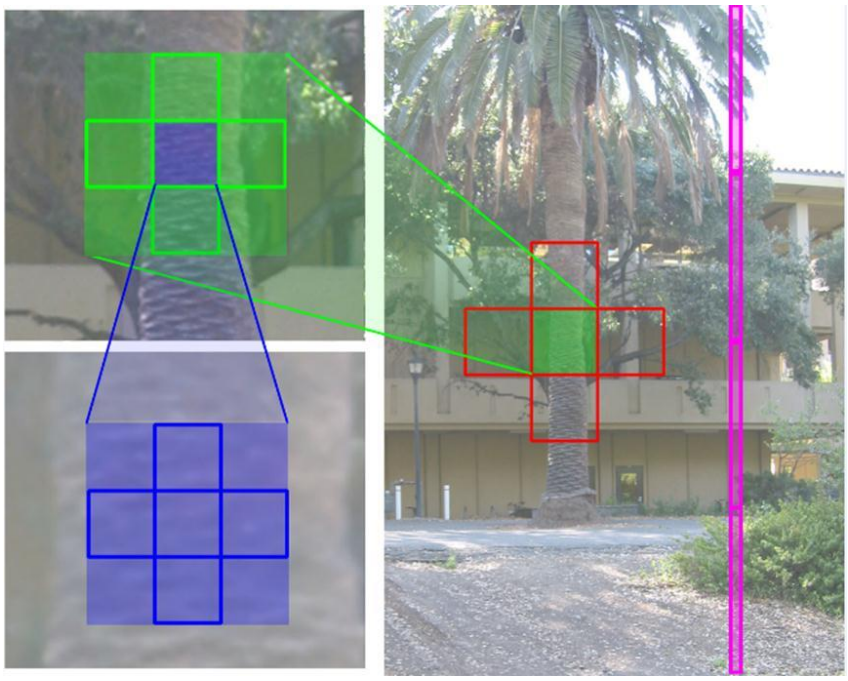


Figure 3. The Multi-Scale Order Pro-Domain System

In the image，Different depths of the scene exhibits a texture change, texture gradient, haze and other characteristics are also different. Therefore，In this study, a given set of filtered is used. (Figure 4), To obtain the depth features of the image according to these three clues. The first nine template is a template of a set of Laws' Mask with 3*3 size proposed by Laws，can be used to detect image changes in local mean edge, spots, texture information；the following six mask with 5 * 5 templates are directional detectors，can be used to extract image texture gradient information. For YUV test sequence，Most texture changes and texture gradient information exists in the Y channel. Atmospheric scattering

more obvious in the low frequency part of the U, V-channel，therefore a filter is used acting on the Y channel of the image to obtain texture changes and texture gradient； Take advantage of the first Laws template acting on the image U, V channels to obtain haze. The 17 filters $F_n(x,y)$,$(n = 1,\cdots,17)$ convolution with the image, and seeking the absolute energy and the square of the energy,

$$E_i(n) = \sum_{(x,y)} \in patch(i)|I(x,y) * F_n(x,y)|^k \quad k = 1,2$$

for each center macro-block i obtained feature vector by the 646-dimensional data .

(a) Laws template (b) the direction of the detector

Figure 4. Laws Template and the Direction of the Detector

4.3. Experimental Results

SVM processing stage. Firstly, the nuclear function, parameter and data preprocessing mode of SVM used should be determined. At present, kernel function and parameter selection of SVM need to be determined by experience or experiment compared. In addition，in order to avoid the numerical calculation difficulties caused by calculate the inner product, the data usually be scaled to [-1,1] or [0,1]. Experiment with the method of cross-validation experiments compared choose nuclear functions, parameters, and data scaling way.

In order to verified the effective of proposed algorithm. The depth map of the fiftieth frame of the test sequence kendo_1, newspaper, lovebird were extracted. Data preparation was completed by using matlab，Libsvm package was used to training and prediction. The experimental results shown in Figure 5, from left to right are the original image, and the real depth map, depth map after quantization, and a depth map after predicted. The accuracy of experimental results were 82.55%、85.7992%、94.0592%. Experimental results show that, the proposed algorithm can get better results in these types of cases.

a) kendo_1

b) newspaper

c) lovebird

Figure 5. Experimental Results Contrast

5. Conclusions

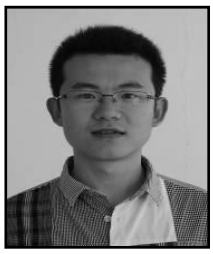
In this paper, the depth extraction algorithm based on the SVM is proposed. The combing rules of multiple clue of SVM training samples is used. Texture changes, texture gradient and haze clues is combined together to determine depth information，thereby the accuracy of the depth of the extracted is improved. In this paper, the experiments result improved the feasibility of the program. Compared to the single clue for depth extraction algorithm，the algorithm can get better depth map in different scenes. For SVM can be increased more clues by the increase in the number of dimensions of the feature vector, extracted depth cues from the two-dimensional image and how to choose a more effective depth characteristics as the characteristic matrix of SVM is the next focus of research. For the booming of 3D video applications, how to effectively take the advantage of depth information is important and worthy of further study. Future works will mainly focus on two parts. The first is extracting more depth cues to further enhance the accuracy of depth map, the second is improving the influence way of visual attention in 3D perception.

References

[1] W. A. IJsselsteijn, P. J. H. Seuntiëns and L. M. J. Meesters, “State-of-the-art in Human Factors and Quality Issues of Stereoscopic Broadcast Television”, Deliverable ATTEST/WP5/01, Eindhoven University of Technology, the Netherlands, http://www.hitech-projects.com/euprojects/attest/deliverables/Attest-D01.pdf, **(2002)**.
[2] L. McMillan Jr., “An image-based approach to three-dimensional computer graphics, PhD thesis, **(1997)**, Chapel Hill, NC, USA.
[3] F. Christoph, “A 3D-TV Approach Using Depth-Image-Based Rendering (DIBR)”, Visualization, Imaging, and Image Processing (VIIP), (**2003**) Benalmadena, Spain.
[4] C. Chaochung, L. Chengte and H. Posun, “A Block-based 2D-to-3D Conversion System with Bilateral Filte”, International Conference on Consumer Electronics, **(2009),** Las Vegas, NV, USA.
[5] X. Feng, E. Guihua and X. Xudong, “2D-to-3D Conversion Based on Motion and Color Mergence”, 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, **(2008)**, Istanbul, Turkey.
[6] Z. Shaojie and S. Terence, “Defocus map estimation from a single image”, Pattern Recognition, vol. 5, **(2011)**.
[8] S. Ashuotsh, S. Jamie and Y. N. Andrew, “Depth estimation using monocular and stereo cues”, International Joint Conference on Artificial Intelligence (IJCAI), **(2007)**, Hyderabad, India.
[9] K. Laws, “Texture energy measures”, Proceedings of Image Understanding Workshop, **(1979)**.

[10] M. L. Samb, F. Camara, S. Ndiaye, Y. Slimani and M. A. Esseghir, "International Journal of Advanced Science and Technology, vol. 43, **(2012)**.
[11] D. Ben Ayed Mezghani, S. Zribi Boujelbene and N. Ellouze, International Journal of Hybrid Information Technology, vol. 3, no. 3, **(2010)**.
[12] T Hoang Le and Len Bui, International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 4, no. 3, **(2011)**.

Authors



Zhipeng Fan, Master, work in Computer and Information Engineering, Harbin University of Commerce, China; lecturer. The major research fields: Image and Signal Processing, Communications and Networking, Computational Science and Technology.Email:fzp369@163.com.

Mingjun Li, Master, work in Computer and Information Engineering, Harbin University of Commerce, China; associate professor. The major research fields: artificial intelligence, information services; e-commerce and e-government.

Ying Lu, Master, work in Computer and Information Engineering, Harbin University of Commerce, China; lecturer. The major research fields: artificial intelligence, information services; e-commerce and e-government.