CLAIMS

1. A binocular matching method, comprising:

obtaining an image to be processed, wherein the image is a two-dimensional (2D) image comprising a left image and a right image;

constructing a three-dimensional (3D) matching cost feature of the image by using extracted features of the left image and extracted features of the right image, wherein the 3D matching cost feature comprises a group-wise cross-correlation feature, or comprises a feature obtained by concatenating the group-wise cross-correlation feature and a connection feature; and

determining the depth of the image by using the 3D matching cost feature.

2. The method according to claim 1, wherein the constructing a 3D matching cost feature of the image by using extracted features of the left image and extracted features of the right image comprises:

determining the group-wise cross-correlation feature by using the extracted features of the left image and the extracted features of the right image; and determining the group-wise cross-correlation feature as the 3D matching cost feature; or

determining the group-wise cross-correlation feature and the connection feature by using the extracted features of the left image and the extracted features of the right image; and determining the feature obtained by concatenating the group-wise cross-correlation feature and the connection feature as the 3D matching cost feature; wherein the connection feature is obtained by concatenating the features of the left image and the features of the right image in a feature dimension.

3. The method according to claim 2, wherein the determining the group-wise cross-correlation feature by using the extracted features of the left image and the extracted features of the right image comprises:

grouping the extracted features of the left image and the extracted features of the right image respectively, and determining cross-correlation results of the grouped features of the left image and the grouped features of the right image under different parallaxes; and

concatenating the cross-correlation results to obtain a group-wise cross-correlation feature.

4. The method according to claim 3, wherein the grouping the extracted features of the left image and the extracted features of the right image respectively, and determining cross-correlation results of the grouped features of the left image and the grouped features of the right image under different parallaxes comprises:

grouping the extracted features of the left image to form a first preset number of first feature groups;

grouping the extracted features of the right image to form a second preset number of second feature groups, wherein the first preset number is the same as the second preset number; and

determining a cross-correlation result of the g-th first feature group and the g-th second feature group under each of the different parallaxes, wherein g is a natural number greater than or equal to 1 and less than or equal to the first preset number; the different parallaxes comprise: a zero parallax , a maximum parallax , and any parallax between the zero parallax and the maximum parallax ; and the maximum parallax is a maximum parallax in the usage scenario corresponding to the image to be processed.

5. The method according to any one of claims 1 to 4, wherein the determining the depth of the image by using the 3D matching cost feature comprises:

determining, by a 3D neural network, a probability of each of different parallaxes corresponding to each pixel point in the 3D matching cost feature;

determining a weighted mean of probabilities of the different parallaxes corresponding to the pixel point;

determining the weighted mean as a parallax of the pixel point; and

determining the depth of the pixel point according to the parallax of the pixel point.

6. A training method for a binocular matching network, comprising:

determining, by a binocular matching network, a 3D matching cost feature of an obtained sample image, wherein the sample image comprises left image and right image with depth annotation information, the left image and right image are the same in size; and the 3D matching cost feature comprises a group-wise cross-correlation feature, or comprises a feature obtained by

concatenating the group-wise cross-correlation feature and a connection feature;

determining, by the binocular matching network, a predicted parallax of the sample image according to the 3D matching cost feature;

comparing the depth annotation information with the predicted parallax to obtain a loss function of binocular matching; and

training the binocular matching network by using the loss function.

7. The method according to claim 6, wherein the determining, by a binocular matching network, a 3D matching cost feature of an obtained sample image comprises:

determining, by a full convolutional neural network in the binocular matching network, 2D concatenated features of the left image and 2D concatenated features of the right image respectively; and

constructing the 3D matching cost feature by using the 2D concatenated features of the left image and the 2D concatenated features of the right image.

8. The method according to claim 7, wherein the determining, by a full convolutional neural network in the binocular matching network, 2D concatenated features of the left image and 2D concatenated features of the right image respectively comprises:

extracting, by the full convolutional neural network in the binocular matching network, 2D features of the left image and 2D features of the right image respectively;

determining an identifier of a convolution layer for performing 2D feature concatenation;

concatenating the 2D features of different convolution layers in the left image in a feature dimension according to the identifier to obtain first 2D concatenated features; and

concatenating the 2D features of different convolution layers in the right image in the feature dimension according to the identifier to obtain second 2D concatenated features.

9. The method according to claim 8, wherein the determining an identifier of a convolution layer for performing 2D feature concatenation comprises: determining the i-th convolution layer as a convolution layer for performing 2D feature concatenation when the interval rate of the i-th convolution layer changes, wherein i is a natural number greater than

or equal to 1.

10. The method according to claim 8 or 9, wherein the full convolutional neural network is a full convolutional neural network sharing parameters;

the extracting, by the full convolutional neural network in the binocular matching network, 2D features of the left image and 2D features of the right image respectively comprises: extracting, by the full convolutional neural network sharing parameters in the binocular matching network, the 2D features of the left image and the 2D features of the right image respectively, wherein the size of the 2D feature is a quarter of the size of the left image or the right image.

11. The method according to any one of claims 7 to 10, wherein the constructing the 3D matching cost feature by using the 2D concatenated features of the left image and the 2D concatenated features of the right image comprises:

determining the group-wise cross-correlation feature by using the obtained first 2D concatenated features and the obtained second 2D concatenated features; and determining the group-wise cross-correlation feature as the 3D matching cost feature; or

determining the group-wise cross-correlation feature by using the obtained first 2D concatenated features and the obtained second 2D concatenated features; determining the connection feature by using the obtained first 2D concatenated features and the obtained second 2D concatenated features; and concatenating the group-wise cross-correlation feature and the connection feature in a feature dimension to obtain the 3D matching cost feature.

12. The method according to claim 11, wherein the determining the group-wise cross-correlation feature by using the obtained first 2D concatenated features and the obtained second 2D concatenated features comprises:

dividing the obtained first 2D concatenated features into N_g groups to obtain N_g first feature groups;

dividing the obtained second 2D concatenated features into N_g groups to obtain N_g second feature groups, N_g being a natural number greater than or equal to 1;

determining a cross-correlation result of each of the N_g first feature groups and a

respective one of the N_g second feature groups under each parallax d, to obtain $N_g * D_{max}$ cross-correlation maps, wherein the parallax d is a natural number greater than or equal to 0 and less than D_{max}, and D_{max} is the maximum parallax in the usage scenario corresponding to the sample image; and

concatenating the $N_g * D_{max}$ cross-correlation maps in a feature dimension to obtain the group-wise cross-correlation feature.

13. The method according to claim 12, wherein the determining a cross-correlation result of each of the N_g first feature groups and a respective one of the N_g second feature groups under each parallax d, to obtain $N_g * D_{max}$ cross-correlation maps comprises:

determining a cross-correlation result of the g-th first feature group and the g-th second feature group under each parallax d, to obtain D_{max} cross-correlation maps, wherein g is a natural number greater than or equal to 1 and less than or equal to N_g; and

determining cross-correlation results of the N_g first feature groups and the N_g second feature groups under each parallax d, to obtain $N_g * D_{max}$ cross-correlation maps.

14. The method according to claim 11, wherein the determining the connection feature by using the obtained first 2D concatenated features and the obtained second 2D concatenated features comprises:

determining a concatenation result of the obtained first 2D concatenated features and the obtained second 2D concatenated features under each parallax d, to obtain D_{max} concatenation maps, wherein the parallax d is a natural number greater than or equal to 0 and less than D_{max}, and D_{max} is the maximum parallax in the usage scenario corresponding to the sample image; and

concatenating the D_{max} concatenation maps to obtain the connection feature.

15. The method according to claim 6, wherein the determining, by the binocular matching network, a predicted parallax of the sample image according to the 3D matching cost feature comprises:

performing, by the binocular matching network, matching cost aggregation on the 3D matching cost feature; and

performing parallax regression on the aggregated result to obtain the predicted parallax of the sample image.

16. The method according to claim 15, wherein the performing, by the binocular matching network, matching cost aggregation on the 3D matching cost feature comprises:

determining, by a 3D neural network in the binocular matching network, a probability of each different parallax d corresponding to each pixel point in the 3D matching cost feature, wherein the parallax d is a natural number greater than or equal to 0 and less than D_{max}, and D_{max} is the maximum parallax in the usage scenario corresponding to the sample image.

17. The method according to claim 15, wherein the performing parallax regression on the aggregated result to obtain the predicted parallax of the sample image comprises:

determining a weighted mean of probabilities of respective different parallaxes d corresponding to each pixel point as the predicted parallax of the pixel point, to obtain the predicted parallax of the sample image;

wherein each of the parallaxes d is a natural number greater than or equal to 0 and less than D_{max}, and D_{max} is the maximum parallax in the usage scenario corresponding to the sample image.

18. A binocular matching apparatus, comprising:

an obtaining unit, configured to obtain an image to be processed, wherein the image is a two-dimensional (2D) image comprising a left image and a right image;

a constructing unit, configured to construct a three-dimensional (3D) matching cost feature of the image by using extracted features of the left image and extracted features of the right

image, wherein the 3D matching cost feature comprises a group-wise cross-correlation feature, or comprises a feature obtained by concatenating the group-wise cross-correlation feature and a connection feature; and

a determining unit, configured to determine the depth of the image by using the 3D matching cost feature.

19. A computer device, comprising a memory and a processor, wherein the memory stores a computer program running on the processor, wherein when the processor executes the program, steps of the binocular matching method according to any one of claims 1 to 5 are implemented, or steps of the training method for a binocular matching network according to any one of claims 6 to 17 are implemented.

20. A computer readable storage medium having a computer program stored thereon, wherein when the computer program is executed by a processor, steps of the binocular matching method according to any one of claims 1 to 5 are implemented, or steps of the training method for a binocular matching network according to any one of claims 6 to 17 are implemented.