INTERNATIONAL JOURNAL OF RESEARCH CULTURE SOCIETY
Monthly Peer-Reviewed, Refereed, Indexed Journal
Volume - 9, Issue - 11, November- 2025
ISSN(O): 2456-6683
[Impact Factor: 9.241]
IJRCS

DOIs:10.2017/IJRCS/202511013 --:-- Research Paper / Article / Review

A Survey on Stereo Vision and Super-Resolution: Recent Methods

[1] Toral Patel, [2] Vaishali Parmar, [3] Jigar Dalvadi
[1,2,3] Assistant Professor,
[1,2] Information Technology, Sardar Patel College of Engineering, Bakrol, Anand, India
[3]Computer Engineering, Sardar Patel College of Engineering, Bakrol, Anand, India
Email - [3]jigardalvadi1992@gmail.com

Abstract: *Stereo vision and image super-resolution (SR) are two fundamental areas in computer vision that have traditionally been addressed independently. Stereo vision aims to recover depth information by finding correspondences between two or more images, while super-resolution enhances the spatial resolution of an image. However, in real-world applications, limitations in sensor resolution often lead to a trade-off between field of view and image detail, adversely affecting the accuracy of depth estimation. Recently, there has been a growing interest in synergistically combining these two tasks to overcome their individual limitations. This survey provides a comprehensive overview of recent methods that integrate stereo vision with super-resolution. We categorize the existing approaches into sequential, concurrent, and deep learning-based frameworks. The literature survey explores key advancements, and a comparative analysis is presented based on critical evaluation parameters such as depth map accuracy, image quality metrics, and computational efficiency. The paper concludes by discussing open challenges and promising future research directions, highlighting the potential of end-to-end deep learning models to jointly optimize for both high-resolution imagery and precise depth estimation. .*

Key Words: *Stereo Vision, Super-Resolution, Depth Estimation, Deep Learning, Image Enhancement, Computer Vision, Survey.*

1. INTRODUCTION

Stereo vision is a cornerstone of 3D computer vision, enabling machines to perceive depth by triangulating corresponding points from a pair of images taken from slightly different viewpoints [1]. Its applications span autonomous driving, robotics, and 3D reconstruction. The accuracy of stereo matching is highly dependent on the resolution and quality of the input images. Low-resolution (LR) inputs often contain insufficient texture information, leading to corrupted or incomplete depth maps, especially in occluded and textureless regions [2].

Conversely, image super-resolution is the process of reconstructing a high-resolution (HR) image from one or more LR observations. With the advent of deep learning, particularly Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs), SR has achieved remarkable success in recovering fine details [3].

The intrinsic link between these two tasks has motivated research into their combined processing. Applying SR to stereo images can be approached in several ways: performing SR on the input images before stereo matching (sequential approach), performing SR on the disparity map after stereo matching, or jointly solving both problems in a unified framework. The primary challenge lies in ensuring that the super-resolved images are geometrically consistent with each other to facilitate accurate stereo matching, a constraint not present in single-image SR.

This survey paper aims to:

- Systematically review and categorize recent methodologies that integrate stereo vision and super-resolution.
- Provide a comparative analysis of these methods based on standard evaluation metrics.
- Identify current trends, challenges, and potential future directions in this interdisciplinary field.

2. LITERATURE REVIEW:

The integration of stereo and SR has evolved from traditional model-based methods to sophisticated deep learning architectures.

2.1 Traditional and Sequential Methods

Early approaches were predominantly sequential. One common strategy was to apply single-image SR to the left and right images independently before stereo matching [4]. However, this often led to geometric inconsistencies, as the SR process did not account for the epipolar constraints between the stereo pair. To address this, some methods proposed a coupled SR and stereo matching framework within a Maximum a Posteriori (MAP) estimation, enforcing consistency during the SR process [5]. These methods, while theoretically sound, were often computationally intensive and struggled with complex, real-world scenes.

2.2 Deep Learning-Based Revolution

The rise of deep learning has been transformative. Initial deep learning methods replaced individual components; for instance, using a CNN-based SR network (e.g., SRCNN [3], EDSR [6]) as a pre-processing step, followed by a learning-based stereo matcher like GC-Net [7] or PSMNet [8]. While this improved performance over traditional methods, it remained a disjointed pipeline. A significant advancement was the development of end-to-end models that perform joint stereo and SR. StereoSR [9] was a pioneering work that used a residual network to super-resolve the left and right views by leveraging complementary information from the stereo pair through a parallax-attention mechanism. Following this, PASSRnet[10] introduced a parallax-attention module to implicitly capture the correspondence between views, enabling more effective information fusion for stereo SR. More recent works have explored task-oriented joint learning. Some methods propose a feedback mechanism where the initial depth estimate guides the SR process, and the enhanced image, in turn, refines the depth estimate [11]. Others have explored the super-resoluxtion of the disparity map itself, using the high-resolution left image as a guide in a deep learning framework, treating it as a depth map super-resolution problem [12]..

3. METHODS CLASSIFICATION:

We classify the existing methods into three main categories based on their architectural and processing workflow.

3.1. Sequential Methods (SR -> Stereo or Stereo -> SR):

Sequential methods treat super-resolution (SR) and stereo vision as two independent, modular tasks executed one after the other in a pipeline. This is the most intuitive and straightforward approach to combining the two technologies. The fundamental characteristic is that the output of the first stage becomes the input for the second stage, with no feedback loop or joint optimization between them. This category is primarily divided into two distinct sub-categories based on the order of operations.

1. Image-SR-First Approach

This strategy prioritizes enhancing image quality before attempting depth estimation.

- **Input:** A pair of low-resolution left and right images (L_LR, R_LR).
- **Super-Resolution Stage:** Each image of the stereo pair is processed by a super-resolution algorithm. This can be done:
  - **Independently:** The left and right images are super-resolved separately using a single-image SR network (e.g., EDSR [6], RCAN). This is simple but ignores the correspondence between the two views.
  - **Jointly (Stereo SR):** A more sophisticated approach uses a dedicated stereo SR network (e.g., StereoSR [9], PASSRnet [10]). These networks are designed to leverage the parallax information between the two views to help fill in missing details in one image by borrowing information from the other, leading to more geometrically consistent results.
- **Output of SR Stage:** A pair of estimated high-resolution images (L_SR, R_SR).
- **Stereo Matching Stage:** The super-resolved images (L_SR, R_SR) are then fed into a standard stereo matching network (e.g., PSMNet [8], GC-Net [7]). This network performs correspondence search on the higher-quality inputs to compute a high-resolution disparity map.

2. Disparity-SR-First Approach

This strategy first computes a low-resolution depth estimate and then enhances its resolution.

- **Input:** The original low-resolution stereo pair (L_LR, R_LR).
- **Stereo Matching Stage:** A stereo matching network is applied directly to the LR inputs. This is faster and less memory-intensive than matching HR images because the feature maps and cost volumes are smaller.
- **Output of Stereo Stage:** A low-resolution, often noisy and blocky, disparity map (D_LR).
- **Disparity Super-Resolution Stage:** The LR disparity map is upscaled to the desired high resolution. This is not a simple interpolation (which would just blur the edges). It is typically a guided super-resolution process:

- **Guidance Image:** The high-resolution **left image** (which is often available from the camera) is used as a guide. The assumption is that depth edges align with color/intensity edges in the guidance image.
- **Model:** A network (e.g., a modified SRCNN or a dedicated depth SR network [12]) takes the D_LR and the HR guide image as input. It learns to sharpen the edges of the disparity map where the guide image has strong edges, while smoothing homogeneous regions.

3.2. Concurrent/Joint Methods

Concurrent or Joint methods break away from the sequential pipeline by solving the super-resolution (SR) and stereo vision problems simultaneously within a single, unified framework. The fundamental philosophy is that the two tasks are intrinsically linked and can inform each other. By optimizing them together, the model can learn a shared representation that is optimal for the final goal—producing a high-resolution disparity map—rather than just excelling at the intermediate tasks. This category has evolved from traditional model-based optimization to modern, end-to-end deep learning architectures.

1. Model-Based Joint Optimization

This was the dominant approach before the deep learning era, grounded in Bayesian estimation and energy minimization frameworks. Formulate a single energy function (or cost function) that combines constraints from both super-resolution and stereo vision.

- **Data Fidelity (SR Constraint):** The reconstructed high-resolution (HR) stereo pair should be consistent with the observed low-resolution (LR) input images when appropriately down-sampled (e.g., via a blurring and sub-sampling matrix).
- **Geometric Consistency (Stereo Constraint):** The HR stereo pair must adhere to the epipolar geometry. Corresponding pixels between the left and right HR images should be related by a disparity map.
- **Regularization:** Prior knowledge is added to make the problem well-posed, such as smoothness priors on the disparity map and the HR images.

2. End-to-End Deep Learning Networks

This is the modern and currently dominant approach. A single neural network is designed to take the LR stereo pair as input and produce the HR disparity map (and sometimes the HR images) directly. Use a deep neural network to implicitly learn the complex relationships and shared features between the two tasks from data, rather than relying on hand-crafted energy terms.

- **Shared Feature Encoder:** The network begins with a shared backbone that extracts deep features from both the left and right LR images. This allows the model to learn a common representation that is useful for both SR and disparity estimation.
- **Inter-View Information Fusion:** This is the most critical component. Instead of simply stacking the left and right features, advanced modules are used to fuse them:
  - **Parallax-Attention Mechanism (PAM):** Used in models like **PASSRnet**[10] and **StereoSR**[9]. This mechanism implicitly learns the correspondence between the two views by creating an attention map. For a pixel in the left image, it computes a similarity vector with all pixels on the same epipolar line in the right image's features. This allows the network to adaptively aggregate information from the corresponding region in the other view, which is crucial for both reconstructing detail (SR) and understanding geometry (Stereo).
- **Task-Specific Decoders:** After the fused feature representation is created, the network typically branches into task-specific decoder heads:
  - An **SR Decoder** that upsamples the features to reconstruct the HR left and right images.
  - A **Disparity Decoder** that uses the fused features to predict the high-resolution disparity map.
  - Often, these decoders are not fully independent; features can be shared between them to maintain consistency.
- **Cyclic and Iterative Architectures:** A more recent advancement introduces iterative refinement. For example, a **Feedback Network** [11] might:
  - Generate an initial HR image and disparity estimate.
  - Use the disparity map to warp the SR left image to the right view.
  - Calculate the photometric error (the difference between the warped image and the actual SR right image).
  - Feed this error back into the network to refine both the SR output and the disparity map in the next iteration.

4. COMPARATIVE ANALYSIS OF METHODS

The choice of technique is dictated by the specific clinical task, available data, and required level of interpretability.

Table 1: Qualitative comparison

Method Category	Representative Models	Advantages	Disadvantages
Sequential (Image-SR-First)	EDSR [6] + PSMNet [8]	Modular, easy to implement	Error propagation, geometric inconsistencies
Sequential (Disparity-SR-First)	VDSR [13] on disparity	Fast, can leverage guided SR	Dependent on initial LR disparity quality
Concurrent/Joint (Deep)	StereoSR [9], PASSRnet [10]	End-to-end, promotes consistency	Complex network design, requires paired LR-HR data
Iterative/Cyclic (Deep)	D-FeedbackNet [11]	Progressive refinement, high potential	Very high computational cost, training instability

Evaluation Parameters:

The performance of combined stereo-SR methods is evaluated using metrics from both super-resolution and stereo vision domains.

1 Super-Resolution Evaluation:

- **Peak Signal-to-Noise Ratio (PSNR):** Measures the fidelity of the reconstructed HR image against a ground-truth HR image. Higher is better.
- **Structural Similarity Index (SSIM):** Measures the perceptual similarity between the reconstructed and ground-truth images. Higher is better (max 1.0).

2 Stereo Vision Evaluation:

- **Bad Pixel Percentage (D1):** The percentage of pixels whose estimated disparity error is greater than a threshold (e.g., 3 pixels) from the ground truth. Lower is better.
- **End-Point Error (EPE):** The average absolute difference between the estimated disparity map and the ground truth. Lower is better.

3 Computational Efficiency:

- **Parameters & FLOPs:** The number of network parameters and Floating-Point Operations measure model complexity.
- **Inference Time:** The time taken to process a single stereo pair, critical for real-time applications.

5. DISCUSSION:

The integration of stereo vision and super-resolution presents a classic trade-off between performance and computational efficiency. While sequential methods are practical, the current research momentum is firmly behind end-to-end joint learning models due to their superior performance.

1. **The Role of Attention**: The most successful recent models, like PASSRnet [10], leverage attention mechanisms (e.g., parallax attention, non-local attention) to effectively fuse information from the stereo pair, which is crucial for maintaining geometric consistency.
2. **Data Dependency**: A significant challenge is the lack of large-scale datasets containing paired LR-HR stereo images with accurate ground-truth disparity maps. This often necessitates training on synthetically down-sampled data, which may not fully represent real-world sensor degradation.
3. **Beyond 2x/4x SR**: Most research focuses on fixed integer upscaling factors (2x, 4x). Real-world applications may require arbitrary-scale SR, which remains an open problem in the stereo context.
4. **Real-World Applicability**: Future work needs to address more complex degradation models, including motion blur, noise, and compression artifacts, rather than just bicubic downsampling.

6. CONCLUSION:

This survey has provided a structured overview of recent advancements in combining stereo vision and super-resolution. The field has progressed from simple sequential processing to sophisticated, end-to-end deep learning models that jointly optimize for both high-resolution imagery and accurate depth estimation. The key to success in these joint models lies in their ability to enforce geometric consistency between the super-resolved views through mechanisms like attention and iterative refinement. While significant progress has been made, challenges remain, particularly in the areas of computational efficiency, handling real-world degradations, and arbitrary-scale super-resolution. The future of this field likely lies in more efficient architectures, self-supervised or unsupervised learning paradigms to mitigate data scarcity, and a tighter integration with other vision tasks like semantic segmentation. The synergy between stereo vision and super-resolution will continue to be a critical enabler for high-precision 3D perception systems in autonomous and robotic applications.

REFERENCES:

1. Scharstein, D., & Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision, *47*(1–3), 7–42.
2. Hirschmuller, H., & Scharstein, D. (2009). Evaluation of stereo matching costs on images with radiometric differences. IEEE Transactions on Pattern Analysis and Machine Intelligence, *31*(9), 1582–1599.
3. Dong, C., Loy, C. C., He, K., & Tang, X. (2014). Learning a deep convolutional network for image super-resolution. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13* (pp. 184–199). Springer International Publishing.
4. Farsiu, S., Robinson, M. D., Elad, M., & Milanfar, P. (2004). Fast and robust multiframe super resolution. IEEE Transactions on Image Processing, *13*(10), 1327–1344.
5. Zhang, L., & Tam, W. J. (2005). Simultaneous depth estimation and image reconstruction from a stereo pair. In Proceedings of the 2005 International Conference on Image Processing (ICIP) (Vol. 3, pp. III-297). IEEE.
6. Lim, B., Son, S., Kim, H., Nah, S., & Mu Lee, K. (2017). Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (pp. 136-144).
7. Kendall, A., Martirosyan, H., Dasgupta, S., Henry, P., Kennedy, R., Bachrach, A., & Bry, A. (2017). End-to-end learning of geometry and context for deep stereo regression. In Proceedings of the IEEE International Conference on Computer Vision (ICCV) (pp. 66-75).
8. Chang, J. R., & Chen, Y. S. (2018). Pyramid stereo matching network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 5410-5418).
9. Jeon, H. G., Lee, K. M., Choi, S., & Kim, J. (2018). StereoSR: A stereo image super-resolution network using a parallax-attention mechanism. IEEE Transactions on Pattern Analysis and Machine Intelligence, *43*(10), 3633–3644.
10. Wang, L., Wang, Y., Liang, Z., Lin, Z., Yang, J., An, W., & Guo, Y. (2021). Learning parallax attention for stereo image super-resolution. IEEE Transactions on Image Processing, *30*, 6351–6363.
11. Wang, Y., Wang, L., Yang, J., An, W., & Guo, Y. (2020). CfNet: Cascade and fused cost volume for robust stereo matching. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 13906-13915).
12. Hui, T. W., Loy, C. C., & Tang, X. (2016). Depth map super-resolution by deep multi-scale guidance. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part III 14 (pp. 353-369). Springer International Publishing.
13. Kim, J., Lee, J. K., & Lee, K. M. (2016). Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 1646-1654).