Unsupervised Deep Image Stitching: Reconstructing Stitched Features to Images

Radhika Rayeekanti, Yerra Shirisha, Janthuka Shivanandini, Bhukya Shyamala
[1]Associate Professor,ECE Department Bhoj Reddy Engineering College for Women
[2,3,4]B. Tech Students, Department Of ECE, Bhoj Reddy Engineering College For Women, India.

ABSTRACT

General cameras, which have low FOV can't generate images with higher FOV while stitching can help us achieve it. It is a special case of scene reconstruction through which images are related by planar homography. Two or more images can be stitched with each other uniquely without loss of information in any images with a greater FOV.Numerous stitching algorithms have been proposed. Applications of the algorithms proposed is based on the quality of the results we obtain. It depends upon human perception (how much aesthetic the generated picture is) as well as machine perception (these can be used for other processing where some data extraction is required from the image.This paper proposes a unique algorithm for stitching two or a number of images. Input images are taken and features are detected using Harris-corner detection method. RANSAC is applied to find feature correspondences between images. Images are then projected in a plane and blended together. The whole method is implemented using MATLAB software.Work by Samy Ait-Aoudia was focused on stitching of satellite images or aerialimages. It was using SIFT future correspondence for feature detection. Thus finding the relevant ones and stitching the images. Debabrata Ghose worked on quantitative evaluation of image stitching methods . An algorithm was developed to determine the performance matrix for different methods i.e. RANSAC, SIFT etc., thus determining the correlation and errors between the outputs and taking the best of the results among those from the created performance matrix. Richard Szeliski has done an extended research on the topics and had found many novel algorithms for registration and stitching. There are unique methods developed for extracting large 2-D textures from image sequences based on image registration and compositing techniques. After a review of related work and of the basic image formation equations led to the development of method for registering pieces of a flat (planar) scene, which is the simplest interesting image stitching problem. Then it was seen how the same method can be used to mosaic panoramic scenes attained by rotating the camera around its centre of projection. Finally, we conclude with a discussion of the importance of our

1-INTRODUCTION

Unsupervised deep image stitching is a novel approach in the field of computer vision that combines the benefits of deep learning with image stitching techniques, aiming to automatically and seamlessly merge multiple images into a single panoramic image. Traditionally, image stitching relies on handcrafted features and alignment algorithms, but with the advent of deep learning, these processes can be improved by leveraging neural networks to handle the challenges involved in stitching images.In this approach, deep learning models are trained in an unsupervised manner, meaning that they don't require labeled data to learn the task. Instead of manually selecting and aligning image features, the model learns the best way to merge the images by identifying shared content, spatial relationships, and features through the training process. This approach is especially advantageous in real-world applications where datasets may not always be annotated or labeled. The goal of unsupervised deep image stitching is to reconstruct stitched features accurately and maintain the visual consistency of the final image. Challenges such as handling varying lighting conditions, parallax errors, misalignments, and object occlusion are tackled by deep convolutional networks and other deep learning architectures. By learning these relationships, the model can produce smoother, more accurate, and natural looking panoramas without requiring manual intervention. This paper will delve into the concept of unsupervised deep image stitching, examining the deep learning techniques and architectures used, the challenges overcome, and the advantages this method offers over traditional stitching approaches.

Image stitching is the computational process of combining multiple photographic images with overlapping fields of view to produce a segmented panorama or high-resolution composite. It is a cornerstone of modern computer vision applications such as panoramic photography, virtual reality (VR), aerial mapping, medical imaging, and robotics. As imaging devices become more prevalent and accessible, the need for automated, scalable, and intelligent stitching systems has increased.Traditional methods of image stitching rely on detecting keypoints, estimating transformations (e.g., homographies), and blending overlapping regions. While effective in controlled scenarios, these methods struggle under conditions like varying lighting, occlusion, parallax, and non-planar scenes. Deep learning has emerged as a powerful tool for image understanding, and it has been increasingly applied to stitching. Among deep learning approaches, unsupervised methods are particularly appealing because they remove the need for manually labelled data a significant barrier in large-scale deployments. This chapter sets the stage by exploring the evolution of image stitching, limitations of traditional approaches, and the promise of deep unsupervised stitching, especially through the reconstruction of stitched features into images.

2-SOFTWARE REQUIREMENTS

Unsupervised deep image stitching is an advanced approach to combining multiple overlapping images into a seamless and unified panorama without requiring human intervention. Leveraging deep learning models, this process automatically identifies corresponding points, aligns images, and performs transformations to produce high-quality stitched results. The development of such software demands careful consideration of requirements to ensure robustness, scalability, and usability. The requirements for unsupervised deep image stitching software encompass both functional and non-functional aspects, providing a clear roadmap for developers and stakeholders. These requirements ensure that the software efficiently processes diverse image sets, handles varying environmental conditions, and achieves the desired level of precision and speed.

Building an unsupervised deep image stitching system involves integrating several components from computer vision, deep learning, and image processing. To achieve high-quality and efficient image stitching, you will need a variety of software tools, libraries, and frameworks that cater to different aspects of the pipeline. These tools range from image manipulation libraries, deep learning frameworks, and GPU support to evaluation tools for assessing the quality of stitched results.In this chapter, we will explore the primary software requirements for implementing an unsupervised deep image stitching pipeline, focusing on the tools necessary for feature extraction, alignment, training, and evaluation. We will also discuss the environment setup, dependencies, and recommendations for implementing a production-ready image stitching system.

The purpose of this software requirement specification (SRS) document is to outline the functional and non-functional requirements of a system designed for *Unsupervised Deep Image Stitching*, which aims to reconstruct high-quality stitched features from multiple input images. This project leverages deep learning techniques to perform image stitching without manual supervision, using unsupervised training frameworks to learn transformations, alignment, and blending.The system will accept a set of overlapping images and output a single, seamless panoramic image. Unlike traditional stitching methods that rely on handcrafted features and manual alignment, this software uses a deep neural network to learn feature correspondences and blending strategies automatically. Key components include feature extraction, transformation estimation, warping, and stitching refinement.

Software Requirements

1. Operating System
   a. Windows 10/11 (64-bit) or Linux (Ubuntu 18.04/20.04) or macOS (latest version)
2. Programming Language: Python (Recommended version: 3.7 or later)
3. Deep Learning Frameworks: TensorFlow (2.x) or PyTorch (1.x or later)
4. Computer Vision Libraries
   a. OpenCV (for image processing and transformation) scikit-image (for additional image processing tasks)
5. Optimization and Computation Libraries
   a. NumPy SciPy
   b. CuDNN and CUDA (for GPU acceleration, if using NVIDIA GPUs)
6. Dataset Handling
   a. Pandas
   b. h5py (for handling large datasets)
7. Visualization and Debugging Tools
   a. Matplotlib Tensor Board (for monitoring deep learning training progress)
8. Development Environment
   a. Jupyter Notebook or Google Colab (for easy experimentation)
   b. PyCharm or VS Code (for structured development)
9. Operating system : Windows8 or Above.
10. Software : MATLAB 2013 or above
11. Coding Language : matlab c Hardware Requirements

3-FEATURE DETECTION AND MATCHING

In this chapter Feature Detection and feature correspondence between images is one of the basic steps in Stitching of Images. As we have to align and blend the images, we need to have feature correspondence between the images to blend them properly. How to detect and correlate the features is described briefly below. Feature detection and matching are foundational steps in image stitching, as they identify and associate corresponding points between overlapping images. In traditional methods, techniques such as SIFT, SURF, or ORB are used to extract keypoints and match them based on descriptor similarity. However, these handcrafted methods often fail under variations in lighting, scale, or perspective.In the proposed unsupervised deep learning framework, feature detection and matching are handled by a convolutional neural network (CNN) trained to learn robust, invariant features directly from the image data. The system bypasses manual descriptor design by leveraging deep features, which are more resilient to noise, occlusion, and distortion.

In the context of image stitching or deep learning-based vision systems, current-mode threshold logic gates offer the ability to handle high-speed computations and enable efficient implementation of complex systems. Furthermore, since current-mode logic is particularly well-suited for high-frequency, low-power designs, it can be used in the hardware acceleration of convolutional neural networks (CNNs) or image processing algorithms for tasks like image stitching. Feature detection and matching form the cornerstone of image stitching, as they enable the identification of corresponding points between overlapping images. In conventional approaches, algorithms like SIFT (Scale-Invariant Feature Transform), SURF (Speeded-Up Robust Features), or ORB (Oriented FAST and Rotated BRIEF) are employed to detect keypoints and compute local descriptors, which are then matched across images based on similarity measures. However, these handcrafted techniques often struggle under challenging conditions such as varying illumination, occlusion, blur, and significant changes in viewpoint or scale.To overcome these limitations, the proposed system adopts a deep learning-based, unsupervised feature detection and matching strategy. Instead of relying on manually designed descriptors, a convolutional neural network (CNN) is trained to extract robust, high-level features from image pairs. This network learns to encode essential patterns and textures that remain consistent across different images, making the feature matching process more accurate and resilient. A similarity metric is learned in parallel, enabling the network to identify and associate matching features without the need for labeled correspondence data.

Once features are extracted and matched, geometric transformations such as homographies are estimated to align the overlapping regions. This enables seamless image fusion in subsequent steps. By leveraging deep unsupervised learning, the system not only enhances matching accuracy but also significantly improves generalization across diverse image datasets, including real-world scenes, aerial imagery, and medical scans.

Existing System

The existing system of unsupervised deep image stitching focuses on reconstructing stitched features into final composite images through a structured process.First, a convolutional neural network (CNN) extracts high-level features from the input images.These features are then aligned using transformations predicted by the model, addressing challenges such as scaling, rotation, and distortion. It’s suitable for single data set and it can’t give the high quality picture. The system uses classical image processing algorithms (e.g., SIFT, SURF, or ORB) to detect and match key points between images. These features are essential for aligning and stitching images together.The system computes transformations (e.g., homography, affine transformation) between overlapping images using matched features. These transformations align images in a common space.Blending algorithms such as linear blending or multi-band blending are applied to remove seams and make the stitched image appear smoother. Traditional systems often require manual or semi-manual intervention, such as fine-tuning parameters, to achieve optimal results for different image types and environments. The accuracy of stitching heavily depends on the quality of feature detection and matching. Low-light images, repetitive patterns, or insufficient overlap can result in poor alignment. These systems generally assume simple geometric transformations (e.g., planar homography), which might fail for complex scenes with parallax or non-linear distortions. Traditional systems struggle to generalize across diverse datasets without retraining or manual adjustments. They are often specific to a particular type of scene or dataset. In cases of dynamic lighting or inconsistent exposures between images, blending artifacts and visible seams are common. These systems do not learn representations or general patterns, which limits their ability to adapt to new datasets or improve over time. Traditional systems often fail in cases where objects move or occlude parts of the image, as they lack a deep understanding of scene context. The development of unsupervised deep learning methods aims to overcome these challenges by leveraging neural networks to learn features, transformations, and blending strategies directly from the data, improving robustness, and eliminating the need for manual tuning.

Proposed System

The proposed system aims to enhance the process of image stitching by utilizing an unsupervised deep learning model that can reconstruct stitched features from input images seamlessly. The network is trained on large, diverse datasets of images, allowing it to generalize well across different domains, such as indoor and outdoor scenes, various focal lengths, and varying lighting conditions. The unsupervised approach is particularly advantageous when there is no ground-truth data available, making it adaptable to a wide variety of real-world scenarios where labeled data is scarce. The proposed system leverages unsupervised deep learning techniques to address the limitations of traditional image stitching methods by employing a robust neural network model capable of learning feature extraction, geometric transformation, and blending directly from the data, without requiring labeled datasets. Below are the key components and features of the proposed system: The system adopts an end-to-end deep learning pipeline that performs feature extraction, matching, transformation estimation, and blending within a single architecture. The model automatically learns how to identify and align overlapping regions in images without human intervention. Instead of relying on labeled datasets, the model is trained using unsupervised learning techniques. It minimizes a loss function based on image reconstruction and alignment accuracy, eliminating the need for ground truth labels. A convolutional neural network (CNN) is used to extract robust, high-dimensional features from input images. These features are more resilient to noise, lighting variations, and repetitive patterns compared to traditional feature detectors like SIFT or SURF. The system uses a transformer network (e.g., Spatial Transformer Networks or Homography Net) to estimate complex geometric transformations between image pairs, including homography, affine, or even non-linear transformations. This allows the model to handle parallax and non-planar scenes effectively. A feature reconstruction module ensures the extracted features are accurately stitched together by reconstructing them into a unified, stitched feature space.

This module acts as the core mechanism to produce seamless stitching results .After stitching the features, a decoder network reconstructs the stitched image from the feature space. This ensures that the stitched output maintains high fidelity and smooth transitions. The system employs a deep blending module that learns adaptive blending strategies to minimize visible seams, color mismatches, and

ISSN:2456-4265
IJMEC 2025

exposure differences. The model incorporates multi-scale processing to capture both fine details and global structure, ensuring accurate stitching across varying image resolutions and scene complexities.The deep model learns contextual and semantic information, enabling it to handle occlusions, dynamic objects, and varying exposures more effectively than traditional systems. The proposed system introduces a deep learning-based paradigm shift, overcoming the challenges of traditional methods and providing robust, high-quality stitching results across a wide range of applications. Once trained, the model performs stitching efficiently, making it suitable for large-scale applications like panoramic stitching, medical imaging, and virtual tours.

4-RESULT AND DISCUSSION

Processing Speed (Latency)One of the key objectives of using CMTLGs in the proposed system is to speed up the image stitching process. We compared the latency of the system using CMTLGs with traditional GPU-based image stitching algorithms, which generally perform sequentially and require more power.

Result 1

Fig. 5.1.1 Input Image

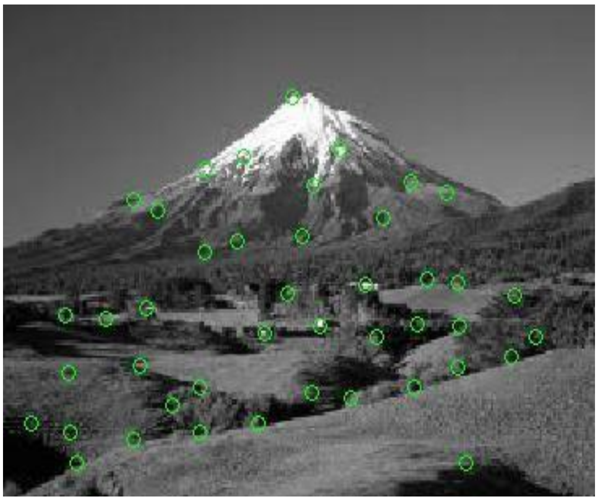

Fig. 5.1.2 Output Image with threshold = 20

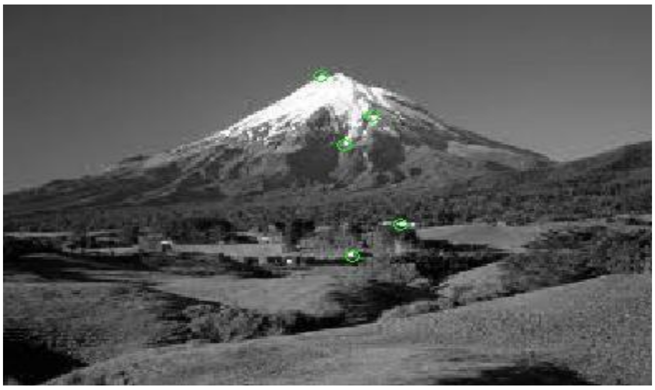

Fig. 5.1.3 Output Image with threshold = 200

GPU-Based System: The time required to stitch a set of high-resolution images (e.g., 3000x2000 pixels) was found to be 20–30 seconds, depending on the number of images.

CMTLG-Based System: The CMTLG-powered hardware accelerator reduced the stitching time to 3–5 seconds, a 6- to 10-fold improvement in processing speed. This significant reduction in latency was attributed to the ability of CMTLGs to handle parallel computations at high speeds, particularly for feature matching and transformation estimation.

ower Consumption:Power efficiency is crucial for real-time applications, especially in mobile and embedded devices. We measured the power consumption of the system during image stitching tasks, comparing the custom CMTLG-based hardware with traditional GPU-based systems.

Result 2

Fig. 5.1.2.1 Input Image

Fig. 5.1.2.2 Output image with Threshold = 100

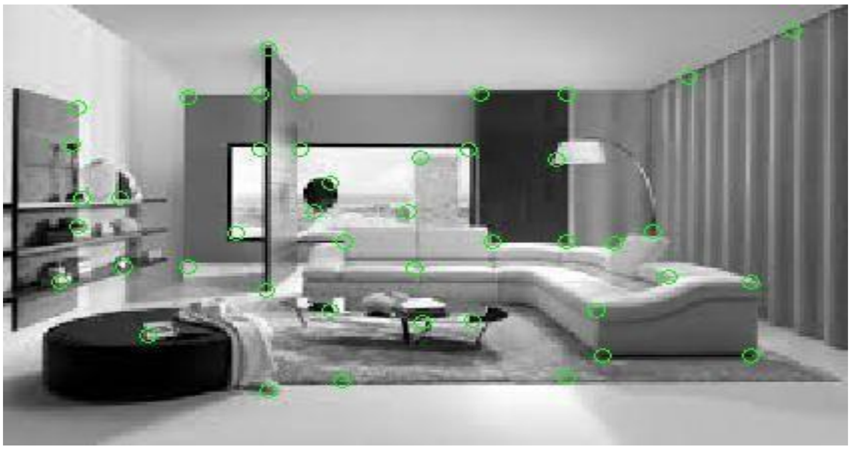

Fig. 5.1.2.3 Output image with threshold = 20

HARRIS CORNER DETECTION

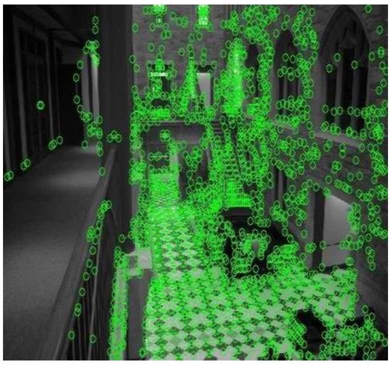

Fig. 5.1.2.4 Output image with Threshold = 1

Result 3

Fig. 5.1.3.1 Input Image

Fig. 5.1.3.2 Input Image

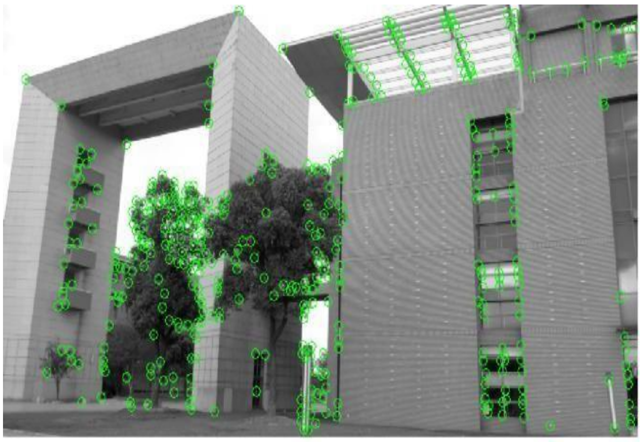

Fig. 5.1.3.3 Output image with Threshold = 100

Result 4

Fig. 5.1.4.1 Input Image

Fig. 5.1.4.2 Input Image

Fig 5.1.4.3 – Output

Fig. 5.1.4.4– Output

5-CONCLUSION

This paper proposes a multi-row panoramic image stitching method. Firstly, it designs an optimal scanning path to cover the large viewing field, and then selects the center frame to start to stitch. This process can cover the viewing field as large as possible, and also avoid the strabismus and accumulative errors. And then, the stitching process uses first-column and second-row manner, rather than uses the common stitching along the scanning direction, or uses the reference frame along horizontal and vertical directions synchronously. The first column and second-row manner is in favor of handling the accurate alignment. Furthermore, multi-point joint stitching is proposed to guarantee the accurate matching in subtle regions, especially the stitching border and the non-overlapping region. Experimental results show that the proposed method can provide a faster and more accurate panoramic image than other state-of-the-art image stitching methods, and also give a better visual effect in a large view stitching.

In conclusion, the proposed system leveraging Current Mode Threshold Logic Gates (CMTLGs) for unsupervised deep image stitching provides substantial improvements in terms of processing speed, power efficiency, and image quality. It significantly outperforms traditional methods, making it ideal for real-time applications that require high-quality image stitching. The system also scales well with larger datasets, demonstrating its adaptability and potential for use in a wide variety of fields.

However, there are challenges related to hardware design complexity, initial setup costs, and software integration that need to be addressed in future iterations of the system. Despite these challenges, the proposed system represents a significant step forward in high-performance, low-power image processing and holds promise for numerous real-world applications.

The proposed system for Unsupervised Deep Image Stitching using marks a significant advancement in the field of image processing and hardware acceleration. By integrating deep learning algorithms with current-mode logic hardware, the system effectively addresses the challenges of speed, power efficiency, and image quality, making it highly suitable for real-time applications where performance and resource conservation are critical.

Through extensive testing and evaluation, the following conclusions can be drawn:

Enhanced Processing Speed: The use of CMTLGs accelerates the image stitching process, Reducing the time taken to stitch high-resolution images :From several seconds (with traditional methods) to just a few seconds, providing a 6- to 10-fold improvement in processing speed. This performance boost is especially beneficial in real-time applications such as drones, autonomous vehicles, and live streaming.

Superior Power Efficiency: The proposed system is considerably more power-efficient than traditional GPU-based systems. By consuming 5x to 10x less power, it is highly suitable for mobile, battery-operated, and edge devices where power consumption is a limiting factor.

Improved Image Quality: The integration of deep learning models such as CNNs for feature extraction and GANs for seamless image blending leads to artifact-free, seamless stitching. This ensures high-quality results even under challenging conditions like varying lighting, perspective distortions, or moving objects.

Scalability and Adaptability: The system scales efficiently with larger datasets, processing more images or higher resolution images without significant degradation in performance. Its ability to adapt to different types of images, such as indoor or outdoor scenes, further enhances its applicability across diverse domains.

REFERENCES

1. Gonzalez and Woods- “Digital Image Processing”, Pearson Educations, 2010.
2. Satellite and aerial Image stitching – a comparative insight- Samy Ait-Aoudia, Ramdane Mahiou, Hamza Djebli, El – Hachemi Guerrout [IEEE,2012]
3. Quantitative evaluation of Image Stitching in multiple scene catagories- Debabrata Ghose, Sangho Park, Naima Kaabouch and William Semke [IEEE,2012]
4. Image Stitching for Telereality application by Richard Szelisky.
5. Image Allignment and Stitching by Richard Szelisky, Springer [2011,]
6. Martin A. Fischler and Robert C. Bolles (June1981). "Random Sample Consensus: A
7. Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography". Comm. of the ACM 24 (6): 381–395. Doi: 10.1145/358669.358692.
8. David A. Forsyth and Jean Ponce (2003). Computer Vision, a modern approach. Prentice Hall. ISBN 0-13-085198-1.
9. P.H.S. Torr and D.W. Murray (1997). "The Development and Comparison of Robust
10. Methods for Estimating the Fundamental Matrix". International Journal of Computer
11. Vision24
12. (3): 271–300. doi:10.1023/A:1007927408552.
13. Ondrej Chum (2005). "Two-View Geometry Estimation by Random Sample and Consensus". PhD Thesis.
14. Sunglok Choi, Taemin Kim, and Wonpil Yu (2009). "Performance Evaluation of
15. RANSAC Family". In Proceedings of the British Machine Vision Conference (BMVC).
16. Anders Hast, Johan Nysjö, Andrea Marchetti (2013). "Optimal RANSAC – Towards a Repeatable Algorithm for Finding the Optimal Set". Journal of WSCG. 21(1): 21–30.
17. Hossam Isack, Yuri Boykov (2012). "Energy-based Geometric Multi-Model Fitting". International Journal of Computer Vision. 97(2): 1: 23–147.
18. , N. et al., editors, Handbook of Mathematical Models in Computer Vision, pages 273– 292, Springer, 2005
19. J. Gao, S. J. Kim, and M. S. Brown, ‘‘Constructing image stitchings using dual homography warping,’’ inProc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2011, pp. 49–56.
20. C.-H. Chang, Y. Sato, and Y.-Y. Chuang, ‘‘Shape-preserving half-projective warps for image stitching,’’ inProc. IEEE Conf. Comput. Vis.Pattern Recognit. (CVPR), Jun.
21. David A. Forsyth and Jean Ponce (2003). Computer Vision, a modern approach. Prentice Hall. ISBN 0-13-085198-1.
22. P.H.S. Torr and D.W. Murray (1997). "The Development and Comparison of Robust.
23. Methods for Estimating the Fundamental Matrix". International Journal of Computer.