Conclusions

In the present work, an attempt is made to overcome few limitations of image registration methods for Augmented Reality (AR) systems in order to achieve accurate view alignment between the virtual and real objects in the view of real scene while achieving low computational complexity of the method. To attain this objective, firstly, performance of six existing and widely used feature detectors (Harris-Affine, Hessain-Affine, MSER, Scale Invariant Feature Transform (SIFT), Affine-SIFT (ASIFT) and Speeded Up Robust Features (SURF)) is compared in terms of number of detected keypoints in an image and number of correct correspondences found between an image pair under varying image quality and varied imaging conditions like viewpoint change, scale change, image blur, illumination change and JPEG compression. Image quality quantification of the images is done in order to formulate an initial base for the performance evaluation of feature detectors and is done using No-Reference Image Quality Assessment (NR-IQA) metrics and Full-Reference Image Quality Assessment (FR-IQA) metrics. The analysis is also weighed upon the Pearson Coefficient for reasoning the behavior of feature detectors in various cases in accordance with their time complexity and robustness to perform well under affine transformations. Conclusions drawn from this comparative study are as follows:

- The number of keypoints detected in an image and the number of matches found between two images can be associated with the quality of that image (when quantified with respect to spatial and luminous features of image pixels) and the similarity index value between two images respectively.
- Interpreting the applicability of feature detectors in an AR system, it is observed that even though ASIFT is able to detect a noticeable higher number of keypoints in images, the computational complexity of ASIFT makes it unsuitable for performing image registration in AR systems. Similarly, for feature matching, ASIFT outperforms both SIFT and SURF, but the time taken for computation is much more than what could be accepted for AR applications. Also, SURF performs better than SIFT in accuracy and has a lower computational complexity than ASIFT, but processing speed is still not appreciable for a real time AR system.

For understanding image and video features that deteriorates the quality of a given image/video, a fundamental analysis of spatial and temporal artifacts in an image is done and an efficient NR-IQA model is designed for better estimation of distortions in an image by combining the efficiency of three diverse NR-IQA metrics (Naturalness Image Quality Evaluator (NIQE), Blind/Reference-less Image Spatial

Quality Evaluator (BRISQUE) and BLind Image Integrity Notator using Discrete Cosine Transform (DCT) Statistics-II (BLIINDS-II)) using a Multi-Linear Regression (MLR) model. The same analysis is carried out for video features and a No-Reference Video Quality Assessment model (NR-VQA) is designed by defining and examining five feature quantification metrics, namely, ringing, frame difference, blocking, clipping and contrast. A Neural Network (NN) model with one middle hidden layer of 15 units is used for fitting appropriate metrics that quantify these distortions in a video. Conclusions drawn from the work are as follows:

- MLR model for blind (No-Reference) estimation of image quality performs marginally better (0.6%) than BRISQUE, which individually has the best performance among the three metrics chosen for MLR. Even low fraction of samples for training, such as 0.5, provide a consistent accuracy over many different training to testing ratios, thus verifying efficient execution of the proposed model.
- The result of the NN fitting is a goodness of fit R=0.8785, a good accuracy, taking into account the limited number of features available to the model. Therefore, as the dataset used for evaluation suffers from loss of generality of features, the proposed model generality too remains a drawback due to lack of availability of features that could fairly determine effects of artifacts and distortions like ringing, freeze frame, etc.

Based on the observations of the comparative evaluation of six feature detectors, an improved method of image registration is proposed using Maximally Stable Extremal Regions (MSER). Basic considerations of MSER are reanalyzed and enhanced for attaining better stability and affine invariance for feature detection procedures in real-time AR applications. The main contribution of the work is the development of a novel detector–descriptor combination using MSER. MSER detector is implemented in two forms, i.e. Linear-MSER and Multi-Scale Linear-MSER (MSLinear-MSER), for extracting stable regions of interest in an image. SIFT and SURF descriptors are used in combination with the two detectors for analyzing the performance of the methods in terms of time complexity, affine invariant property and accurate correspondences between image pairs. Conclusions drawn from this study are as follows:

- The outcome shows that MSLinear-MSER+SIFT (when MSLinear-MSER is executed with six octaves and five levels per each octave) detector–descriptor combination works more efficiently under varying imaging conditions in terms of more number of matches between image pairs with comparable time complexity to the other methods. To demonstrate the efficiency of MSLinear-MSER+SIFT, an AR system prototype is developed using the same approach.
- The efficiency of the proposed AR system in terms of correct augmentation is justified using precision metric yielding an accuracy of up to 0.9729.

In another attempt to design a more invariant and efficient image registration method, a novel feature description procedure is proposed using circular and elliptical sampling of the neighboring pixels of the extracted keypoint. Circular and elliptical sampling for this purpose is implemented in two variations: 1) using polar coordinates, 2) using curve tracking, to increase the robustness of the proposed descriptor in terms of affine invariance and low computational complexity. The performance comparison of the proposed descriptor is done by SIFT feature descriptor for determining and proving its efficiency. Conclusions drawn from this study are as follows:

- Proposed feature descriptor exhibits more number of matches in an image pair as compared to SIFT whenever the quality of images are differing much. Therefore, the proposed method is invariant to various parameters like viewpoint change, scale change, blur change, illumination change and JPEG compression.
- The proposed descriptor is faster than SIFT by a factor of 1.6 on an average basis while producing sufficient number of matches between an image pair to perform image registration in an AR system.

Thus, in the present research, a wide analysis of image registration methods is done for determining their applicability in AR systems. From analyzing the behavior of image registration methods towards different quality of images captured under varying imaging conditions to proposing new methods, the work focuses on improving the overall efficiency of the image registration procedure for an AR system to achieve correct alignment between real and virtual objects with reduced time complexity. Improvements proposed in this work achieve both the accuracies in terms of affine invariance property of image registration methods and low computational complexity.