Computer vision can be utilized as a complementary to conventional sensor-based structural health monitoring (SHM). Such an approach can eliminate some of the challenges in practical and effective applications of SHM to real life structures. For example, practical SHM of highway bridges requires low cost, quick installation for monitoring that has ideally minimal interruptions to operations. Sensor and computer vision-based hybrid monitoring or fully computer vision based monitoring can be a viable solution for real life structural health monitoring and identification, where inputs and responses of the structure are to be captured. In this study, non-contact identification of structures is investigated particularly for bridge structures. Unlike most of the vision-based research studies, which focus either on structural input (vehicle location) estimation or on structural output (structural displacement and strain responses) estimation, the proposed framework in this dissertation combined the vision-based structural input and the structural output from non-contact sensors to overcomes the limitations given above. First, a series of computer vision-based displacement measurement methods are developed for structural response (structural output) monitoring which can be utilized on different infrastructures such as grandstands, stadiums, towers, footbridges, small/medium span concrete bridges, railway bridges, and long span bridges subjected to loading cases such as human crowd, pedestrians, wind, vehicle, etc. By means of the methods and approaches investigated, structural behaviors, modal properties, load carrying capacities and performances are investigated using vision-based methods and verified by comparing with conventional approaches. In addition, this study also investigated the serviceability status of structures by using computer vision-based methods. Then issues and considerations of computer vision-based measurement in field application are discussed and recommendations are provided for better results. Also, to overcome the difficulties of vision-based monitoring under adverse environmental factors such as fog and illumination change, this study proposed a robust vision-based method for displacement measurement using spatio-temporal context learning and Taylor approximation. A critical parameter, load distribution, that is commonly used in bridge engineering practice for load rating was also estimated by using visual tracking. By combining the structural input and output results, the unit influence lines (UIL) of structures were extracted during daily traffic just using cameras. As such, the external loads can be estimated by using just cameras and extracted UIL. The structural condition assessment is accomplished with the vision-based structural input and output with a normalized response irrespective of the type and/or load configurations of the vehicles or human loads.