

(54) Title of the invention : MULTI-LINGUAL TEXT EXTRACTION SYSTEM

(51) International classification	:G06K0009620000, G06F0016330000, G06F0016280000, G06F0017280000, G06K0009030000	(71)Name of Applicant : 1)CGC Technical Campus Jhanjeri Address of Applicant :State Highway 12A Jhanjeri, Sahibzada Ajit Singh Nagar, Punjab 140307, India. Punjab India
(31) Priority Document No	:NA	(72)Name of Inventor : 1)Gangandeep Singh
(32) Priority Date	:NA	
(33) Name of priority country	:NA	
(86) International Application No	:NA	
Filing Date	:NA	
(87) International Publication No	: NA	
(61) Patent of Addition to Application Number	:NA	
Filing Date	:NA	
(62) Divisional to Application Number	:NA	
Filing Date	:NA	

(57) Abstract :

A multi-lingual text extraction system comprise of an image capturing unit for acquiring image(s) (document images, textual images, non-textual images) from a user, a pre-processor for detecting non-textual and textual images and eliminating non-text area from the textual image(s), a feature extraction module for mining out bilingual textual features by multiple classifiers (such as ImageNet2012, AlexNet, VGGNet, and GoogleNet), where the features are clubbed together in multiple feature vectors, a computing unit for decreasing dimension of the vectors, a feature fusion module that ranks the vectors, a storage unit that stores the vectors in an organized format for easy retrieval, a data analysis module for calculating a similarity matrix in order to find multi-lingual textual content from the image(s).

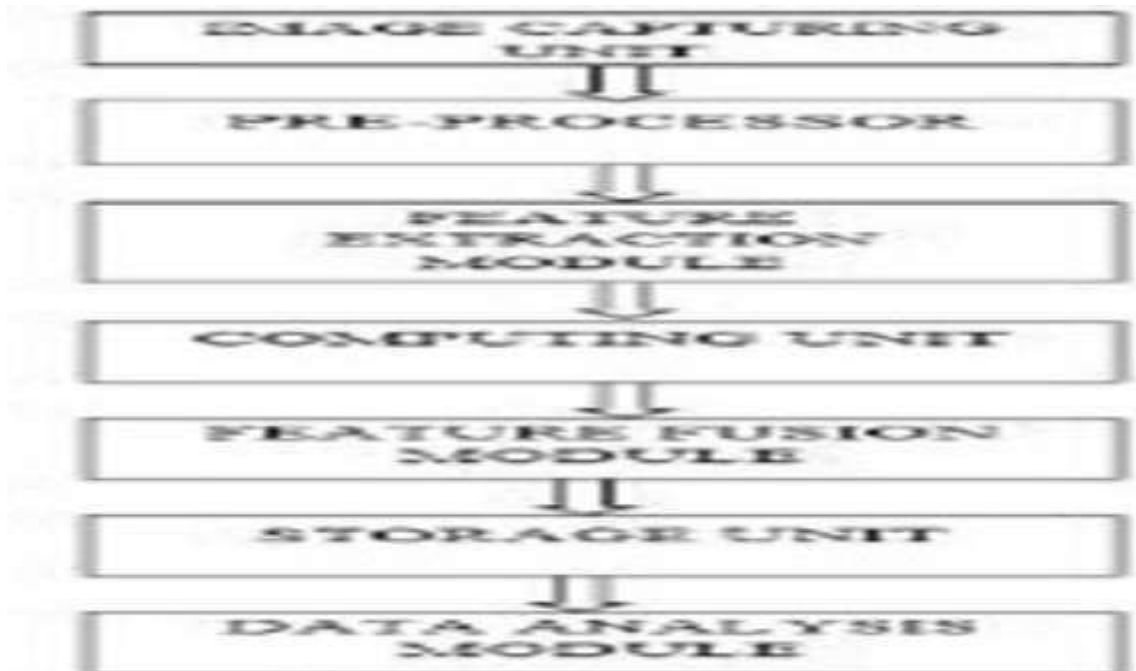


Figure 1