SCIENCE CHINA
Technological Sciences

• News & Views •

December 2015 Vol.58 No.12: 2210–2211
doi: 10.1007/s11431-015-5991-0

# New advances in visual computing for intelligent processing of visual media and augmented reality

WANG JuHong[1], ZHANG SongHai[1*] & MARTIN Ralph R.[2]

[1] *Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China;*
[2] *School of Computer Science & Informatics, Cardiff University, 5 the Parade, Roath Cardiff, Wales CF24 3AA, UK*

Visual imagery constitutes the most important sensory information for humans. The entire field of acquiring, analyzing and synthesizing visual data by means of computers is called visual computing. It has an extraordinarily wide range of applications, including for example: industrial quality control, street view and driver assistance systems, robot navigation, multimedia systems, and computer games. Visual computing comprises four key areas: computer vision and image processing, computer graphics, virtual and augmented reality, and visualization. It requires deep, interdisciplinary scientific knowledge, in particular in computer science, mathematics, physics, engineering, and cognitive sciences. It tackles high level tasks such as editing and composition of visual content [1] and recognition of sematic content [2–5], as well as basic low level problems such as denoising [6,7] and decomposing [8,9] images, video and 3D shapes.

Denosing is often a first step to provide high quality inputs to more complicated tasks such as panoramic image stitching to construct a street view database, and image understanding in computer vision. The object of image denoising is to reduce the noise level, while preserving edges and textures as much as possible. The necessary processing may be done in the image domain, or frequency domain using FFT or wavelets. Recent studies have considered how to represent contour information in images using ridgelets, curvelets and contourlets, as well as finding suitable threshold schemes to remove noise [6]. There is also corresponding work on mesh denosing, which finds structures in terms of positional and normal features [7].

Image decomposition concerns the splitting of an image into two or more components. A fundamental goal in image analysis and computer vision is to extract meaningful components from an image, for tasks such as scene understanding, generation of visual media and many intelligent applications of visual content. Various image decomposition strategies exist. One approach is cartoon and texture decomposition: ref. [9] gave a good survey on the existing decomposition models and extended the nonlinear filter method to decompose an image into three components: the cartoon component, i.e. the main geometric structures, the oscillatory component, or texture, and noise. An image can also be decomposed into a lighting image and a reflectance image known as the intrinsic image. Automatic intrinsic image decomposition remains a significant challenge, particularly for real-world scenes. Recent advances to this longstanding problem are data-driven methods based on large scale datasets of ground truth data. For example, ref. [8] built a dataset of more than 5000 labeled intrinsic images as a basis for a decomposition algorithm, as demonstrated in Figure 1.

Computer understanding of scenes or objects within them is a fundamental problem in computer vision, and is crucial to intelligent applications based on visual content. For example, efficient recognition of objects and scene understanding of a live video captured by cameras on a car allows a driver assistance system to instantly feed appropriate information to the driver, or to the cars controls directly.


**Figure 1** (Color online) Example of image decomposition. (a) Input image; (b) reflectance; (c) shading.

*Corresponding author (email: shz@tsinghua.edu.cn)

Recognition of objects in visual data has long been a popular topic. Both object (including human individual) recognition, and behavior recognition, are key tasks, with widespread applications to areas such as visual surveillance, advanced user interfaces, video meetings, and identification for security purposes. Recent work focuses on extending the traditional bag-of-word models to improve recognition performance even when the features are extremely noisy [5]. In order to endow computers with the ability to recognise and understand scenes, it has become a trend to include the findings of research on the cognitive mechanisms of human vision into computational cognitive modes for recognition and understanding of scenes. For example, the authors of ref. [3] proposed a distributed computational cognitive model for general object recognition, and gave a systematic review of psychological and neurophysiological studies which provide collective evidence for a distributed representation of 3D objects in the human brain. They also presented a computational model simulating the distributed mechanisms of the object vision pathway. Experimental results show that the resulting computational cognitive model outperformed five representative earlier 3D object recognition algorithms. Location recognition is another task of current interest due to the popularity of location based services (LBS) in recent years. For example, ref. [2] presents interesting work that places are recognized in real time from photos taken by mobile devices; it is based on fast geometric image matching and a RANSAC procedure.

Another main goal of visual computing is to provide tools to edit and composite visual data to produce new visual imagery meeting users' requirements. Typical examples of recent work consider image completion [1], as illustrated in Figure 2. As well as 2D images, 3D shapes can also be understood and edited. Augmented reality provides users with mixed imagery by compositing virtual content with a captured image, which can be used in diverse areas such as gaming, and giving instructions to engineers performing maintenance. Recent work has considered such issues as efficient image descriptions to provide real-time performance when searching a large database to find optimal virtual content for a mobile device, and truthful color reproduction of virtual content to achieve coherent and immersive imagery [10]. 2D images and 3D models both can represent the physical world, and joint processing of images and 3D model collections can exploit the strength of each data modality to improve tasks in the other, such as in the recently proposed CROSSLINK system [4], which significantly improves the quality of text based 3D model searching by using side information from an image database.

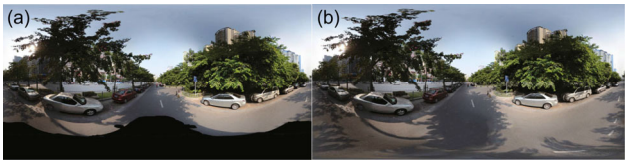


**Figure 2** (Color online) Example of image completion. (a) Input image with missing data at the bottom; (b) completion result.

By considering recent research, we have observed that several trends have emerged in research into visual computing.

(1) The joint analysis of multimodal data, such as images (2D) and 3D geometric data, may achieve better performance than those in individual domain.

(2) Based on the availability of 'Big Data', especially on the Internet, more and more intelligent systems are being built based on knowledge acquisition and learning.

(3) Study of human cognitive mechanisms and using them to build computational cognitive model based on human vision is becoming more important in visual computing research.

1 Zhu Z, Martin R R, Hu S M. Panorama completion for street views. Comput Visual Media, 2015, 1: 49–57
2 Gong M Y, Sun L F, Yang S Q, et al. Find where you are: A new try in place recognition. Visual Comput, 2013, 29: 1211–1220
3 Liu Y J, Fu Q F, Liu Y,et al. A distributed computational cognitive model for object recognition. Sci China Inf Sci, 2013, 56: 092101
4 Hueting M, Ovsjanikov M, Mitra N. CrossLink: Joint understanding of image and 3D model collections through shape and camera pose variations. ACM SIGGRAPH Asia 2015
5 Li H P, Zhang F, Zhang S W. Multi-feature hierarchical topic models for human behavior recognition. Sci China Inf Sci, 2014, 57: 092107
6 Shen X H, Wang K, Guo Q. Local thresholding with adaptive window shrinkage in the contourlet domain for image denoising. Sci China Inf Sci, 2013, 56: 092107
7 FanH Q, Peng Q S, Yu Y Z. A robust high-resolution details preserving denoising algorithm for meshes. Sci China Inf Sci, 2013, 56: 092104
8 Bell S, Bala K, Snavely N. Intrinsic images in the wild. ACM T Graphic, 2014, 33: 159
9 Xu J L, Feng X C, Hao Y, et al. Adaptive variational models for image decomposition. Sci China Inf Sci, 2014, 57: 028102
10 Menk C, Koch R. Truthful color reproduction in spatial augmented reality applications. IEEE T Vis Comput Gr, 2013, 19: 236–248