摘　要

立体视觉注意是人类视觉在信息处理过程中一个重要阶段，可以让人有效地去处理有意义的信息，自动过滤无意义或较少意义的信息。因为视觉注意的重要性，视觉注意分析得到了很多研究机构的关注，成为计算机视觉和相关领域近年来的研究热点。目前已有的视觉显著计算模型以及对应的评测数据集主要是针对二维图像和视频，随着三维摄影摄像技术的成熟，多媒体应用开始进入了立体影像时代。已有一些立体视觉显著计算模型相继被提出，但是至今仍缺乏一个多种类、多数量的立体图像公开评测数据集，因而很难对不同模型进行定量对比。为了探究 2D 与 3D 图像在眼动的区别，以及加入深度信息后视觉注意的分布变化，本文首先构建了 2D 与 3D 图像眼动对比实验，并且分别针对 2D 图像和 3D 图像提出了视觉显著计算模型。本文具体主要研究内容如下：

第一，本文首先构建了一个包含 1000 张不同种类的立体图像数据集，并且采用眼动仪收集了 20 位被试的眼动数据。同时设计了 3D 图像眼动和 2D 图像眼动的对比实验。实验分析表明，在自然情况下观看 3D 图像时对比 2D 图像具有注视点多，而且每个注视点关注时间更短的特点。从眼动分布对比可以看出，2D 与 3D 两种情况下眼动分布存在一定的差异。本次实验选取了 100 张复杂场景和 100 张简单场景的图像，可以发现不同复杂度场景下眼动存在明显的差异，并且在复杂场景中深度信息对视觉注意分布影响更为明显。

第二，基于生理学、心理学和神经科学的结论，构建了一个多尺度多方向自底向上的二维视觉显著计算模型，该模型主要采用 2D Log-Gabor 滤波金字塔模拟单眼感受野模型，并且加入“中心偏爱”现象的影响，形成一个针对 2D 图像的视觉显著计算模型。基于本文构造的二维图像数据集即相应的眼动数据，和已有立体视觉显著计算模型对比，结果表明本文中提出的二维视觉显著计算模型的有效性。

第三，在二维视觉显著计算模型基础上，提出了一个新的多尺度、多方向、双眼融合、自底向上的立体视觉显著计算模型。首先借鉴视网膜的生理特性，提取颜色特征与亮度特征作为输入，用 2D Log-Gabor 滤波金字塔模拟单眼的感受野模型并且加入左右眼相位差因素，然后用 Two-Stage 模型模拟左右眼刺激与抑制等相互作用进行双眼融合。之后加入“中心偏爱”和“前景偏爱”这两个常见

现象对立体视觉注意的影响，最终获得立体图像的显著区域。采用本文构造的多种类规模化的立体图像数据集以及相应的眼动数据，与已有立体视觉显著计算模型对比，结果证明本文提出的立体视觉显著计算模型在效果上有了很大的改进，能更有效提取立体图像中感兴趣区域。

综上所述，本文针对 2D 与 3D 图像眼动对比、二维和立体视觉显著计算模型三个重要问题进行了研究。本文的研究工作为立体视觉显著计算模型进一步研究奠定了基础。

关键字：Log-Gabor 滤波，双眼融合，中心偏爱和前景偏爱，二维和立体视觉显著计算模型，2D 与 3D 眼动对比实验。

The Comparison Analysis of Visual Saliency Computational Models for 2D and Stereoscopic Images

Min Wang (Computer Applied Technology)
Directed by Tiejun Huang

Stereoscopic visual attention is a very important stage in biological information processing, and it makes people to be more effective to deal with meaningful information, while automatically filter out unimportant information. Inspired by the importance of visual attention, visual attention analysis gets the attention of many research institutions and become hot spot in computer vision and related fields. So far, most of visual saliency computational models and the benchmark data sets are based on the 2D images or videos. With the upsurge of the 3D images and videos, 3D era has come. Although some stereoscopic visual saliency computational models have been proposed, none of the evaluation result is based on a public 3D images/videos data set, so the quantity contrast is unreachable. On the other hand, when people watch 2D images and 3D images, what's the difference in eye movements? How does the visual attention change after adding the depth information? Trying to solve those problems, we first design a contrast experiment of the eye movements of 2D and 3D images, and then propose two visual saliency detection models for 2D and 3D images. Specifically, this thesis mainly includes the following issues:

Firstly, we construct a data set which includes 1000 3D images of different categories, and collect the eye movements from 20 subjects. We design a contrast experiment of the eye movements of 2D and 3D images based on the 3D images experiment. The experiment proves that the number of fixation point is more when watching 3D images, and the average of fixation duration is shorter in nature scene. We can deduce that the distribution of 3D and 2D eye movements are different, from the ROC curves and ROC areas, so the depth information has a great impact on 3D visual attention. We also select 100 complex images and 100 simple images from the 1000 images, and the result shows that different scenes can produce different influence when we watch

3D images, and the depth information plays a greater role in complex scene.

Secondly, based on the conclusion of physiology, psychology and neuroscience, we propose a simple bottom-up 2D visual saliency detection model. In our computational model, we apply 2D Log-Gabor wavelet pyramid to detect the low-level irregularities in various pre-attentive features. These irregularities are then integrated by considering the influence of the "center bias" to construct the bottom-up saliency map. Based on the 2D images data set and the corresponding ground-truth, we compare our model with other four state-of-the-art approaches. Experimental results show that our approach outperforms other approaches remarkably.

Thirdly, we also propose a new multi-scale, multi-directional, binocular summation and bottom-up stereoscopic visual saliency computational model, based on the 2D visual saliency detection model. In our stereoscopic computational model, we extract color and luminance features as input based on the characteristics of the retina. We also use 2D Log-Gabor wavelet pyramid to simulate monocular receptive field, and consider the phase disparity between left and right eye, then we use the Two-stage model to simulate the binocular summation. At last we consider the influence of the "center bias" and "foreground bias", then we obtain the saliency map of 3D images finally. Based on the 3D images data set and the corresponding ground-truth, we compare our model with other five state-of-the-art approaches. Experimental results show that our approach outperforms other approaches remarkably.

In conclusion, the thesis focuses on the comparison of eye movements on 2D and 3D, 2D and 3D visual saliency models. This study lays a foundation for a great interest of research in stereoscopic visual saliency field.

Keywords: Log-Gabor filters, binocular summation, center bias and foreground bias, 2D and stereoscopic visual saliency detection, contrast experiment of the eye movements of 2D and 3D images.