# 同义图像融合系统设计与优化

王宇新 陆国际 郭 禾 何昌钦 杨元生
(大连理工大学计算机科学与工程系 大连 116023)

**摘 要** 图像无缝融合技术是图像编辑领域的重要方向,传统的方法是将目标物体从源图像中分离出来后直接嵌入到目标图像中,而未考虑源图像和目标图像在语义上是否匹配。实现了一种新的图像融合系统,它用Gist特征描述图像的场景语义,并将场景语义匹配应用于无缝融合之中,使图像融合更具有现实意义,该系统称之为同义图像融合系统。鉴于语义匹配时间复杂度较大,采用异构多核环境下CUDA并行编程和OpenMP多核多线程方法进行优化,有效提高了系统的整体性能。实验表明,该系统不仅用户参与度较高,而且实用性较强,可以达到用户满意的融合速度和效果。

**关键词** 图像编辑,场景识别,语义匹配,无缝融合
**中图法分类号** TP391.41 **文献标识码** A

## Design and Optimization of Synonymous Image Cloning System

WANG Yu-xin LU Guo-ji GUO He HE Chang-qin YANG Yuan-sheng
(Department of Computer Science and Engineering, Dalian University of Technology, Dalian 116023, China)

**Abstract** Image seamless cloning technique is an important filed in image editing region. In previous work, researchers separated the target object from source image and then embed it into the target image without consideration of whether the source image and the target one are semantic matching. A novel and complete image cloning system was presented, in which Gist descriptors were used to describe image's scene semantics and semantic matching technique was applied to arbitrary cloning method to give it a realistic significance, which is called synonymous image cloning system. In consideration of the large time complexity of the semantic matching process, CUDA parallel programming model and OpenMP multithread programing method of the heterogeneous multicore environment were used to accelerate the process and effectively improve the performance of overall system. Experiments show that this system guarantees higher user involvement, stronger practicability and ensures the users' satisfaction of the cloning speed and effect.

**Keywords** Image editing, Scene recognition, Semantic matching, Arbitrary cloning

自从Patrick P'erez等人将泊松方法用于图像融合并取得了较好的效果后[1],各种各样基于边缘检测和颜色融合技术的融合算法层出不穷。Zeng Yun,Chen Wei等人提出了一种标准变分图像编辑模型[2],通过定义梯度项,得到了视觉上更好的融合效果。本课题组在2008年提出了自由融合算法,通过改进泊松图像编辑与抠像技术,将抠像技术和图像融合技术结合成一个整体,给出了一个完整的图像融合解决方案[3],该方案成为本文的工作基础。

严格说来,图像融合技术的现实合理性源自源图像和目标图像在语义上的匹配。毕竟,鸭子在城市的人行道上游泳,或是人在海面上骑马(见图1)是不具现实意义的。因而理想的方法是首先找到一些与源图像场景语义相似的图像,再从中选择用户满意的一幅作为目标图像进行融合——我们将上述过程称之为"同义图像融合"。

寻找相似语义场景的算法很多,文献[4]论证了在一系列图像的快速呈现过程中,观察者能够分辨出每幅图像的语义类别和图像中的一些目标物体以及它们的属性特征。类似这种快速理解现象我们经常能体会,比如电影预告片大多采用众多主要场景间的快速切换来完成,一幅图像只需看一眼,即便没有记住细节,也能明白每个镜头要表达的含义[5]。通常在一瞥(大约200ms)内感知的全部信息就是场景的Gist特征[6]。

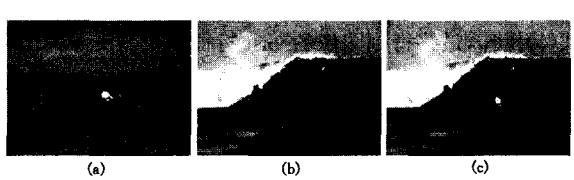


图1 源图像和目标图像语义不一致时的融合效果图

另一个问题在于,场景语义匹配的过程非常费时,特别是要从数据库数以百万计的图像中找到与源图像场景语义接近

到稿日期:2009-09-23 返修日期:2009-12-17
王宇新(1973-),男,博士生,讲师,CCF会员,主要研究方向为图像处理、计算机系统结构,E-mail:wyx@dlut.edu.cn;陆国际(1987-),男,硕士生,主要研究方向为图像处理;郭 禾(1955-),男,教授,博士生导师,CCF高级会员,主要研究方向为计算机系统结构、计算机视觉;何昌钦(1986-),男,硕士生,主要研究方向为图像识别;杨元生(1946-),男,教授,博士生导师,主要研究方向为算法设计与分析。

的目标图像。Hays 等人采用 15 台机器组成集群加速匹配过程[7]，而本文采取了更便捷更有效的加速方法。

## 1 图像场景语义匹配

图像场景语义匹配就是从给定的图像数据库中找到与源图像语义信息相似的目标图像集。本文引入 Gist 场景描述符[8]，通过计算比较 Gist 特征实现源图像与图像数据库中图像的实时匹配。

### 1.1 Gist 特征提取

Gist 场景描述符是一种基于场景中心的语义理解方法，聚合了多尺度下面向边缘响应的图像空间特征。为了使局部边缘特征不同尺度下的检测最优化，采取多尺度的滤波器对图像进行锐化预处理，再利用 Gabor 变换来提取图像的 Gist 特征。

二维 Gabor 变换的核函数可以表示为

$$g(x,y)=\frac{1}{2\pi\sigma_x\sigma_y}\exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)+2\pi jWx\right] \tag{1}$$

它的傅里叶变换为：

$$G(u,v)=\exp\left\{-\frac{1}{2}\left[\frac{(u-W)^2}{\sigma_u^2}+\frac{v^2}{\sigma_v^2}\right]\right\} \tag{2}$$

式中，$\sigma_u=1/2\pi\sigma_x$，$\sigma_v=1/2\pi\sigma_y$。

对 Gabor 变换核进行适当尺度变换和旋转变换就可得到自相似的一组滤波器，称为 Gabor 滤波器组。

$$g_{mn}(x,y)=a^{-m}g(x',y'),a>1,m,n\in Z \tag{3}$$

式中，$x'=a^{-m}(x\cos\theta+y\sin\theta)$，$y'=a^{-m}(-x\sin\theta+y\cos\theta)$，这里 $\theta=n\pi/k$，$k$ 表示方向个数($n\in[0,K]$)，$a-m$ 为尺度因子，在上式中用来确保其总能量与 $m$ 无关。通过改变 $m$ 和 $n$ 的值，便可以得到一组尺度和方向都不同的滤波器。

给定一幅图像 $I$，它的 Gabor 小波变换可以定义为：

$$W_{mn}(x,y)=\iint I(x,y)g_{mn}^{*}(x-x_1,y-y_1)\mathrm{d}x_1\mathrm{d}y_1 \tag{4}$$

式中，$g_{mn}^{*}$ 是变换核 $g_{mn}$ 的复共轭形式，中心在$(x_1,y_1)$，$W_{mn}$ 为图像 $I$ 与 $g_{mn}^{*}$ 的卷积运算结果。

上式卷积操作较复杂，借助卷积定理将其转化为时间复杂度相对较低的傅里叶和反傅里叶变换以及频域中矩阵点乘运算：

$$W_{mn}=I*g_{mn}^{*}=ifft(fft(I)\cdot fft(g_{mn}^{*})) \tag{5}$$

在计算 Gist 特征的过程中，首先需要创建 Gabor 滤波器组，3 个尺度上的方向数为[8,8,4]，总计 20 个滤波器。然后对图像进行锐化预处理，使得图像的边缘、轮廓线以及图像细节变得更清晰。将处理过的图像与所有的 20 个滤波器分别执行式(5)操作，最终得到图像的 Gist 特征向量。尺寸为 128 * 128 * 3的图像，处理后得到 320 * 3 的二维数组，进而转化为 960 * 1 维的 Gist 向量作为图像的场景描述符。

### 1.2 Gist 特征匹配

提取源图像的 Gist 特征后，需要从图像数据库中找到与该图像语义信息相似的目标图像集。直观比较就可以看出，图 2 中(a)和(b)具有相似的 Gist 特征描述符，(c)和(d)则差别很大。

计算中用欧氏距离(也称为欧几里得距离)来判断两幅图像的语义相似程度[9]。$L(x,y)$表示样本 $x$ 与 $y$ 之间的欧氏距离，样本维数为 $k$，则

$$L(x,y)=\left[\sum_{i=1}^{k}(x_i-y_i)^2\right]^{\frac{1}{2}} \tag{6}$$

计算完成源图像与所有图像的欧氏距离后，将结果进行比较排序，选出前几个最小的作为候选目标图像，供用户自由选择，以便能满足相似或相同语义图像融合的要求。

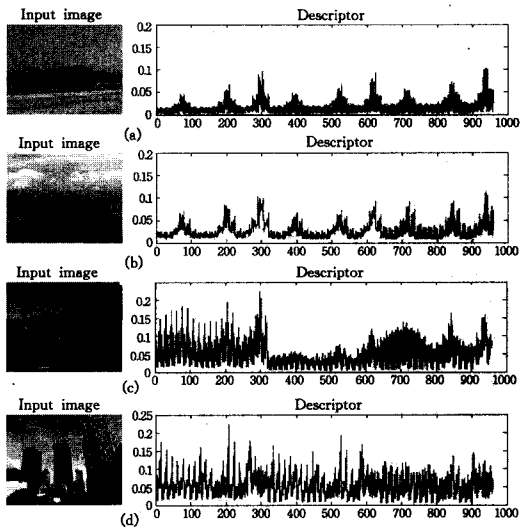


图 2 Gist 特征描述符

## 2 自由融合技术

选定好目标图像后，采用文献[3]中提出的自由融合方法进行高质量的无缝融合。首先利用抠像技术提取源图像的前景映射图，用户选定该区域在目标图像中的位置，就可完成融合操作。

### 2.1 前景映射图的提取

抠像技术假设图像满足下面的模型：

$$I_i=\alpha_iF_i+(1-\alpha_i)B_i \tag{7}$$

式中，$\alpha_i$ 是前景层的透明度，$F_i$ 为前景层，$B_i$ 为背景层。对本文的融合算法来说，只需提取 $\alpha$ 值作为泊松方程的梯度域标识，并不需要花费额外的计算代价去求出 $F$ 和 $B$。

对于灰度图像，假设在每一个像素周围的小窗口内，$F$ 和 $B$ 近似不变，那么抠像模型式(7)可以改写为 $\alpha_i\approx aI_i+b,\forall i\in w$。其中，$a=\frac{1}{F-B}$，$b=-\frac{B}{F-B}$，$w$ 是任一个小窗口。最终目标是找到 $\alpha,a,b$ 来最小化下面的代价函数[10]：

$$J(\alpha,a,b)=\sum_{j\in I}\left(\sum_{i\in w_j}(\alpha_i-a_jI_i-b_j)^2+\varepsilon a_j^2\right) \tag{8}$$

上式是关于 $\alpha,a,b$ 的二次函数，很难被直接最优化。根据文献[10]中的推导，可化简为 $J(\alpha)=\min_{a,b}J(\alpha,a,b)=\alpha^{\mathrm{T}}L\alpha$。其中 $\alpha$ 为 $N\times1$ 的向量，$N$ 为像素个数。$L$ 是一个 $N\times N$ 的矩阵，其元素可以通过周围窗口内像素点的灰度均值和方差计算出来。

对于彩色图像，可以把上面的结论扩展为 RGB 通道，也可以放松上面的基本假设。好处在于，这个颜色模型会符合大部分彩色图像，因而远远扩大了该算法的适用范围。以上详细分析与计算过程请参考文献[3]。

在提取 $\alpha$ 过程中，还需要用户在源图像上粗略标识前景层和背景层，分别用白色和黑色表示，见图 3(b)，再把这些约束添加到最优化问题中，最终需要最优化二次函数：

$$C(\alpha)=\alpha^{\mathrm{T}}L\alpha+\lambda(\alpha^{\mathrm{T}}-b_s^{\mathrm{T}})D_s(\alpha-b_s) \tag{9}$$

式中第 2 项是用户添加的约束；$\lambda$ 为一个很大的数，它决定用户约束所起的作用。$D_s$ 是一个对角阵，对角元素为 1 表示用

户标识了的像素。$b_s$ 为一个向量，对于用户标识了的像素，它的相应元素为用户标识的值(0 或 1)。把此函数所有偏导数置 0，组成线性方程组 $(L+\lambda D_s)\alpha=\lambda b_s$，求解便得到前景映射图 $\alpha$，见图 3(c)。


图 3　前景映射图的提取

### 2.2　无缝融合

引用泊松方法[1]，将源图像的前景图融合到目标图像的指定位置，就是要解决下面的最优化问题：

$$\min_f \iint_\Omega |\nabla f-v|^2 \text{ with } f|_{\partial\Omega}=I_{target}|_{\partial\Omega} \tag{10}$$

式中，$\Omega$ 为源图像中待融合区域，$f$ 是定义在 $\Omega-\partial\Omega$ 上的未知函数，$v$ 是定义在 $\Omega$ 上的已知向量域，$\nabla=\left[\frac{\partial}{\partial x},\frac{\partial}{\partial y}\right]$ 是梯度运算符。根据变分法理论中的奥氏方程，它的解就是带有狄利克雷边界条件的泊松方程的解：

$$\nabla^2 f=\text{div } v \text{ over}\Omega, \text{with } f|_{\partial\Omega}=I_{target}|_{\partial\Omega} \tag{11}$$

$\text{div } v=\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}$ 是 $v=(u,v)$ 的散度，$\nabla^2=\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}$ 是拉普拉斯运算符。

接下来通过像素点把问题离散化。对于每个像素 $p$，$N_p$ 表示 $p$ 的 4 邻域点集，$\langle p,q\rangle$ 表示一个像素对，使 $q\in N_p$。首先确定梯度域函数，取 $v=\nabla I_{source}|_\Omega$。再根据拉普拉斯 5 点有限差分公式，将其离散化为：

$$|N_p|f_p-\sum_{q\in N_p\cap\Omega} f_q=\sum_{q\in N_p\cap\partial\Omega} I_q^{target}+\sum_{q\in N_p} v_{pq},$$
$$\text{for all } p\in\Omega \tag{12}$$

当 $\Omega$ 包含图像边缘上的像素时，4 邻域自然而然就减少为 3 或 2 邻域，此时 $|N_p|<4$。而对于在 $\Omega$ 内部的像素，式(12)的右侧并没有边界条件，改写成为：

$$|N_p|f_p-\sum_{q\in N_p} f_q=\sum_{q\in N_p} v_{pq} \tag{13}$$

由于边界条件的任意性，很难利用这个方程建立一个整体的线性方程组，因此采用经典的高斯塞德尔线性迭代方法来求解 $f$。

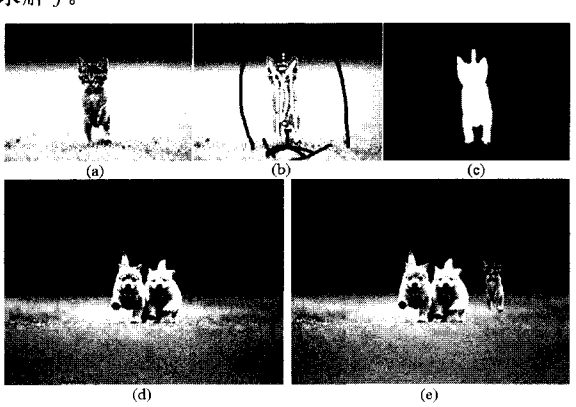


图 4　无缝融合效果图

文献[3]中大量实验对比证明，直接选取 $\Omega$ 所对应的原图像区域作为梯度域，效果非常好。用户在目标图像中选定好位置，将前景映射图嵌入该位置，通过若干次高斯赛德尔迭代完成无缝融合过程。用户不但可以控制前景映射图的走向，还可以自行控制迭代次数，达到多种多样的融合效果。图 4(a)，(b)，(c)，(d)分别代表原图像、用户提示图像、前景映射图和目标图像，(e)是采用本文梯度域进行迭代 100 次得到的融合结果。

## 3　同义语义融合系统

传统的自由融合技术没有考虑源图像和目标图像场景语义的相似性和合理性，可能产生毫无现实意义的融合结果。本文在图像无缝融合之前引入语义场景匹配方法，形成一套融合语义相近图像的完整系统，即同义图像融合系统。

### 3.1　系统实现过程

首先，图像数据库中所含全部图像都附带 Gist 特征信息，它是在追加和更新时计算而一并存入的。在语义匹配时，用户选定源图像后，系统实时提取该图像的 Gist 特征，并将其与图像数据库中所有图像分别匹配，得到语义相近的图像集。用户从中选择希望融合的目标图像。然后，系统提示用户在源图像上用白色和黑色线条粗略标识前景层与背景层，利用抠像技术结合用户约束得到待融合的前景映射图。用户在目标图像上选好位置，系统自动将前景图嵌入。用户通过控制迭代次数获得满意的效果后，图像同义融合过程终止。

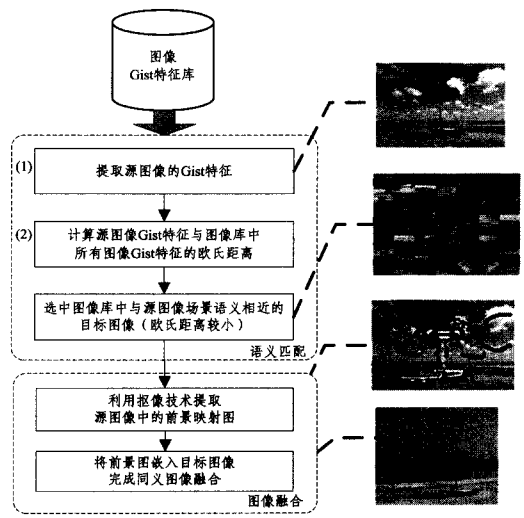


图 5　同义图像融合

语义匹配的过程是比较费时的，尤其图 5 中(1)处实时提取源图像的 Gist 特征和(2)处将源图像与图像库中所有图像的场景语义进行匹配，是整个系统运行的瓶颈。针对其不同的特点，分别采用 CUDA 并行编程和 OpenMP 支持的多核多线程方法进行优化，以满足用户的实时要求。

### 3.2　系统性能优化

NVIDIA CUDA 是一种可编程 GPU 上的并行编程模型和软件环境，可以用来对多种复杂的计算问题进行加速处理，尤其是图像领域的复杂问题。为保证源图像 Gist 特征的实时提取，在 CUDA 编程环境中调用 CUFFT 库函数加速 1.1 节中式(5)的傅里叶和反傅里叶变换，利用 CUDA 的群核运算优势加速频域中矩阵点乘。由图 6 可以看出，使用 CUDA 计算源图像的 Gist 特征要比单独的 CPU 运算效率高很多，而且图像维度增大，加速越明显。新建或更新图像数据库时，采用 CUDA 加速多幅图像 Gist 特征提取的效果更明显。

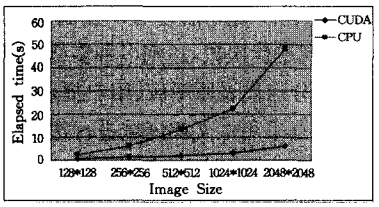


图 6　CUDA 加速 Gist 特征提取的计算时间

Gist 特征是 960 维的向量。在进行特征匹配时，若图像数据库很庞大，将所有图像 Gist 特征传至显存需要巨大的传输代价，因此对匹配过程的加速不使用 CUDA，而采用 OpenMP。后者是一种基于多线程的并行编程模型，可以较方便地应用于多核处理器平台上对程序代码进行加速。

本文采用 OpenMP 并行技术加速源图像与图像数据库之间的场景语义匹配过程，即计算 Gist 特征向量间的欧氏距离。由图 7 可以看出，在 Intel Q8200 四核处理器上使用 OpenMP 处理匹配的速度接近单线程实现的 4 倍。即使当匹配图片数量达到百万张，OpenMP 处理所需的时间也仅为 3.11s。

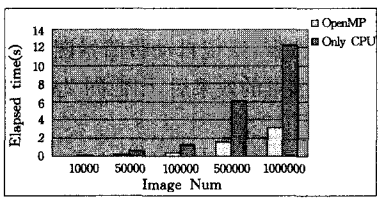


图 7　OpenMP 加速语义匹配的计算时间(图像大小为 128 * 128)

图 8 验证了系统性能优化的整体效果。同时采用 CUDA 和 OpenMP 加速技术处理场景语义匹配过程比单独 CPU 实现快约 3.5 倍，基本满足实时要求，大幅度地增强了系统的实用性。实验所使用的测试源图像大小为 128 * 128，图像数据库的大小分别为 $10^4$，$10^5$ 以及 $10^6$。

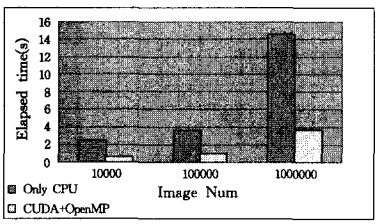


图 8　语义匹配计算总时间

以上所有的实验结果都是在配置为 Intel Q8200 CPU，2GB 内存，NVIDIA GeForce 9800 GTX＋的计算机上得到的。大量的实验证明，在这种异构多核环境下同时运用上述两种方法加速语义匹配，不仅优化了系统的整体性能，还能有效地节约宝贵的 CPU 计算资源。特别是当图片维度很大、图像数据库中图片数量很多时，优势更为明显。

## 4　系统应用效果

实验 1 显示一次完整的同义图像融合的效果。图 9(a)为源图像，(b)为用户选定的目标图像，(c)为用户提示图像，(d)为前景映射图，(f)是场景语义匹配的目标图像集合，(e)为最终同义融合结果。由于源图像和目标图像场景接近，融合效果比较自然，边缘区域也几近完美，实现了令人满意的同义融合效果。

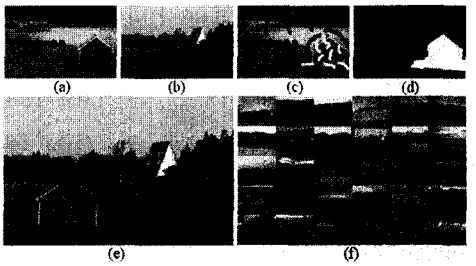


图 9　实验 1 同义融合效果图

实验 2 显示另一次完整的同义图像融合的效果，见图 10，图中排列同上。

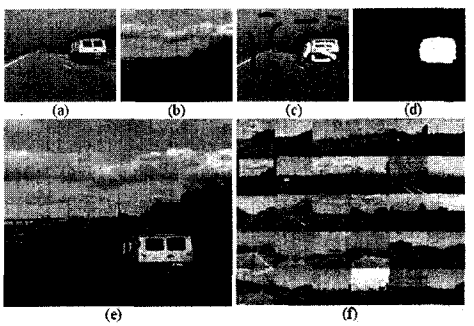


图 10　实验 2 同义融合效果图

**结束语**　本文设计并实现了一种新颖而实用的同义图像融合系统，通过引入 Gist 场景识别技术，将场景语义匹配与自由融合技术有效结合，并在语义匹配过程中应用 CUDA 并行计算和 OpenMP 多线程编程技术进行优化，最终达到用户满意的速度和效果。实验证明，由于对匹配过程的优化使得该系统在整体效率上并不低于一般的自由融合系统，而场景语义匹配的引入使现有的自由融合技术更具有现实意义。

## 参 考 文 献

[1] P′erez P, Gangnet M, Blake A. Poisson image editing[C]// ACM SIGGRAPH 2003. San Diego, 2003: 313-318

[2] Zeng Yun, Chen Wei, Peng Qunsheng. A novel variational image model: Towards a unified approach to image editing[J]. Journal of Computer Science and Technology, 2006, 21(2): 224-231

[3] 付新元，郭禾，王宇新，等. 基于抠像技术的图像无缝融合算法[J]. 中国图像图形学报，2008，13(6)：1082-1089

[4] Potter M C, Staub A, O'Connor D H. Pictorial and Conceptual Representation of Glimpsed Pictures[J]. Journal of Experimental Psychology: Human Perception and Performance, 2004, 30(3): 478-489

[5] Maljkovic V, Martini P. Short-term memory for scenes with affective content[J]. Journal of Vision, 2005, 5(3): 215-229

[6] Torralba A, Murphy K P, Freeman W T, et al. Context-based vision system for place and object recognition [C]// Proceedings of Ninth IEEE International Conference on Computer Vision, Nice. 2003: 273-280

[7] Hays J, Efros A A. Scene completion using millions of photographs[J]. Communications of the ACM, 2008, 51(10): 87-94

[8] Oliva A, Torralba A. Building the gist of a scene: The role of global image features in recognition[J]. Progress in Brain Research, 2006, 155: 23-36

[9] Torralba A, Fergus R, Freeman W T. 80 million tiny images: a large dataset for non-parametric object and scene recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(11): 1958-1970

[10] Levin A, Lischinski D, Weiss Y. A Closed Form Solution to Natural Image Matting[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, 2006, 1: 61-68