



(12) 发明专利

(10) 授权公告号 CN 103632372 B

(45) 授权公告日 2016. 02. 24

(21) 申请号 201310651561. 3

(22) 申请日 2013. 12. 05

(73) 专利权人 宁波大学

地址 315211 浙江省宁波市江北区风华路
818 号

(72) 发明人 邵枫 潘云峰 蒋刚毅 郁梅
李福翠 彭宗举

(74) 专利代理机构 宁波奥圣专利代理事务所
(普通合伙) 33226

代理人 周珏

(51) Int. Cl.

G06T 7/00(2006. 01)

审查员 郭明华

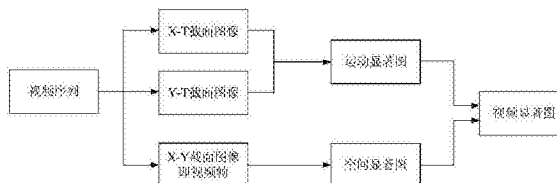
权利要求书4页 说明书9页 附图3页

(54) 发明名称

一种视频显著图提取方法

(57) 摘要

本发明公开了一种视频显著图提取方法,其首先通过对二维视频在时域上、在水平方向及在垂直方向进行采样,分别得到 X-Y 截面图像、X-T 截面图像和 Y-T 截面图像,然后通过对 X-T 截面图像和 Y-T 截面图像进行低秩矩阵分解提取出 X-Y 截面图像的运动显著图,对 X-Y 截面图像的特征矢量进行低秩矩阵分解提取出 X-Y 截面图像的空间显著图,最后对运动显著图 and 空间显著图进行融合,得到最终的视频显著图,优点是所获得的视频显著图能够较好地反映视频的静态和动态区域的显著变化情况,符合人眼运动显著语义的特征。



1. 一种视频显著图提取方法,其特征在于包括以下步骤:

①对呈三维立方体的二维视频在时域上进行采样,得到二维视频的 T 个 X - Y 截面图像,即得到二维视频的 T 个时刻的视频帧,将二维视频中的 t 时刻的视频帧记为 $\{I_t(x, y)\}$,将 $\{I_t(x, y)\}$ 中坐标位置为 (x, y) 的像素点的像素值记为 $I_t(x, y)$,其中, T 表示二维视频中包含的视频帧的总帧数, $1 \leq t \leq T, 1 \leq x \leq W, 1 \leq y \leq H$, W 表示二维视频中各时刻的视频帧的宽, H 表示二维视频中各时刻的视频帧的高;

同样,对呈三维立方体的二维视频在 X 方向上进行采样,得到二维视频的 W 个 Y - T 截面图像,将二维视频中的第 x 个 Y - T 截面图像记为 $\{I_x(y, t)\}$,将 $\{I_x(y, t)\}$ 中坐标位置为 (y, t) 的像素点的像素值记为 $I_x(y, t)$, $I_x(y, t) = I_t(x, y)$;

同样,对呈三维立方体的二维视频在 Y 方向上进行采样,得到二维视频的 H 个 X - T 截面图像,将二维视频中的第 y 个 X - T 截面图像记为 $\{I_y(x, t)\}$,将 $\{I_y(x, t)\}$ 中坐标位置为 (x, t) 的像素点的像素值记为 $I_y(x, t)$, $I_y(x, t) = I_t(x, y)$;

②对二维视频中的每个 Y - T 截面图像进行低秩矩阵分解,并获取二维视频中的每个 Y - T 截面图像的最优垂直方向运动矩阵,同样对二维视频中的每个 X - T 截面图像进行低秩矩阵分解,并获取二维视频中的每个 X - T 截面图像的最优水平方向运动矩阵;然后根据二维视频中的每个 Y - T 截面图像的最优垂直方向运动矩阵和每个 X - T 截面图像的最优水平方向运动矩阵,获取二维视频中每个时刻的视频帧的运动显著图,将二维视频中的 t 时刻的视频帧 $\{I_t(x, y)\}$ 的运动显著图记为 $\{S_{t, \text{motion}}(x, y)\}$,其中, $S_{t, \text{motion}}(x, y)$ 表示 $\{S_{t, \text{motion}}(x, y)\}$ 中坐标位置为 (x, y) 的像素点的像素值;

所述的步骤②的具体过程为:

②-1、对二维视频中的每个 Y - T 截面图像进行低秩矩阵分解,得到二维视频中的每个 Y - T 截面图像的所有垂直方向背景矩阵和所有垂直方向运动矩阵;

②-2、利用最小化概率密度函数,获取二维视频中的每个 Y - T 截面图像的最优垂直方向背景矩阵和最优垂直方向运动矩阵,将二维视频中的第 x 个 Y - T 截面图像 $\{I_x(y, t)\}$ 的最优垂直方向背景矩阵和最优垂直方向运动矩阵对应记为 \mathbf{B}_x^* 和 \mathbf{M}_x^* ,将 \mathbf{B}_x^* 和 \mathbf{M}_x^* 的组 合记

为 $(\mathbf{B}_x^*, \mathbf{M}_x^*)$, $(\mathbf{B}_x^*, \mathbf{M}_x^*) = \arg \min_{(\mathbf{B}_x, \mathbf{M}_x) \in \Omega_x} [\|\mathbf{B}_x\|_* + \lambda \|\mathbf{M}_x\|_1]$, 其中, S_x 为二维视频中的第 x 个 Y - T 截面

图像 $\{I_x(y, t)\}$ 的矩阵形式表示, S_x 、 \mathbf{B}_x^* 和 \mathbf{M}_x^* 的维数均为 $H \times T$, $\arg \min []$ 表示最小化概率密度函数, Ω_x 表示对 S_x 进行低秩矩阵分解得到的所有的垂直方向背景矩阵和垂直方向运动矩阵的组合 $(\mathbf{B}_x, \mathbf{M}_x)$ 的集合, $(\mathbf{B}_x, \mathbf{M}_x) \in \Omega_x$, 符号 “ $\|\cdot\|_*$ ” 为求取矩阵核的范数符号, 符号 “ $\|\cdot\|_1$ ” 为求取矩阵的 1- 范数符号, λ 为拉格朗日乘子;

②-3、根据二维视频的 W 个 Y - T 截面图像各自的最优垂直方向运动矩阵,通过在 X 方向上进行叠加,构造得到 Y - T 截面图像的立方体,记为 S_{cubeYT} ,其中, S_{cubeYT} 的维数为 $W \times H \times T$;

②-4、对二维视频中的每个 X - T 截面图像进行低秩矩阵分解,得到二维视频中的每个 X - T 截面图像的所有水平方向背景矩阵和所有水平方向运动矩阵;

②-5、利用最小化概率密度函数,获取二维视频中的每个 X - T 截面图像的最优水平方向背景矩阵和最优水平方向运动矩阵,将二维视频中的第 y 个 X - T 截面图像 $\{I_y(x, t)\}$ 的

最优水平方向背景矩阵和最优水平方向运动矩阵对应记为 \mathbf{B}_y^* 和 \mathbf{M}_y^* , 将 \mathbf{B}_y^* 和 \mathbf{M}_y^* 的组合记为 $(\mathbf{B}_y^*, \mathbf{M}_y^*)$, $(\mathbf{B}_y^*, \mathbf{M}_y^*) = \arg \min_{(\mathbf{B}_y, \mathbf{M}_y) \in \Omega_y} [\|\mathbf{B}_y\|_* + \lambda \|\mathbf{M}_y\|_1]$, 其中, S_y 为二维视频中的第 y 个 X-T 截面图像 $\{I_y(x, t)\}$ 的矩阵形式表示, S_y 、 \mathbf{B}_y^* 和 \mathbf{M}_y^* 的维数均为 $W \times T$, Ω_y 表示对 S_y 进行低秩矩阵分解得到的所有的水平方向背景矩阵和水平方向运动矩阵的组合 $(\mathbf{B}_y, \mathbf{M}_y)$ 的集合, $(\mathbf{B}_y, \mathbf{M}_y) \in \Omega_y$;

②-6、根据二维视频的 H 个 X-T 截面图像各自的最优水平方向运动矩阵, 通过在 Y 方向上进行叠加, 构造得到 X-T 截面图像的立方体, 记为 S_{cubeXT} , 其中, S_{cubeXT} 的维数为 $W \times H \times T$;

②-7、计算 $\text{norm}(S_{\text{cubeXT}} \cdot * S_{\text{cubeYT}})$ 作为二维视频的初步的运动显著图, 其中, $\text{norm}(\cdot)$ 表示归一化操作, $\text{norm}(S_{\text{cubeXT}} \cdot * S_{\text{cubeYT}})$ 中的符号“ $\cdot *$ ”为矩阵点乘符号;

②-8、采用高斯滤波器对二维视频的初步的运动显著图中每个时刻的 X-Y 截面图像进行平滑操作, 得到二维视频中每个时刻的视频帧的运动显著图, 将二维视频中的 t 时刻的视频帧 $\{I_t(x, y)\}$ 的运动显著图记为 $\{S_{t, \text{motion}}(x, y)\}$, 将 $\{S_{t, \text{motion}}(x, y)\}$ 中坐标位置为 (x, y) 的像素点的像素值记为 $S_{t, \text{motion}}(x, y)$, $S_{t, \text{motion}}(x, y) = S'_{t, \text{motion}}(x, y) * G_\sigma(x, y)$, 其中, $S'_{t, \text{motion}}(x, y)$ 表示二维视频的初步的运动显著图中的 t 时刻的 X-Y 截面图像, “ $*$ ”为卷积操作符号, $G_\sigma(x, y)$ 表示标准差为 σ 的高斯函数, $G_\sigma(x, y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(x^2 + y^2)}{2\sigma^2}\right)$, $\exp(\cdot)$ 为以 e 为底的指数函数, e 表示自然基数;

③提取二维视频中每个时刻的视频帧中的每个像素点的特征矢量; 然后对提取得到的所有像素点的特征矢量构成的矩阵进行低秩矩阵分解, 并获取二维视频中每个时刻的视频帧中的所有像素点的特征矢量构成的矩阵的最优背景矩阵和最优运动矩阵; 再根据二维视频中每个时刻的视频帧中的所有像素点的特征矢量构成的矩阵的最优运动矩阵, 获得二维视频中每个时刻的视频帧的空间显著图, 将二维视频中的 t 时刻的视频帧 $\{I_t(x, y)\}$ 的空间显著图记为 $\{S_{t, \text{spatial}}(x, y)\}$, 其中, $S_{t, \text{spatial}}(x, y)$ 表示 $\{S_{t, \text{spatial}}(x, y)\}$ 中坐标位置为 (x, y) 的像素点的像素值;

所述的步骤③的具体过程为:

③-1、将当前正在处理的二维视频中的 t 时刻的视频帧 $\{I_t(x, y)\}$ 定义为当前视频帧;

③-2、提取当前视频帧中的每个像素点的红颜色分量、绿颜色分量、蓝颜色分量、色调分量和饱和度分量, 由提取出的每个像素点的红颜色分量、绿颜色分量、蓝颜色分量、色调分量和饱和度分量构成对应像素点的第一特征矢量, 其中, 当前视频帧中的每个像素点的第一特征矢量的维数为 $5 \times (W \times H)$;

③-3、提取当前视频帧中的每个像素点经 4 个方向和 3 个尺度的可控金字塔滤波后得到的振幅, 由提取出的每个像素点对应的 12 个振幅构成对应像素点的第二特征矢量, 其中, 当前视频帧中的每个像素点的第二特征矢量的维数为 $12 \times (W \times H)$;

③-4、提取当前视频帧中的每个像素点经 12 个方向和 3 个尺度的 Gabor 滤波后得到的振幅, 由提取出的每个像素点对应的 36 个振幅构成对应像素点的第三特征矢量, 其中, 当前视频帧中的每个像素点的第三特征矢量的维数为 $36 \times (W \times H)$;

③-5、将当前视频帧中的每个像素点的第一特征矢量、第二特征矢量和第三特征矢量

按序重组构成对应像素点的特征矢量, 然后对当前视频帧中的所有像素点的特征矢量 构成的矩阵进行低秩矩阵分解, 得到当前视频帧中的所有像素点的特征矢量构成的矩阵的所有背景矩阵和所有运动矩阵;

③ -6、利用最小化概率密度函数, 获取当前视频帧中的所有像素点的特征矢量构成的矩阵的最优背景矩阵和最优运动矩阵, 对应记为 \mathbf{B}_t^* 和 \mathbf{M}_t^* , 将 \mathbf{B}_t^* 和 \mathbf{M}_t^* 的组合记为 $(\mathbf{B}_t^*, \mathbf{M}_t^*)$, $(\mathbf{B}_t^*, \mathbf{M}_t^*) = \arg \min_{(\mathbf{B}_t, \mathbf{M}_t) \in \Omega_t} [\|\mathbf{B}_t\|_* + \lambda \|\mathbf{M}_t\|_1]$, 其中, F 为当前视频帧中的所有像素点的特征矢量构成的矩阵, F 、 \mathbf{B}_t^* 和 \mathbf{M}_t^* 的维数均为 $53 \times (W \times H)$, $\arg \min[\]$ 表示最小化概率密度函数, Ω_t 表示对 F 进行低秩矩阵分解得到的所有的背景矩阵和运动矩阵的组合 $(\mathbf{B}_t, \mathbf{M}_t)$ 的集合, $(\mathbf{B}_t, \mathbf{M}_t) \in \Omega_t$, 符号“ $\|\ \ \|_*$ ”为求取矩阵核的范数符号, 符号“ $\|\ \ \|_1$ ”为求取矩阵的 1-范数符号, λ 为拉格朗日乘子;

③ -7、采用超像素分割技术将当前视频帧分割成 M 个互不重叠的区域, 然后将 \mathbf{M}_t^* 表示为 $\mathbf{M}_t^* = [(\mathbf{M}_{t,1}^*)^T, (\mathbf{M}_{t,2}^*)^T, \dots, (\mathbf{M}_{t,M}^*)^T]^T$, 其中, $M \geq 1$, $\mathbf{M}_{t,1}^*$ 表示由 \mathbf{M}_t^* 中属于第 1 个区域内的所有像素点的特征矢量构成的矩阵, $\mathbf{M}_{t,2}^*$ 表示由 \mathbf{M}_t^* 中属于第 2 个区域内的所有像素点的特征矢量构成的矩阵, $\mathbf{M}_{t,M}^*$ 表示由 \mathbf{M}_t^* 中属于第 M 个区域内的所有像素点的特征矢量构成的矩阵, $(\mathbf{M}_{t,1}^*)^T$ 为 $\mathbf{M}_{t,1}^*$ 的转置矢量, $(\mathbf{M}_{t,2}^*)^T$ 为 $\mathbf{M}_{t,2}^*$ 的转置矢量, $(\mathbf{M}_{t,M}^*)^T$ 为 $\mathbf{M}_{t,M}^*$ 的转置矢量, $[(\mathbf{M}_{t,1}^*)^T, (\mathbf{M}_{t,2}^*)^T, \dots, (\mathbf{M}_{t,M}^*)^T]^T$ 为 $[(\mathbf{M}_{t,1}^*)^T, (\mathbf{M}_{t,2}^*)^T, \dots, (\mathbf{M}_{t,M}^*)^T]$ 的转置矢量, 符号“ $[\]$ ”为矢量表示符号;

③ -8、计算当前视频帧中的每个区域的空间显著值, 将当前视频帧中的第 k 个区域的空间显著值记为 $f_{t,k}$, $f_{t,k} = \|\mathbf{M}_{t,k}^*\|_1$, 然后将当前视频帧中的每个区域的空间显著值作为对应区域内的所有像素点的空间显著值, 得到当前视频帧中的每个像素点的空间显著值, 从而得到当前视频帧的空间显著图, 记为 $\{S_{t, \text{spatial}}(x, y)\}$, 其中, $1 \leq k \leq M$, $\mathbf{M}_{t,k}^*$ 表示由 \mathbf{M}_t^* 中属于第 k 个区域内的所有像素点的特征矢量构成的矩阵, $S_{t, \text{spatial}}(x, y)$ 表示 $\{S_{t, \text{spatial}}(x, y)\}$ 中坐标位置为 (x, y) 的像素点的像素值;

③ -9、令 $t = t+1$, 将二维视频中下一个时刻的视频帧作为当前视频帧, 然后返回步骤 ③ -2 继续执行, 直至二维视频中的所有视频帧处理完毕, 得到二维视频中每个时刻的视频帧的空间显著图, 其中, $t = t+1$ 中的“=”为赋值符号;

④对二维视频中每个时刻的视频帧的运动显著图和空间显著图进行融合, 得到二维视频中每个时刻的视频帧的视频显著图, 将二维视频中 t 时刻的视频帧 $\{I_t(x, y)\}$ 的视频显著图记为 $\{S_{\text{video}}(x, y)\}$, 将 $\{S_{\text{video}}(x, y)\}$ 中坐标位置为 (x, y) 的像素点的像素值记为 $S_{\text{video}}(x, y)$, $S_{\text{video}}(x, y) = (S_{t, \text{motion}}(x, y))^\alpha \times (S_{t, \text{spatial}}(x, y))^{1-\alpha}$, 其中, α 为权重系数。

2. 根据权利要求 1 所述的一种视频显著图提取方法, 其特征在于所述的步骤② -2 和所述的步骤② -5 中取 $\lambda = 0.06$; 所述的步骤② -8 中取 $\sigma = 5$ 。

3. 根据权利要求 1 所述的一种视频显著图提取方法, 其特征在于所述的步骤③ -6 中取

$\lambda = 0.06$;所述的步骤③-7 中取 $M = 200$ 。

4. 根据权利要求 3 所述的一种视频显著图提取方法,其特征在於所述的步骤④中取 $\alpha = 0.3$ 。

一种视频显著图提取方法

技术领域

[0001] 本发明涉及一种视频信号的处理方法,尤其是涉及一种视频显著图提取方法。

背景技术

[0002] 在人类视觉接收与信息处理中,由于大脑资源有限以及外界环境信息重要性区别,因此在处理过程中人脑对外界环境信息并不是一视同仁的,而是表现出选择特征。人们在观看图像或者视频片段时注意力并非均匀分布到图像的每个区域,而是对某些显著区域关注度更高。如何将视频中视觉注意度高的显著区域检测并提取出来是计算机视觉以及基于内容的视频检索领域的一个重要的研究内容。

[0003] 目前,通常采用视频分割、光流法、运动估计等方法来提取视频显著区域,然而这些方法并不能很好地将运动的物体和静态背景进行分离,这是因为:一方面,通过运动估计或帧差法得到的运动信息只反映运动趋势,并不能很好地反映运动显著语义特征;另一方面,视频中的静态场景也会影响视觉注意力的判断,在某些情况下,空间显著图会起主要的作用,因此,如何提取反映运动显著语义特征的运动显著图,如何对运动显著图和空间显著图进行结合,都是视频显著图提取中需要研究解决的问题。

发明内容

[0004] 本发明所要解决的技术问题是提供一种符合人眼运动显著语义特征,且提取精度高的视频显著图提取方法。

[0005] 本发明解决上述技术问题所采用的技术方案为:一种视频显著图提取方法,其特征在于包括以下步骤:

[0006] ①对呈三维立方体的二维视频在时域上进行采样,得到二维视频的 T 个 X - Y 截面图像,即得到二维视频的 T 个时刻的视频帧,将二维视频中的 t 时刻的视频帧记为 $\{I_t(x, y)\}$,将 $\{I_t(x, y)\}$ 中坐标位置为 (x, y) 的像素点的像素值记为 $I_t(x, y)$,其中, T 表示二维视频中包含的视频帧的总帧数, $1 \leq t \leq T$, $1 \leq x \leq W$, $1 \leq y \leq H$, W 表示二维视频中各时刻的视频帧的宽, H 表示二维视频中各时刻的视频帧的高;

[0007] 同样,对呈三维立方体的二维视频在 X 方向上进行采样,得到二维视频的 W 个 Y - T 截面图像,将二维视频中的第 x 个 Y - T 截面图像记为 $\{I_x(y, t)\}$,将 $\{I_x(y, t)\}$ 中坐标位置为 (y, t) 的像素点的像素值记为 $I_x(y, t)$, $I_x(y, t) = I_t(x, y)$;

[0008] 同样,对呈三维立方体的二维视频在 Y 方向上进行采样,得到二维视频的 H 个 X - T 截面图像,将二维视频中的第 y 个 X - T 截面图像记为 $\{I_y(x, t)\}$,将 $\{I_y(x, t)\}$ 中坐标位置为 (x, t) 的像素点的像素值记为 $I_y(x, t)$, $I_y(x, t) = I_t(x, y)$;

[0009] ②对二维视频中的每个 Y - T 截面图像进行低秩矩阵分解,并获取二维视频中的每个 Y - T 截面图像的最优垂直方向运动矩阵,同样对二维视频中的每个 X - T 截面图像进行低秩矩阵分解,并获取二维视频中的每个 X - T 截面图像的最优水平方向运动矩阵;然后根据二维视频中的每个 Y - T 截面图像的最优垂直方向运动矩阵和每个 X - T 截面图像的

最优水平方向运动矩阵,获取二维视频中每个时刻的视频帧的运动显著图,将二维视频中的 t 时刻的视频帧 $\{I_t(x, y)\}$ 的运动显著图记为 $\{S_{t, \text{motion}}(x, y)\}$,其中, $S_{t, \text{motion}}(x, y)$ 表示 $\{S_{t, \text{motion}}(x, y)\}$ 中坐标位置为 (x, y) 的像素点的像素值;

[0010] ③提取二维视频中每个时刻的视频帧中的每个像素点的特征矢量;然后对提取得到的所有像素点的特征矢量构成的矩阵进行低秩矩阵分解,并获取二维视频中每个时刻的视频帧中的所有像素点的特征矢量构成的矩阵的最优背景矩阵和最优运动矩阵;再根据二维视频中每个时刻的视频帧中的所有像素点的特征矢量构成的矩阵的最优运动矩阵,获得二维视频中每个时刻的视频帧的空间显著图,将二维视频中的 t 时刻的视频帧 $\{I_t(x, y)\}$ 的空间显著图记为 $\{S_{t, \text{spatial}}(x, y)\}$,其中, $S_{t, \text{spatial}}(x, y)$ 表示 $\{S_{t, \text{spatial}}(x, y)\}$ 中坐标位置为 (x, y) 的像素点的像素值;

[0011] ④对二维视频中每个时刻的视频帧的运动显著图和空间显著图进行融合,得到二维视频中每个时刻的视频帧的视频显著图,将二维视频中 t 时刻的视频帧 $\{I_t(x, y)\}$ 的视频显著图记为 $\{S_{\text{video}}(x, y)\}$,将 $\{S_{\text{video}}(x, y)\}$ 中坐标位置为 (x, y) 的像素点的像素值记为 $S_{\text{video}}(x, y)$, $S_{\text{video}}(x, y) = (S_{t, \text{motion}}(x, y))^\alpha \times (S_{t, \text{spatial}}(x, y))^{1-\alpha}$,其中, α 为权重系数。

[0012] 所述的步骤②的具体过程为:

[0013] ②-1、对二维视频中的每个 Y-T 截面图像进行低秩矩阵分解,得到二维视频中的每个 Y-T 截面图像的所有垂直方向背景矩阵和所有垂直方向运动矩阵;

[0014] ②-2、利用最小化概率密度函数,获取二维视频中的每个 Y-T 截面图像的最优垂直方向背景矩阵和最优垂直方向运动矩阵,将二维视频中的第 x 个 Y-T 截面图像 $\{I_x(y, t)\}$ 的最优垂直方向背景矩阵和最优垂直方向运动矩阵对应记为 \mathbf{B}_x^* 和 \mathbf{M}_x^* ,将 \mathbf{B}_x^* 和 \mathbf{M}_x^* 的组合

记为 $(\mathbf{B}_x^*, \mathbf{M}_x^*)$, $(\mathbf{B}_x^*, \mathbf{M}_x^*) = \arg \min_{(\mathbf{B}_x, \mathbf{M}_x) \in \Omega_x} [\|\mathbf{B}_x\|_* + \lambda \|\mathbf{M}_x\|_1]$,其中, S_x 为二维视频中的第 x 个 Y-T 截

面图像 $\{I_x(y, t)\}$ 的矩阵形式表示, $S_x = \mathbf{B}_x^* + \mathbf{M}_x^*$, S_x 、 \mathbf{B}_x^* 和 \mathbf{M}_x^* 的维数均为 $H \times T$, $\arg \min[\]$ 表示最小化概率密度函数, Ω_x 表示对 S_x 进行低秩矩阵分解得到的所有的垂直方向背景矩阵和垂直方向运动矩阵的组合 $(\mathbf{B}_x, \mathbf{M}_x)$ 的集合, $(\mathbf{B}_x, \mathbf{M}_x) \in \Omega_x$, 符号“ $\|\ \ \|_*$ ”为求取矩阵核的范数符号,符号“ $\|\ \ \|_1$ ”为求取矩阵的 1-范数符号, λ 为拉格朗日乘子;

[0015] ②-3、根据二维视频的 W 个 Y-T 截面图像各自的最优垂直方向运动矩阵,通过在 X 方向上进行叠加,构造得到 Y-T 截面图像的立方体,记为 S_{cubeYT} ,其中, S_{cubeYT} 的维数为 $W \times H \times T$;

[0016] ②-4、对二维视频中的每个 X-T 截面图像进行低秩矩阵分解,得到二维视频中的每个 X-T 截面图像的所有水平方向背景矩阵和所有水平方向运动矩阵;

[0017] ②-5、利用最小化概率密度函数,获取二维视频中的每个 X-T 截面图像的最优水平方向背景矩阵和最优水平方向运动矩阵,将二维视频中的第 y 个 X-T 截面图像 $\{I_y(x, t)\}$ 的最优水平方向背景矩阵和最优水平方向运动矩阵对应记为 \mathbf{B}_y^* 和 \mathbf{M}_y^* ,将 \mathbf{B}_y^* 和 \mathbf{M}_y^* 的组合

记为 $(\mathbf{B}_y^*, \mathbf{M}_y^*)$, $(\mathbf{B}_y^*, \mathbf{M}_y^*) = \arg \min_{(\mathbf{B}_y, \mathbf{M}_y) \in \Omega_y} [\|\mathbf{B}_y\|_* + \lambda \|\mathbf{M}_y\|_1]$,其中, S_y 为二维视频中的第 y 个 X-T

截面图像 $\{I_y(x, t)\}$ 的矩阵形式表示, $S_y = \mathbf{B}_y^* + \mathbf{M}_y^*$, S_y 、 \mathbf{B}_y^* 和 \mathbf{M}_y^* 的维数均为 $W \times T$, Ω_y

表示对 S_y 进行低秩矩阵分解得到的所有的水平方向背景矩阵和水平方向运动矩阵的组合 (B_y, M_y) 的集合, $(B_y, M_y) \in \Omega_y$;

[0018] ②-6、根据二维视频的 H 个 X - T 截面图像各自的最优水平方向运动矩阵, 通过在 Y 方向上进行叠加, 构造得到 X - T 截面图像的立方体, 记为 S_{cubeXT} , 其中, S_{cubeXT} 的维数为 $W \times H \times T$;

[0019] ②-7、计算 $\text{norm}(S_{\text{cubeXT}} * S_{\text{cubeYT}})$ 作为二维视频的初步的运动显著图, 其中, $\text{norm}()$ 表示归一化操作, $\text{norm}(S_{\text{cubeXT}} * S_{\text{cubeYT}})$ 中的符号“ $*$ ”为矩阵点乘符号;

[0020] ②-8、采用高斯滤波器对二维视频的初步的运动显著图中每个时刻的 X - Y 截面图像进行平滑操作, 得到二维视频中每个时刻的视频帧的运动显著图, 将二维视频中的 t 时刻的视频帧 $\{I_t(x, y)\}$ 的运动显著图记为 $\{S_{t, \text{motion}}(x, y)\}$, 将 $\{S_{t, \text{motion}}(x, y)\}$ 中坐标位置为 (x, y) 的像素点的像素值记为 $S_{t, \text{motion}}(x, y)$, $S_{t, \text{motion}}(x, y) = S'_{t, \text{motion}}(x, y) * G_\sigma(x, y)$, 其中, $S'_{t, \text{motion}}(x, y)$ 表示二维视频的初步的运动显著图中的 t 时刻的 X - Y 截面图像, “ $*$ ”为卷积

操作符号, $G_\sigma(x, y)$ 表示标准差为 σ 的高斯函数, $G_\sigma(x, y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(x^2 + y^2)}{2\sigma^2}\right)$, $\exp()$

为以 e 为底的指数函数, e 表示自然基数。

[0021] 所述的步骤②-2 和所述的步骤②-5 中取 $\lambda = 0.06$; 所述的步骤②-8 中取 $\sigma = 5$ 。

[0022] 所述的步骤③的具体过程为:

[0023] ③-1、将当前正在处理的二维视频中的 t 时刻的视频帧 $\{I_t(x, y)\}$ 定义为当前视频帧;

[0024] ③-2、提取当前视频帧中的每个像素点的红颜色分量、绿颜色分量、蓝颜色分量、色调分量和饱和度分量, 由提取出的每个像素点的红颜色分量、绿颜色分量、蓝颜色分量、色调分量和饱和度分量构成对应像素点的第一特征矢量, 其中, 当前视频帧中的每个像素点的第一特征矢量的维数为 $5 \times (W \times H)$;

[0025] ③-3、提取当前视频帧中的每个像素点经 4 个方向和 3 个尺度的可控金字塔滤波后得到的振幅, 由提取出的每个像素点对应的 12 个振幅构成对应像素点的第二特征矢量, 其中, 当前视频帧中的每个像素点的第二特征矢量的维数为 $12 \times (W \times H)$;

[0026] ③-4、提取当前视频帧中的每个像素点经 12 个方向和 3 个尺度的 Gabor 滤波后得到的振幅, 由提取出的每个像素点对应的 36 个振幅构成对应像素点的第三特征矢量, 其中, 当前视频帧中的每个像素点的第三特征矢量的维数为 $36 \times (W \times H)$;

[0027] ③-5、将当前视频帧中的每个像素点的第一特征矢量、第二特征矢量和第三特征矢量按序重组构成对应像素点的特征矢量, 然后对当前视频帧中的所有像素点的特征矢量构成的矩阵进行低秩矩阵分解, 得到当前视频帧中的所有像素点的特征矢量构成的矩阵的所有背景矩阵和所有运动矩阵;

[0028] ③-6、利用最小化概率密度函数, 获取当前视频帧中的所有像素点的特征矢量构成的矩阵的最优背景矩阵和最优运动矩阵, 对应记为 B_t^* 和 M_t^* , 将 B_t^* 和 M_t^* 的组合记为

(B_t^*, M_t^*) , $(B_t^*, M_t^*) = \arg \min_{(B_t, M_t) \in \Omega_t} [\|B_t\|_* + \lambda \|M_t\|_F]$, 其中, F 为当前视频帧中的所有像素点的特征

矢量构成的矩阵, $F = B_t^* + M_t^*$, F 、 B_t^* 和 M_t^* 的维数均为 $53 \times (W \times H)$, $\arg \min[]$ 表示最小化

概率密度函数, Ω_t 表示对 F 进行低秩矩阵分解得到的所有的背景矩阵和运动矩阵的组合 (B_t, M_t) 的集合, $(B_t, M_t) \in \Omega_t$, 符号 “ $\| \cdot \|_*$ ” 为求取矩阵核的范数符号, 符号 “ $\| \cdot \|_1$ ” 为求取矩阵的 1- 范数符号, λ 为拉格朗日乘子;

[0029] ③ -7、采用超像素分割技术将当前视频帧分割成 M 个互不重叠的区域, 然后将 M_t^* 表示为 $M_t^* = \left[(M_{t,1}^*)^T, (M_{t,2}^*)^T, \dots, (M_{t,M}^*)^T \right]^T$, 其中, $M \geq 1$, $M_{t,1}^*$ 表示由 M_t^* 中属于第 1 个区域内的所有像素点的特征矢量构成的矩阵, $M_{t,2}^*$ 表示由 M_t^* 中属于第 2 个区域内的所有像素点的特征矢量构成的矩阵, $M_{t,M}^*$ 表示由 M_t^* 中属于第 M 个区域内的所有像素点的特征矢量构成的矩阵, $(M_{t,1}^*)^T$ 为 $M_{t,1}^*$ 的转置矢量, $(M_{t,2}^*)^T$ 为 $M_{t,2}^*$ 的转置矢量, $(M_{t,M}^*)^T$ 为 $M_{t,M}^*$ 的转置矢量, $\left[(M_{t,1}^*)^T, (M_{t,2}^*)^T, \dots, (M_{t,M}^*)^T \right]^T$ 为 $\left[(M_{t,1}^*)^T, (M_{t,2}^*)^T, \dots, (M_{t,M}^*)^T \right]$ 的转置矢量, 符号 “[]” 为矢量表示符号;

[0030] ③ -8、计算当前视频帧中的每个区域的空间显著值, 将当前视频帧中的第 k 个区域的空间显著值记为 $f_{t,k}$, $f_{t,k} = \left\| M_{t,k}^* \right\|$, 然后将当前视频帧中的每个区域的空间显著值作为对应区域内的所有像素点的空间显著值, 得到当前视频帧中的每个像素点的空间显著值, 从而得到当前视频帧的空间显著图, 记为 $\{S_{t, \text{spatial}}(x, y)\}$, 其中, $1 \leq k \leq M$, $M_{t,k}^*$ 表示由 M_t^* 中属于第 k 个区域内的所有像素点的特征矢量构成的矩阵, $S_{t, \text{spatial}}(x, y)$ 表示 $\{S_{t, \text{spatial}}(x, y)\}$ 中坐标位置为 (x, y) 的像素点的像素值;

[0031] ③ -9、令 $t=t+1$, 将二维视频中下一个时刻的视频帧作为当前视频帧, 然后返回步骤③ -2 继续执行, 直至二维视频中的所有视频帧处理完毕, 得到二维视频中每个时刻的视频帧的空间显著图, 其中, $t=t+1$ 中的 “=” 为赋值符号。

[0032] 所述的步骤③ -6 中取 $\lambda = 0.06$; 所述的步骤③ -7 中取 $M=200$ 。

[0033] 所述的步骤④中取 $\alpha = 0.3$ 。

[0034] 与现有技术相比, 本发明的优点在于:

[0035] 1) 本发明方法首先通过对二维视频在时域上、在水平方向及在垂直方向进行采样, 分别得到 X-Y 截面图像、X-T 截面图像和 Y-T 截面图像, 然后通过对 X-T 截面图像和 Y-T 截面图像进行低秩矩阵分解提取出 X-Y 截面图像的运动显著图, 对 X-Y 截面图像的特征矢量进行低秩矩阵分解提取出 X-Y 截面图像的空间显著图, 最后对运动显著图和空间显著图进行融合, 得到最终的视频显著图, 所获得的视频显著图能够较好地反映视频的静态和动态区域的显著变化情况。

[0036] 2) 本发明方法采用低秩矩阵分解将图像分解为背景矩阵和运动矩阵, 并从运动矩阵中提取出显著信息, 这样能够提取符合人眼运动显著语义的特征信息, 有效地提高了特征提取的鲁棒性, 从而提高了视频显著图的提取精度。

附图说明

[0037] 图 1 为本发明方法的总体实现框图;

[0038] 图 2a 为 “video5” 视频序列的第 100 帧视频图像;

- [0039] 图 2b 为“video5”视频序列的第 100 帧视频图像的运动显著图；
- [0040] 图 2c 为“video5”视频序列的第 100 帧视频图像的空间显著图；
- [0041] 图 2d 为“video5”视频序列的第 100 帧视频图像的视频显著图；
- [0042] 图 3a 为“video8”视频序列的第 100 帧视频图像；
- [0043] 图 3b 为“video8”视频序列的第 100 帧视频图像的运动显著图；
- [0044] 图 3c 为“video8”视频序列的第 100 帧视频图像的空间显著图；
- [0045] 图 3d 为“video8”视频序列的第 100 帧视频图像的视频显著图；
- [0046] 图 4a 为“video10”视频序列的第 100 帧视频图像；
- [0047] 图 4b 为“video10”视频序列的第 100 帧视频图像的运动显著图；
- [0048] 图 4c 为“video10”视频序列的第 100 帧视频图像的空间显著图；
- [0049] 图 4d 为“video10”视频序列的第 100 帧视频图像的视频显著图；
- [0050] 图 5a 为“video21”视频序列的第 100 帧视频图像；
- [0051] 图 5b 为“video21”视频序列的第 100 帧视频图像的运动显著图；
- [0052] 图 5c 为“video21”视频序列的第 100 帧视频图像的空间显著图；
- [0053] 图 5d 为“video21”视频序列的第 100 帧视频图像的视频显著图。

具体实施方式

[0054] 以下结合附图实施例对本发明作进一步详细描述。

[0055] 本发明提出的一种视频显著图提取方法,其总体实现框图如图 1 所示,其包括以下步骤:

[0056] ①由于二维视频可以看作是一个三维阵列即三维立方体,其在各个坐标轴上的语义特征是完全不一样的,因此本发明分别在时域上、在 X 方向上和 Y 方向上进行采样,即对呈三维立方体的二维视频在时域上进行采样,得到二维视频的 T 个 X-Y 截面图像,即得到二维视频的 T 个时刻的视频帧,将二维视频中的 t 时刻的视频帧记为 $\{I_t(x, y)\}$,将 $\{I_t(x, y)\}$ 中坐标位置为 (x, y) 的像素点的像素值记为 $I_t(x, y)$,其中, T 表示二维视频中包含的视频帧的总帧数, $1 \leq t \leq T$, $1 \leq x \leq W$, $1 \leq y \leq H$, W 表示二维视频中各时刻的视频帧的宽, H 表示二维视频中各时刻的视频帧的高。

[0057] 同样,对呈三维立方体的二维视频在 X 方向上进行采样,得到二维视频的 W 个 Y-T 截面图像,将二维视频中的第 x 个 Y-T 截面图像记为 $\{I_x(y, t)\}$,将 $\{I_x(y, t)\}$ 中坐标位置为 (y, t) 的像素点的像素值记为 $I_x(y, t)$, $I_x(y, t) = I_t(x, y)$,即两者相等。

[0058] 同样,对呈三维立方体的二维视频在 Y 方向上进行采样,得到二维视频的 H 个 X-T 截面图像,将二维视频中的第 y 个 X-T 截面图像记为 $\{I_y(x, t)\}$,将 $\{I_y(x, t)\}$ 中坐标位置为 (x, t) 的像素点的像素值记为 $I_y(x, t)$, $I_y(x, t) = I_t(x, y)$,即两者相等。

[0059] 在三维立方体中, (x, y, t) 是三维立方体中的任意点的坐标位置。

[0060] ②由于 Y-T 截面图像和 X-T 截面图像包含了二维视频在垂直和水平方向的运动信息,因此本发明对二维视频中的每个 Y-T 截面图像进行低秩矩阵分解(low-rank matrix decomposition),并获取二维视频中的每个 Y-T 截面图像的最优垂直方向运动矩阵,同样对二维视频中的每个 X-T 截面图像进行低秩矩阵分解,并获取二维视频中的每个 X-T 截面图像的最优水平方向运动矩阵;然后根据二维视频中的每个 Y-T 截面图像的最优垂直

方向运动矩阵和每个 X-T 截面图像的最优水平方向运动矩阵,获取二维视频中每个时刻的视频帧的运动显著图,将二维视频中的 t 时刻的视频帧 $\{I_t(x, y)\}$ 的运动显著图记为 $\{S_{t, \text{motion}}(x, y)\}$,其中, $S_{t, \text{motion}}(x, y)$ 表示 $\{S_{t, \text{motion}}(x, y)\}$ 中坐标位置为 (x, y) 的像素点的像素值。

[0061] 在此具体实施例中,步骤②的具体过程为:

[0062] ②-1、对二维视频中的每个 Y-T 截面图像进行低秩矩阵分解,得到二维视频中的每个 Y-T 截面图像的所有垂直方向背景矩阵和所有垂直方向运动矩阵。

[0063] ②-2、利用最小化概率密度函数,获取二维视频中的每个 Y-T 截面图像的最优垂直方向背景矩阵和最优垂直方向运动矩阵,将二维视频中的第 x 个 Y-T 截面图像 $\{I_x(y, t)\}$ 的最优垂直方向背景矩阵和最优垂直方向运动矩阵对应记为 \mathbf{B}_x^* 和 \mathbf{M}_x^* ,将 \mathbf{B}_x^* 和 \mathbf{M}_x^* 的组合

记为 $(\mathbf{B}_x^*, \mathbf{M}_x^*)$, $(\mathbf{B}_x^*, \mathbf{M}_x^*) = \arg \min_{(\mathbf{B}_x, \mathbf{M}_x) \in \Omega_x} [\|\mathbf{B}_x\|_* + \lambda \|\mathbf{M}_x\|_1]$,其中, S_x 为二维视频中的第 x 个 Y-T 截

面图像 $\{I_x(y, t)\}$ 的矩阵形式表示, $S_x = \mathbf{B}_x^* + \mathbf{M}_x^*$, S_x 、 \mathbf{B}_x^* 和 \mathbf{M}_x^* 的维数均为 $H \times T$, $\arg \min[]$ 表示最小化概率密度函数, Ω_x 表示对 S_x 进行低秩矩阵分解得到的所有的垂直方向背景矩阵和垂直方向运动矩阵的组合 $(\mathbf{B}_x, \mathbf{M}_x)$ 的集合, $(\mathbf{B}_x, \mathbf{M}_x) \in \Omega_x$, 符号“ $\|\cdot\|_*$ ”为求取矩阵核的范数符号,符号“ $\|\cdot\|_1$ ”为求取矩阵的 1-范数符号, λ 为拉格朗日乘子,在本实施例中取 $\lambda = 0.06$ 。

[0064] 在本实施例中,采用鲁棒主成分分析方法求解最小化概率密度函数。

[0065] ②-3、根据二维视频的 W 个 Y-T 截面图像各自的最优垂直方向运动矩阵,通过在 X 方向上进行叠加,构造得到 Y-T 截面图像的立方体,记为 S_{cubeYT} ,其中, S_{cubeYT} 的维数为 $W \times H \times T$ 。

[0066] ②-4、对二维视频中的每个 X-T 截面图像进行低秩矩阵分解,得到二维视频中的每个 X-T 截面图像的所有水平方向背景矩阵和所有水平方向运动矩阵。

[0067] ②-5、利用最小化概率密度函数,获取二维视频中的每个 X-T 截面图像的最优水平方向背景矩阵和最优水平方向运动矩阵,将二维视频中的第 y 个 X-T 截面图像 $\{I_y(x, t)\}$ 的最优水平方向背景矩阵和最优水平方向运动矩阵对应记为 \mathbf{B}_y^* 和 \mathbf{M}_y^* ,将 \mathbf{B}_y^* 和 \mathbf{M}_y^* 的组合

记为 $(\mathbf{B}_y^*, \mathbf{M}_y^*)$, $(\mathbf{B}_y^*, \mathbf{M}_y^*) = \arg \min_{(\mathbf{B}_y, \mathbf{M}_y) \in \Omega_y} [\|\mathbf{B}_y\|_* + \lambda \|\mathbf{M}_y\|_1]$,其中, S_y 为二维视频中的第 y 个 X-T

截面图像 $\{I_y(x, t)\}$ 的矩阵形式表示, $S_y = \mathbf{B}_y^* + \mathbf{M}_y^*$, S_y 、 \mathbf{B}_y^* 和 \mathbf{M}_y^* 的维数均为 $W \times T$, Ω_y 表示对 S_y 进行低秩矩阵分解得到的所有的水平方向背景矩阵和水平方向运动矩阵的组合 $(\mathbf{B}_y, \mathbf{M}_y)$ 的集合, $(\mathbf{B}_y, \mathbf{M}_y) \in \Omega_y$, 在本实施例中取 $\lambda = 0.06$ 。

[0068] ②-6、根据二维视频的 H 个 X-T 截面图像各自的最优水平方向运动矩阵,通过在 Y 方向上进行叠加,构造得到 X-T 截面图像的立方体,记为 S_{cubeXT} ,其中, S_{cubeXT} 的维数为 $W \times H \times T$ 。

[0069] ②-7、计算 $\text{norm}(S_{\text{cubeXT}} \cdot S_{\text{cubeYT}})$ 作为二维视频的初步的运动显著图,其中, $\text{norm}()$ 表示归一化操作, $\text{norm}(S_{\text{cubeXT}} \cdot S_{\text{cubeYT}})$ 中的符号“ \cdot ”为矩阵点乘符号。

[0070] ②-8、采用高斯滤波器对二维视频的初步的运动显著图中每个时刻的 X-Y 截

面图像进行平滑操作,得到二维视频中每个时刻的视频帧的运动显著图,将二维视频中的 t 时刻的视频帧 $\{I_t(x, y)\}$ 的运动显著图记为 $\{S_{t, \text{motion}}(x, y)\}$, 将 $\{S_{t, \text{motion}}(x, y)\}$ 中坐标位置为 (x, y) 的像素点的像素值记为 $S_{t, \text{motion}}(x, y)$, $S_{t, \text{motion}}(x, y) = S'_{t, \text{motion}}(x, y) * G_\sigma(x, y)$, 其中, $S'_{t, \text{motion}}(x, y)$ 表示二维视频的初步的运动显著图中的 t 时刻的 X-Y 截面图像, “*” 为卷积操作符号, $G_\sigma(x, y)$ 表示标准差为 σ 的高斯函数, 在本实施例中取

$$\sigma = 5, G_\sigma(x, y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(x^2 + y^2)}{2\sigma^2}\right), \exp() \text{ 为以 } e \text{ 为底的指数函数, } e \text{ 表示自然基数,}$$

$e = 2.718281828$ 。

[0071] ③由于静态场景也会吸引人的视觉注意力,但在静态场景中,各种运动几乎为零,无法采用基于运动的方法获得静态注意力区域,因此本发明方法提取二维视频中每个时刻的视频帧中的每个像素点的特征矢量;然后对提取得到的所有像素点的特征矢量构成的矩阵进行低秩矩阵分解,并获取二维视频中每个时刻的视频帧中的所有像素点的特征矢量构成的矩阵的最优背景矩阵和最优运动矩阵;再根据二维视频中每个时刻的视频帧中的所有像素点的特征矢量构成的矩阵的最优运动矩阵,获得二维视频中每个时刻的视频帧的空间显著图,将二维视频中的 t 时刻的视频帧 $\{I_t(x, y)\}$ 的空间显著图记为 $\{S_{t, \text{spatial}}(x, y)\}$, 其中, $S_{t, \text{spatial}}(x, y)$ 表示 $\{S_{t, \text{spatial}}(x, y)\}$ 中坐标位置为 (x, y) 的像素点的像素值。

[0072] 在此具体实施例中,步骤③的具体过程为:

[0073] ③-1、将当前正在处理的二维视频中的 t 时刻的视频帧 $\{I_t(x, y)\}$ 定义为当前视频帧。

[0074] ③-2、提取当前视频帧中的每个像素点的红颜色分量、绿颜色分量、蓝颜色分量、色调分量和饱和度分量,由提取出的每个像素点的红颜色分量、绿颜色分量、蓝颜色分量、色调分量和饱和度分量构成对应像素点的第一特征矢量,其中,当前视频帧中的每个像素点的第一特征矢量的维数为 $5 \times (W \times H)$ 。

[0075] ③-3、提取当前视频帧中的每个像素点经 4 个方向和 3 个尺度的可控金字塔 (steerable pyramids) 滤波后得到的振幅,由提取出的每个像素点对应的 12 个振幅构成对应像素点的第二特征矢量,其中,当前视频帧中的每个像素点的第二特征矢量的维数为 $12 \times (W \times H)$ 。

[0076] ③-4、提取当前视频帧中的每个像素点经 12 个方向和 3 个尺度的 Gabor 滤波后得到的振幅,由提取出的每个像素点对应的 36 个振幅构成对应像素点的第三特征矢量,其中,当前视频帧中的每个像素点的第三特征矢量的维数为 $36 \times (W \times H)$ 。

[0077] ③-5、将当前视频帧中的每个像素点的第一特征矢量、第二特征矢量和第三特征矢量按序重组构成对应像素点的特征矢量,然后对当前视频帧中的所有像素点的特征矢量构成的矩阵进行低秩矩阵分解,得到当前视频帧中的所有像素点的特征矢量构成的矩阵的所有背景矩阵和所有运动矩阵。

[0078] ③-6、利用最小化概率密度函数,获取当前视频帧中的所有像素点的特征矢量构成的矩阵的最优背景矩阵和最优运动矩阵,对应记为 \mathbf{B}_t^* 和 \mathbf{M}_t^* , 将 \mathbf{B}_t^* 和 \mathbf{M}_t^* 的组合记为

$$(\mathbf{B}_t^*, \mathbf{M}_t^*), (\mathbf{B}_t^*, \mathbf{M}_t^*) = \arg \min_{(\mathbf{B}_t, \mathbf{M}_t) \in \Omega_t} [\|\mathbf{B}_t\|_* + \lambda \|\mathbf{M}_t\|_1], \text{ 其中, } F \text{ 为当前视频帧中的所有像素点的特征}$$

s.t. $F = \mathbf{B}_t + \mathbf{M}_t$

矢量构成的矩阵, $\mathbf{F} = \mathbf{B}_t^* + \mathbf{M}_t^*$, \mathbf{F} 、 \mathbf{B}_t^* 和 \mathbf{M}_t^* 的维数均为 $53 \times (W \times H)$, $\text{argmin}[\cdot]$ 表示最小化概率密度函数, Ω_t 表示对 \mathbf{F} 进行低秩矩阵分解得到的所有的背景矩阵和运动矩阵的组合 $(\mathbf{B}_t, \mathbf{M}_t)$ 的集合, $(\mathbf{B}_t, \mathbf{M}_t) \in \Omega_t$, 符号 “ $\|\cdot\|_*$ ” 为求取矩阵核的范数符号, 符号 “ $\|\cdot\|_1$ ” 为求取矩阵的 1-范数符号, λ 为拉格朗日乘子, 在本实施例中取 $\lambda = 0.06$ 。

[0079] ③-7、采用超像素 (Superpixel) 分割技术将当前视频帧分割成 M 个互不重叠的区域, 然后将 \mathbf{M}_t^* 表示为 $\mathbf{M}_t^* = [(\mathbf{M}_{t,1}^*)^T, (\mathbf{M}_{t,2}^*)^T, \dots, (\mathbf{M}_{t,M}^*)^T]^T$, 其中, $M \geq 1$, $\mathbf{M}_{t,1}^*$ 表示由 \mathbf{M}_t^* 中属于第 1 个区域内的所有像素点的特征矢量构成的矩阵, $\mathbf{M}_{t,2}^*$ 表示由 \mathbf{M}_t^* 中属于第 2 个区域内的所有像素点的特征矢量构成的矩阵, $\mathbf{M}_{t,M}^*$ 表示由 \mathbf{M}_t^* 中属于第 M 个区域内的所有像素点的特征矢量构成的矩阵, $(\mathbf{M}_{t,1}^*)^T$ 为 $\mathbf{M}_{t,1}^*$ 的转置矢量, $(\mathbf{M}_{t,2}^*)^T$ 为 $\mathbf{M}_{t,2}^*$ 的转置矢量, $(\mathbf{M}_{t,M}^*)^T$ 为 $\mathbf{M}_{t,M}^*$ 的转置矢量, $[(\mathbf{M}_{t,1}^*)^T, (\mathbf{M}_{t,2}^*)^T, \dots, (\mathbf{M}_{t,M}^*)^T]^T$ 为 $[(\mathbf{M}_{t,1}^*)^T, (\mathbf{M}_{t,2}^*)^T, \dots, (\mathbf{M}_{t,M}^*)^T]$ 的转置矢量, 符号 “ $[\cdot]$ ” 为矢量表示符号, 在本实施例中取 $M=200$ 。

[0080] ③-8、计算当前视频帧中的每个区域的空间显著值, 将当前视频帧中的第 k 个区域的空间显著值记为 $f_{t,k}$, $f_{t,k} = \|\mathbf{M}_{t,k}^*\|$, 然后将当前视频帧中的每个区域的空间显著值作为对应区域内的所有像素点的空间显著值, 得到当前视频帧中的每个像素点的空间显著值, 从而得到当前视频帧的空间显著图, 记为 $\{S_{t,\text{spatial}}(x, y)\}$, 其中, $1 \leq k \leq M$, $\mathbf{M}_{t,k}^*$ 表示由 \mathbf{M}_t^* 中属于第 k 个区域内的所有像素点的特征矢量构成的矩阵, $S_{t,\text{spatial}}(x, y)$ 表示 $\{S_{t,\text{spatial}}(x, y)\}$ 中坐标位置为 (x, y) 的像素点的像素值。

[0081] ③-9、令 $t=t+1$, 将二维视频中下一个时刻的视频帧作为当前视频帧, 然后返回步骤③-2 继续执行, 直至二维视频中的所有视频帧处理完毕, 得到二维视频中每个时刻的视频帧的空间显著图, 其中, $t=t+1$ 中的 “=” 为赋值符号。

[0082] ④对二维视频中每个时刻的视频帧的运动显著图和空间显著图进行融合, 得到二维视频中每个时刻的视频帧的视频显著图, 将二维视频中 t 时刻的视频帧 $\{I_t(x, y)\}$ 的视频显著图记为 $\{S_{\text{video}}(x, y)\}$, 将 $\{S_{\text{video}}(x, y)\}$ 中坐标位置为 (x, y) 的像素点的像素值记为 $S_{\text{video}}(x, y)$, $S_{\text{video}}(x, y) = (S_{t,\text{motion}}(x, y))^\alpha \times (S_{t,\text{spatial}}(x, y))^{1-\alpha}$, 其中, α 为权重系数, 在本实施例中取 $\alpha = 0.3$ 。

[0083] 以下就利用本发明方法对比利时蒙斯理工学院 TCTS 实验室提供的 ASCMN 数据库中 video5、video8、video10 和 video21 四个视频序列的视频显著图进行提取。图 2a 给出了 “video5” 视频序列的第 100 帧视频图像, 图 2b 给出了 “video5” 视频序列的第 100 帧视频图像的运动显著图, 图 2c 给出了 “video5” 视频序列的第 100 帧视频图像的空间显著图, 图 2d 给出了 “video5” 视频序列的第 100 帧视频图像的视频显著图; 图 3a 给出了 “video8” 视频序列的第 100 帧视频图像, 图 3b 给出了 “video8” 视频序列的第 100 帧视频图像的运动显著图, 图 3c 给出了 “video8” 视频序列的第 100 帧视频图像的空间显著图, 图 3d 给出了 “video8” 视频序列的第 100 帧视频图像的视频显著图; 图 4a 给出了 “video10” 视频序列的第 100 帧视频图像, 图 4b 给出了 “video10” 视频序列的第 100 帧视频图像的运动显著图, 图 4c 给出了 “video10” 视频序列的第 100 帧视频图像的空间显著图, 图 4d 给出了 “video10”

视频序列的第 100 帧视频图像的视频显著图；图 5a 给出了“video21”视频序列的第 100 帧视频图像，图 5b 给出了“video21”视频序列的第 100 帧视频图像的运动显著图，图 5c 给出了“video21”视频序列的第 100 帧视频图像的空间显著图，图 5d 给出了“video21”视频序列的第 100 帧视频图像的视频显著图。从图 2a 至图 5d 可以看出，采用本发明方法得到的视频显著图由于考虑了空间显著图和运动显著图，因此能够较好地反映视频的静态和动态的显著变化情况，符合视频显著语义的特征。

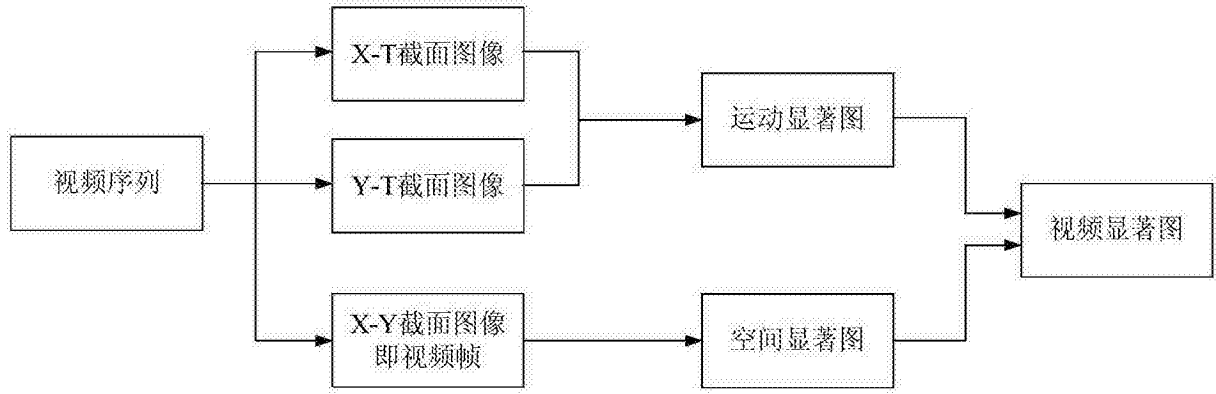


图 1

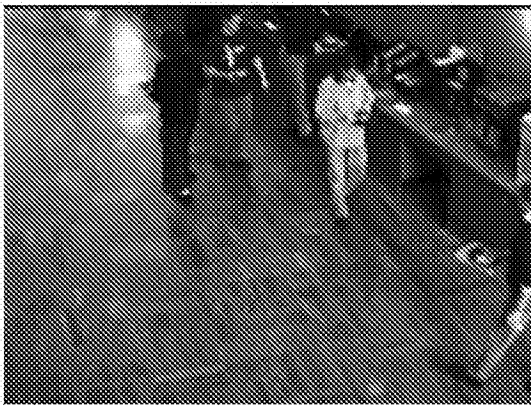


图 2a



图 2b

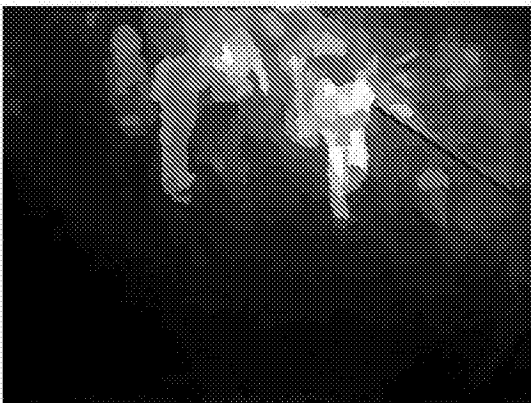


图 2c

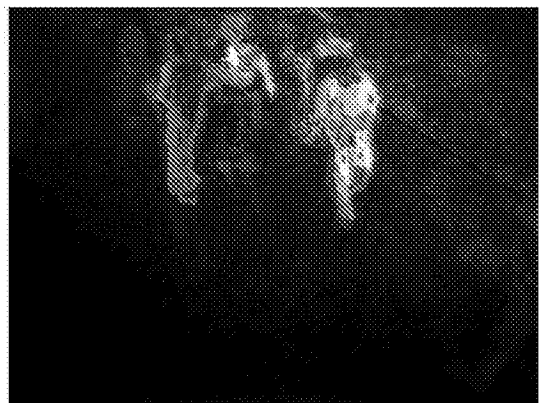


图 2d

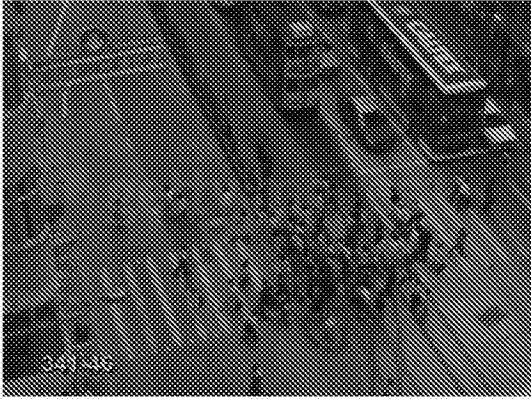


图 3a

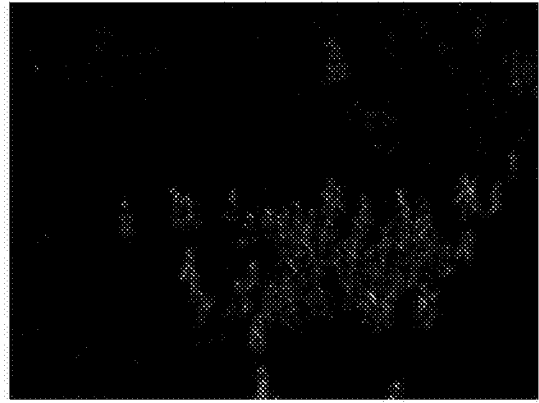


图 3b

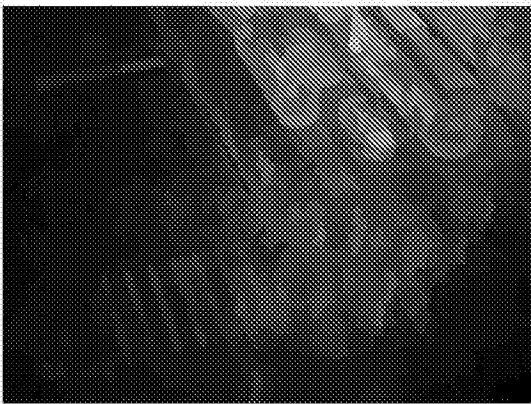


图 3c

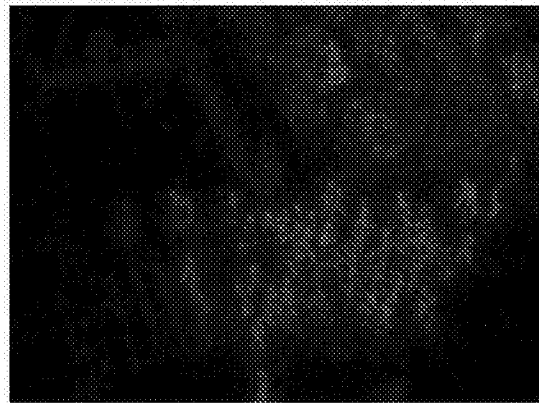


图 3d



图 4a

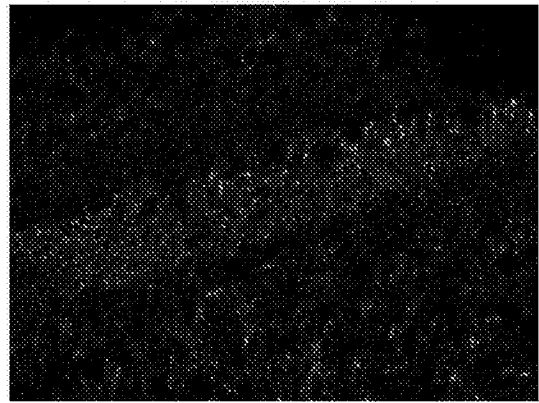


图 4b

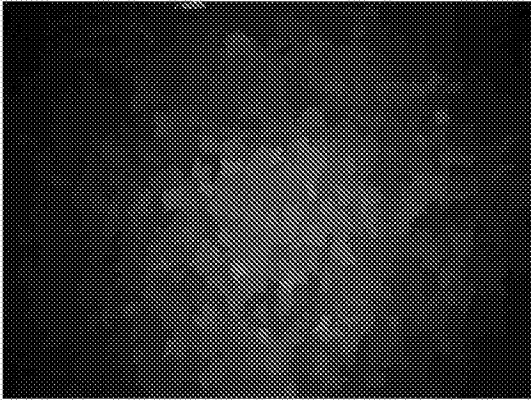


图 4c

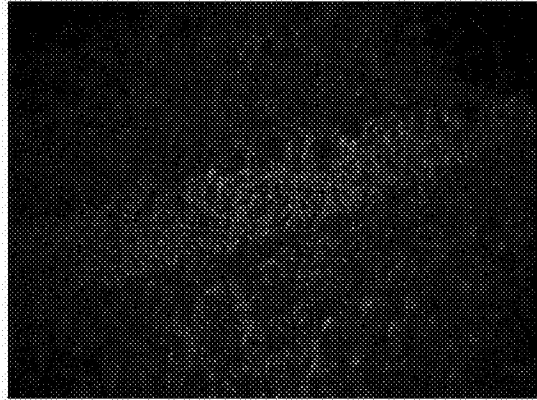


图 4d

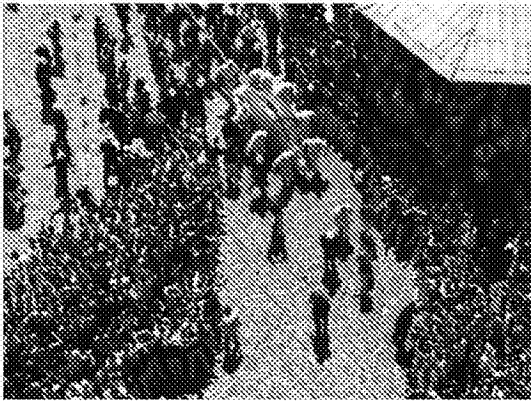


图 5a

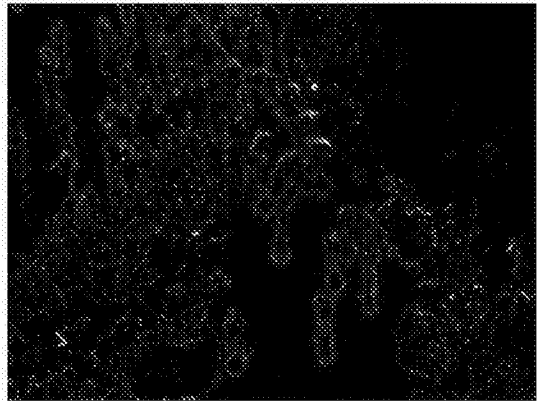


图 5b

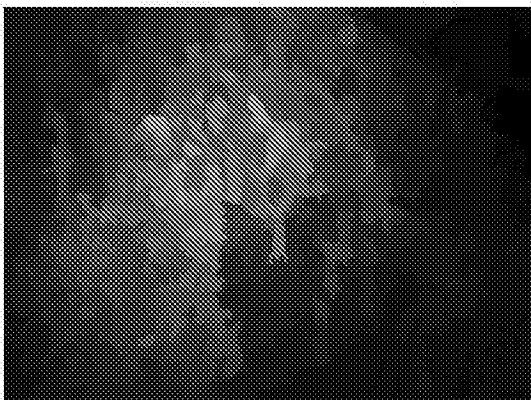


图 5c

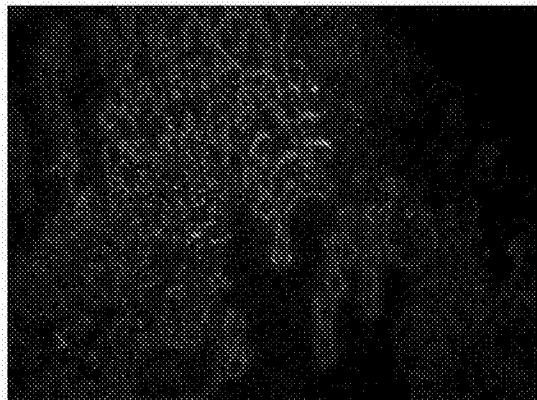


图 5d