Stereoscopic Image Generation Based on Depth Images for 3D TV

Liang Zhang and Wa James Tam

***Abstract*—A depth-image-based rendering system for generating stereoscopic images is proposed. One important aspect of the proposed system is that the depth maps are pre-processed using an asymmetric filter to smoothen the sharp changes in depth at object boundaries. In addition to ameliorating the effects of blocky artifacts and other distortions contained in the depth maps, the smoothing reduces or completely removes newly exposed (disocclusion) areas where potential artifacts can arise from image warping which is needed to generate images from new viewpoints. The asymmetric nature of the filter reduces the amount of geometric distortion that might be perceived otherwise. We present some results to show that the proposed system provides an improvement in image quality of stereoscopic virtual views while maintaining reasonably good depth quality.**

***Index Terms*—Asymmetric filter, depth-image-based rendering, stereoscopic image, stereoscopic image generation, three-dimensional television.**

I. Introduction

DEPTH-IMAGE-BASED Rendering (DIBR) techniques have recently received much attention in the broadcast research community as a promising technology for three-dimensional television (3D TV) systems [1]–[3]. Whereas, the classical approach requires the transmission of two streams of video images [4], [5], one for each eye, 3D TV systems based on DIBR will require a single stream of monoscopic images and a second stream of associated images, usually termed depth images or depth maps, that convey per-pixel depth information. A depth map is essentially a two-dimensional (2D) function that gives the depth, with respect to the camera position, of a point in the visual scene as a function of the image coordinates. Since the depth of every point in an original image is known, a virtual image of any nearby viewpoint can be rendered by projecting the pixels of the original image to their proper 3D locations and re-projecting them onto the virtual image plane. Thus, DIBR permits the creation of novel images, using information from the depth maps, as if they were captured with a camera from different viewpoints. A further advantage of the DIBR approach is that depth maps can be coded more efficiently than two streams of natural images, thereby reducing the bandwidth required for transmission. In this vein, it is not only suitable for 3D TV but also for other 3D applications such as multimedia systems [6].

Manuscript received August 30, 2004; revised December 22, 2004.

The authors are with the Communications Research Centre Canada, Ottawa, Ontario, Canada, K2H 8S2 (e-mail: liang.zhang@crc.ca; james.tam@crc.ca).

One disadvantage of the DIBR approach is that with this type of data representation, one or more "virtual" images of the 3D scene have to be generated at the receiver side in real time. In addition, it is not an easy task to create new, virtual, images with high image quality.

The most significant problem in DIBR is how to deal with newly exposed areas (holes) appearing in the virtual images. Holes are due to the accretion (disocclusion) of portions/regions of objects or background that would have been visible only from the new viewpoint but not from the original location that was used in capturing the original image. There is no information in the original image for these disoccluded regions and, therefore, they would appear empty, like holes, in the new virtual image. A simple way to 'fill' these holes is to map a pixel in the original image to several pixels in the virtual image by simple interpolation of pixel information in the foreground or background. More complex extrapolation technique might also be used [3]. However, these filling techniques are known to produce visible disocclusion artifacts in the virtual images, whose severity depends on the scene layout.

To deal with these disocclusion artifacts in the virtual images several approaches have been suggested. One approach, termed the layered-depth-image (LDI) [7], uses a set of original images of a scene and their associated depth maps. The images and depth maps store not only what is visible in the original image, but also what is behind the visible surface. Note that while this approach is likely to produce very accurate virtual images, it is more computationally demanding and it requires more bandwidth for transmission. An alternative approach involves pre-processing of the depth maps. Recently, we adopted this latter approach and pre-processed depth maps using a symmetric 2D Gaussian filter, so that the disocclusion artifacts were incrementally removed as the smoothing of depth maps became stronger [8], [9]. Experimental results using formal subjective evaluation techniques indicated that this technique (*symmetric smoothing*) could be used to significantly improve the image quality of novel stereoscopic views especially when there are blocky artifacts or noise in the depth maps and potential distortions in the newly generated images as a result of disocclusion [8], [9]. The notion of smoothing depth maps to remove disocclusion artifacts has been advocated by other authors as well [10].

In this paper, we propose a new system for stereoscopic image generation based on depth images to deal with the disocclusion artifact in virtual images [11]. Different types of artifacts and distortions that could appear in the virtual images are then exper-

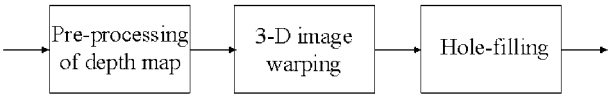

Fig. 1. Flowchart of the proposed depth-image-based rendering system.

on the investigation, we present a new concept of *asymmetric smoothing* of depth maps for DIBR that can reduce artifacts and distortions in the virtual images and provide an improvement in the image quality.

The remaining portions of this paper are organized as follow. In Section II, we illustrate the proposed rendering system. Section III is devoted to experimental investigation with different system setups. In Section IV we propose the concept of *asymmetric smoothing* of depth maps. Section V provides a discussion of experimental results using natural depth images. Conclusions can be found in Section VI.

II. Depth-Image-Based Rendering System

A flowchart describing the proposed depth-image-based rendering system is illustrated in Fig. 1. This system consists of three parts: (i) pre-processing of depth maps, (ii) 3D image warping and (iii) hole-filling. Note that part (iii) is not necessary if there are no holes to fill as a result of optimal pre-processing of depth maps. In the following, these three parts will be addressed in detail.

A. Pre-Processing of Depth Maps

The pre-processing of depth maps includes two issues: choosing the *convergence distance* Z_c (so-called *zero-parallax setting* (*ZPS*)) [12] and smoothing the depth maps.

There are several methods that can be used to establish a *ZPS* [12]. In the so-called *toed-in* approach, the *ZPS* is chosen by a joint inward rotation of the left-eye and right-eye cameras. In the so-called *shift-sensor* approach, a plane of convergence is established by a small lateral shift h of the CCD sensors in the pair of parallel cameras. Different from these two methods, in the present rendering system the *ZPS* is chosen by "shifting" the depth map. Without loss of generality, we choose

$$Z_c = \frac{(Z_{\text{near}} - Z_{\text{far}})}{2}$$

as the *ZPS* plane, where Z_{near} and Z_{far} are the nearest clipping plane and the farthest clipping plane of the depth map. In an 8-bit depth map, $Z_{\text{near}} = 255$ and $Z_{\text{far}} = 0$ (Fig. 2). After that, the depth map is further normalized with the factor of 255, so that the values of the depth map lie in the interval of $[-0.5, 0.5]$, values that are required by the image-warping algorithm.

The second issue in the pre-processing step is to smooth the depth maps. To this end, different filter types can be used. For simplicity, a Gaussian filter $g(x, \sigma)$

$$g(x,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{x^2}{\sigma^2}\right\}, \quad \text{for} \quad -\frac{w}{2} \le x \le \frac{w}{2}, \quad (1)$$

is employed, where w is the filter's window size and σ the standard deviation. Then, depth value $\hat{s}(x, y)$ after smoothing using a Gaussian filter is equal to

$$\frac{\sum_{\upsilon=-\frac{w}{2}}^{\frac{w}{2}} \left\{ \sum_{\mu=-\frac{w}{2}}^{\frac{w}{2}} \left(s(x-\mu, y-\upsilon) g(\mu, \sigma_\mu)\right) g(\upsilon, \sigma_\upsilon) \right\}}{\sum_{\upsilon=-\frac{w}{2}}^{\frac{w}{2}} \left\{ \sum_{\mu=-\frac{w}{2}}^{\frac{w}{2}} g(\mu, \sigma_\mu) g(\upsilon, \sigma_\upsilon) \right\}}. \quad (2)$$

It is expected that different values of w and σ have different impact on the quality of the virtual images generated from the original center image. We will discuss this issue in the next section. In the following experiment, we let the filter's window size w be equal to 3σ.

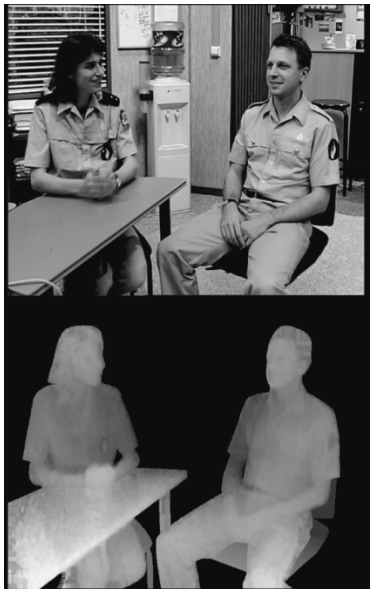

Fig. 2. The test image: "Interview." The original image is on the top and its associated unprocessed depth map is on the bottom. A lower luminance value in the depth map means that the objects are farther away from the camera.

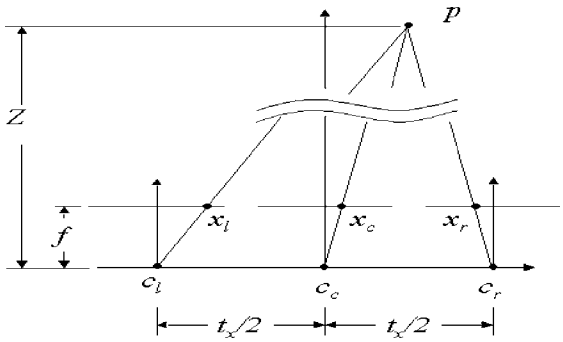

Fig. 3. Camera configuration used for generation of virtual stereoscopic images.

B. 3-D Image Warping

For simplicity, we only consider the commonly used parallel camera configuration for generating virtual stereoscopic images from one center image associated with one depth map for 3D TV (Fig. 3). In this case, the vertical coordinate of the projection

and c_r the viewpoint of the virtual left-eye and right-eye images to be generated. f is the focal length of three cameras. t_x is the baseline distance between two virtual cameras.

Under this camera configuration, one point p with the depth Z in the world (of dimensions X, Y, Z) is projected onto the image plane of three cameras at pixel (x_l, y), (x_c, y) and (x_r, y), respectively. From the geometry shown in Fig. 3, we have

$$x_l = x_c + \frac{t_x}{2}\frac{f}{Z} \quad x_r = x_c - \frac{t_x}{2}\frac{f}{Z} \tag{3}$$

where information about x_c and Z is given in the center image and the associated depth map, respectively. Therefore, with formulation (3) for 3D image warping, the virtual left-eye and right-eye images can be generated from the original center image and its depth map by providing the value of the baseline distance t_x and focal length f. Without loss of generality we choose the focal length f to be equal to one in the experiments.

Based on the *ZPS* defined in Section II and pre-processing of depth maps, the value of the baseline distance t_x also indicates the depth range appearing in the generated stereoscopic image pair. According to the image warping formulation (3), the disparity $(x_l - x_r)$ involved in the rendered left-eye and right-eye images is proportional to the baseline distance t_x. A large disparity value indicates that the object point in the real world is far away from the *ZPS*, while a small value means that the object point is close to the *ZPS*.

C. Disocclusion and Hole-Filling

Due to a difference in viewpoints, some areas that are occluded in the original image might become visible in the virtual left-eye or the right-eye images. These newly exposed areas, referred to as "disocclusion" in the computer graphics literature, have no texture after 3D image warping because information about the disocclusion area is available neither in the center image nor in the accompanying depth map. We fill in the newly exposed areas by averaging textures from neighborhood pixels, and this process is called *hole-filling*.

III. Investigation of Different System Setups

This section is devoted to investigating the performance of the proposed rendering system using natural images. As an example, only results with the test image "Interview" (Fig. 2) are shown. The image and its corresponding depth map were generously supplied by Fraunhofer HHI (Heinrich-Hertz-Institut), Germany. In the following investigation, the distance between the two virtual left-eye and right-eye cameras is fixed at 48 pixels for illustration.

First, we investigate the performance of this system *without* smoothing the depth maps. Fig. 4 shows an example of the results of the virtual left-eye image. The depth map without pre-processing is shown in Fig. 4(a). The image after 3D image warping is illustrated in Fig. 4(b). In the figure, the white areas, i.e., the holes, are the newly exposed areas. Recall that these holes are produced because information about the previously

Fig. 4. Virtual left-eye image generated *without* smoothing of the depth map. (a) Depth map without smoothing; (b) image after 3D image warping. White areas represent newly exposed areas; (c) image after *hole-filling*; (d) and (e) artifacts clearly seen in enlarged segments of the image from (c).

From the figure, we can see that the newly exposed areas are located mainly along the boundaries of objects and also the right margin of the whole image. After *hole-filling*, as shown in Fig. 4(c), significant texture artifacts appear at object boundaries in the virtual image. Fig. 4(d) and (e) show more clearly the artifacts by enlarging segments of Fig. 4(c).

We then evaluate the performance of the proposed system *with* smoothing of the depth maps. Similar to [8], [9], we let σ_μ and σ_υ of two Gaussian filters that are applied separately along the vertical and horizontal directions, respectively, have the same value. We term this process *symmetric smoothing*. Fig. 5 shows the results of the virtual left-eye image with $\sigma_\mu = \sigma_\upsilon = 30$. The depth map after symmetric smoothing is shown in Fig. 5(a). The image after 3D image warping is illustrated in Fig. 5(b). White areas represent newly exposed areas. Compared to Fig. 4(b), we can see from Fig. 5(b) that the newly exposed areas along object boundaries almost disappear except for the area in the right margin of the whole image. This can be explained as follows. Due to the smoothing of the depth map, there are no more sharp depth discontinuities. In other words, the disocclusion areas have become sparse because of smoothing and even disappear as the smoothing becomes stronger. Fig. 6 shows the relation between the newly exposed areas and the depth smoothing strength (as determined by σ) in the virtual left-eye

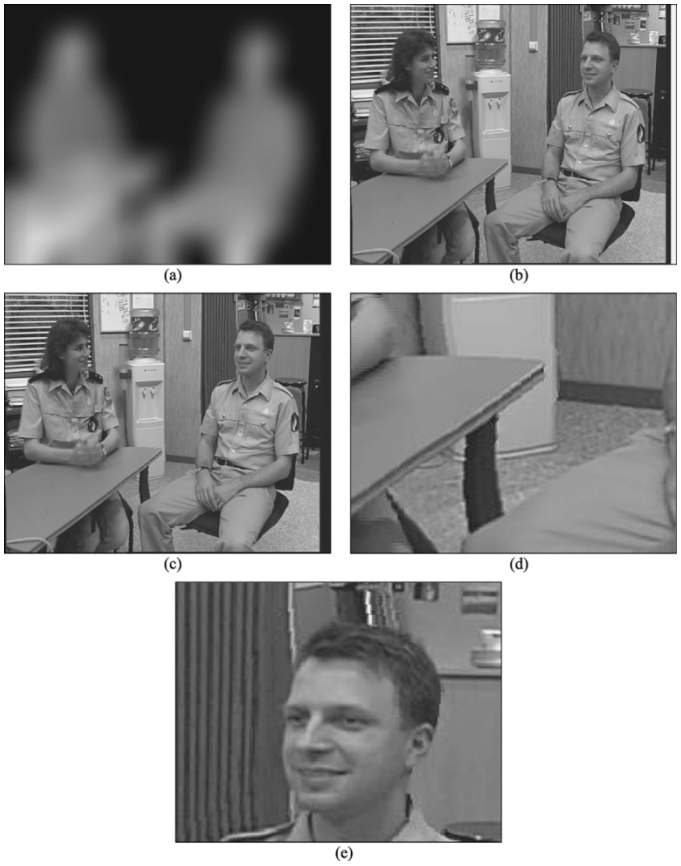

Fig. 5. Intermediate steps in the generation of a virtual left-eye image using *symmetric smoothing* of the depth map with $\sigma_\mu = \sigma_v = 30$. (a) Depth map after symmetric smoothing; (b) image after 3D image warping. White areas (right margin of image) are newly exposed areas; (c) image after *hole-filling*; (d) and (e) enlarged segments of the image shown in (b). Notice the curved table leg in (d) and the curved vertical lines in (e), even though the overall output is significantly better than that without smoothing [cf. Fig. 4(e)].

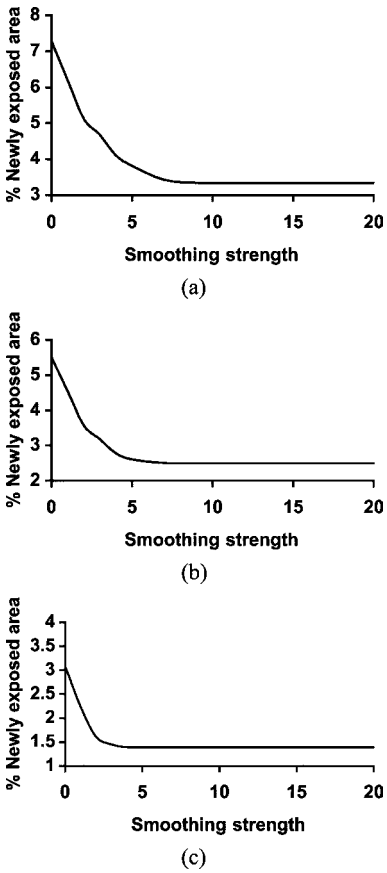

Fig. 6. Relation between depth smoothing strength and newly exposed areas as a percentage of the total area for the test image "Interview." Three graphs are shown for the three baseline distances that were used: (a) 48 pixels, (b) 36 pixels, and (c) 20 pixels. Note that in each figure the percentage of newly exposed area decreases with smoothing strength.

pixels over the total number of pixels in the image. We term this ratio the disocclusion ratio. As the depth smoothing strength becomes stronger, the disocclusion ratio decreases gradually until it reaches a constant value. This constant value is simply due to the persistence of newly exposed areas at the image margins. Also, from Fig. 6 it can be seen that the minimum depth smoothing strength to reach a constant value of newly exposed areas is dependent on the baseline distances t_x. For the test image "Interview," it is approximately one quarter of the baseline distance.

Comparison of Fig. 4 to Fig. 5 shows that simple *hole-filling* produces significant *texture artifacts* whereas *symmetric smoothing* virtually eliminates these artifacts. However, *symmetric smoothing* still produces some distortion. Specifically, vertically straight object boundaries now can become curved, depending on the depth in neighboring regions. This can be more clearly seen in Fig. 5(d) and (e), which show enlarged segments of the image in Fig. 5(c). We call this type of distortion, *geometric distortion*. The origin of this type of distortion can be explained as follows. Let us examine the table leg in Fig. 5(d). In the unprocessed depth map, this table leg has the same depth value along its vertical length and, at the bottom of processing, due to smoothing of the depth map in the horizontal direction and at a level that is as strong as that in the vertical direction, the bottom of the table leg has a slightly larger value than that of its top [Fig. 5(a)]. This creates a curved table leg after 3D image warping.

IV. Asymmetric Smoothing of Depth Map

The analysis of the underlying reason for the geometric distortions in the previous section suggests that the strength of the smoothing of depth maps in the horizontal direction should be less than that of the smoothing in the vertical direction, so that vertical objects, e.g. the table leg, have similar depth values throughout after depth smoothing. We call this *asymmetric smoothing*.

The concept of *asymmetric smoothing* is consistent with known characteristics of the binocular system of the human eyes. The human visual system obtains depth cues from disparity mainly from horizontal differences rather than vertical differences between the images that are projected to the left and the right eyes. This allows us to filter the depth map stronger in the vertical than in the horizontal direction. In other words, we can use an asymmetric filter to smoothen the sharp depth changes in a manner that will overcome the disocclusion

Fig. 7. Virtual images generated using *asymmetric smoothing* of the depth map with $\sigma_\mu = 10$ and $\sigma_\upsilon = 90$. (a) Depth map after asymmetric smoothing; (b) virtual left-eye image; (c) virtual right-eye image; (d) and (e) enlarged segments from the image shown in (b). Note vertical lines are now straight compared to Fig. 5(e).

Fig. 7 shows the results of rendering using *asymmetric smoothing* of the depth map with $\sigma_\mu = 10$ and $\sigma_\upsilon = 90$. The depth map after asymmetric smoothing is shown in Fig. 7(a). The virtual left-eye and right-eye images generated from the original center image (Fig. 2) and the processed depth map are shown in Fig. 7(b) and (c). Two enlarged segments from the left-eye image are shown in Fig. 7(d) and (e) for clarity. It can be seen from Fig. 7(d) and (e) that *geometric distortions* are strongly reduced compared to Fig. 5(d) and (e). Also, no *texture artifacts* appear. In general *asymmetric smoothing* results in virtual images that have sharper texture and higher image quality. When viewed in a stereoscopic display, they also create reasonably good and stable depth.

V. Experimental Results and Discussions

To further evaluate the performance of the proposed rendering system with asymmetric smoothing, experiments with additional natural depth image sequences were carried out.

A. Test Depth Image Sequences

Samples of three additional stereo video sequences and their

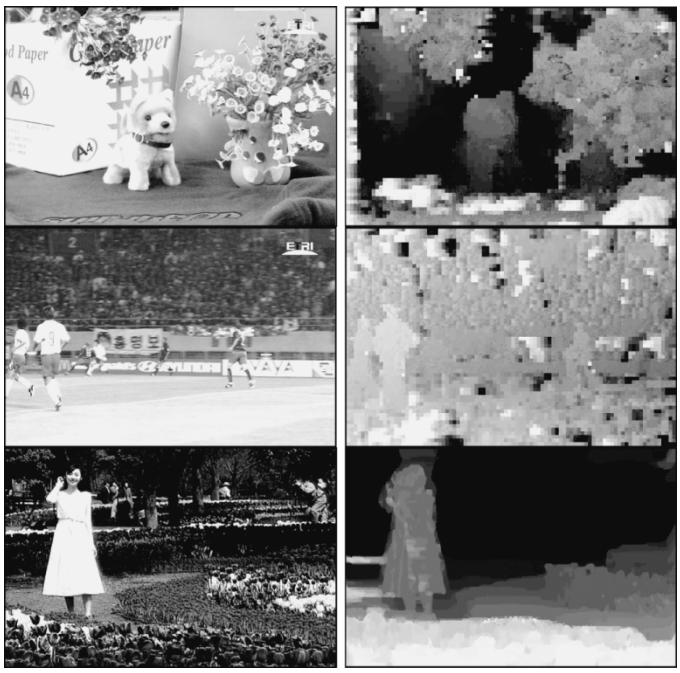

Fig. 8. Sample of three image sequences and their corresponding depth maps. From top to bottom are "Puppy," "Soccer," and "Tulips."

"Soccer," and "Tulips." The depth maps for the first two images were obtained from the same institution that generously provided the source images [Electronics and Telecommunications Research Institute (ETRI), Korea]. These depth maps had 8×8-block depth resolution and were not as stable in that there were blocky artifacts that appeared and disappeared over time. The depth map of the third sequence, "Tulips," was estimated using our own in-house developed software for disparity estimation [13]. It had pixel depth resolution with pel accuracy and was relatively stable, although it contained some inaccuracies appearing in the left side of the walking woman. The image size of all of the depth maps was 720×480 pixels.

B. Parameter Selection

The depth range for the rendered virtual stereoscopic images was selected so that the image was comfortable to view. Several studies suggest [14] that the maximum depth range that is still comfortable for viewing is 1° disparity or approximately 5% of the width of a standard 4×3 image at a viewing distance of 4H (four times image height). Therefore, we chose the baseline distance of 36 pixels to render the virtual left-eye and right-eye images for our images, which had a spatial resolution of 720×480. For depth smoothing strength, we chose σ_μ to be equal to 9 using the empirical relation found in Section III that the depth smoothing strength is approximately equal to one quarter of the baseline distance. In the case of symmetric smoothing, the smoothing strength in the vertical direction was the same as that in the horizontal direction. In the case of asymmetric smoothing, the smoothing strength was chosen to be five times the value in

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.