Automatic Projector Calibration Using Self-Identifying Patterns

Mark Fiala
National Research Council of Canada
NRC 1200 Montreal RD, Ottawa, Canada K1A-0R6
mark.fiala@nrc-cnrc.gc.ca

Abstract

Calibrating multiple monitor or projector display elements to provide a composite image can be a time-consuming task if performed manually. Ideally the user would like to roughly aim a number of projectors at a surface, define the desired display corners, and have some automatic method to align the display. A digital camera and computer vision can be used to calibrate the projectors with the assistance of self-identifying patterns*. To account for distortion effects and to equalize brightness, it is desirable to know the mapping of many points within each projector image. A small set of images can be projected from each display element if a self-identifying pattern is used. An array of* ARTag *markers are used as a self-identifying pattern which is displayed in turn by each of the display monitors or projectors and recognized in the camera image. In this way an ad-hoc arrangement of projectors can be calibrated in seconds. Experimental results are shown validating this architecture.*

1. Introduction

A large scale display can be created from an array of display monitors or by a number of projectors aiming onto a common projections surface. Using several display elements (projectors or monitors) together to create a seamless combined image is useful in such applications as *display walls* and where it is advantageous to use an array of less expensive units for a large display area than a single high power projector. The calibration of the spatial *geometric* and intensity *photometric* parameters to produce a seamless mosaic is a challenging task.

Using a camera or cameras and computer vision to automatically align projector or monitor arrays is an attractive method to avoid the complex and tedious task of attempting to set up a composite display [8]. Raskar *et al* [7] use cameras to achieve composite displays using structured light methods.

Our method described hereing uses self-identifying patterns to reduce the number of image frames that need to be displayed on each display element, quite often only two or three calibration patterns per display element. This results in a faster autocalibration time as opposed to binary encoding techniques that need to display more calibration patterns. Our concentrates on planar display surfaces by finding *homographies* instead of projection matrices, but the method can be extended to three-dimensional display surfaces.

In our method the camera is calibrated before hand, a process that itself is automatic and fast. Chen *et al* [2] use an uncalibrated camera to save on this camera calibration step but only use one radial distortion parameter and rely on an iterative annealing process to find the camera parameters. Our camera is calibrated according to Zhang's [10] model with four distortion parameters and no numerical iterations (other than those in Zhang's proven algorithm for the camera) are necessary removing the reliance on a possibly non-converging procedure.

2. System Configuration

With a composite display system there is one video source, a special box to perform the compositing called a *compositor* herein, and a number of output video signals sent to several projectors or monitors. This time varying imagery can be a virtual display and need not actually pass through one of the video protocols, but is described as taking a video signal as input as a practical system implemented today would likely use. The *compositor* is responsible for breaking up this image into sections to give to each projector or monitor it the composite display. The compositor must alter both the spatial location and intensity of the input image content to provide a large composite display that appears as a single large seamless display to the human viewer. This paper proposes a set of algorithms for the compositor to use.

The compositor can be in two modes; auto-calibration mode or running mode. After the individual displays are set up, the system is put into the auto-calibration mode where the system calculates how to later generate the output signals for each display. The system is then put into running mode for the duration of time it is in regular operation viewed by users. The display elements (projectors or monitors) are expected to be stationary throughout, and the system must be put back into auto-calibration mode if the display elements or projection surface is moved.

The system proposed in this paper requires a digital video camera aimed at the composite display area, the camera must be stationary during autocalibration but can be removed when the system is in running mode.

Hardware-wise, the system can consist of a simple hardware engine or computer with multiple video outputs. The complex auto-calibration algorithms can be performed on a laptop computer. The laptop and digital video camera can be removed after calibration is performed.

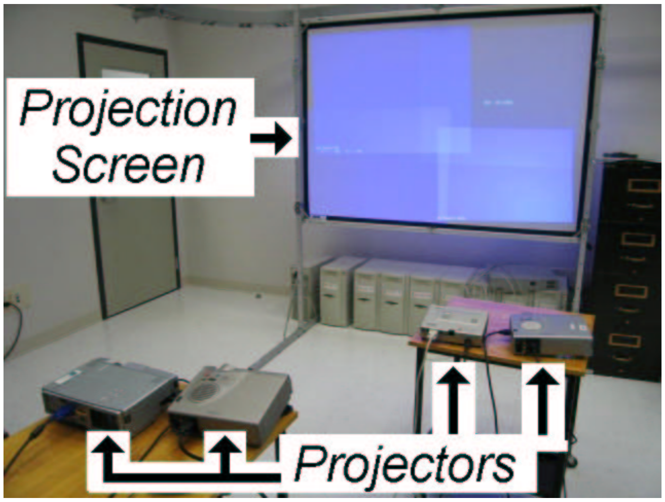

Figure 1: *Composite display. Four different different display elements (projectors or monitors) are loosely arranged, all are showing full white images in this picture. A digital camera (not shown) is used to automatically calibrate the composite display.*

In running mode, the compositor takes the single input video signal as input and outputs N video signals. In autocalibration mode the output signals contain either one of the self-identifying calibration patterns or are all black. The video signals are usually in the form of a VGA, DVI-D, DVI-I, or NTSC. Of these only DVI-D is a digital standard, the others are analog and must be resampled into digital in the *compositor*. This compositor takes one video signal from a computer, DVD/VHS player, or studio video equipment as input and produces N signals as output which drive N projectors or monitors, typically these output signals will be synchronous to the input by using the input's timing information. In video signals, either digital or analog, the image is *serialized* and each image pixel is sent in turn. With a digital signal a pixel's intensity and colour information can be determined by sampling the correct position in the digital stream, or by a certain time after a synchronization pulse in an analog signal.

2.1. Compositing

The compositing operation contains both a *spatial* and *intensity* component corresponding to *geometric* and *photometric* calibration respectively. The spatial position of each point in the input image must be mapped properly to the correct point in one of the output displays to maintain a consistent geometry. The radiance of points in the composite display must be scaled to accomodate the varying brightness between displays, and within each display so that display boundaries are not visible.

The compositing operation would in its most efficient form be performed by a look-up table (LUT) which is has memory storing an entry for each pixel for each of the N outputs. This allows a real time system to use simple hardware or GPU (graphics processor in many modern computers) as opposed to more complex schemes such as in [7] where a two-pass rendering scheme performs 3D graphics operations for every image frame.

With the LUT implementation, a counter generates addresses sequentially $0 \leq i \leq W * H - 1$ (where W, H are the output image width and height in pixels respectively), one for each output pixel $\mathbf{p_{ni}}$ (where n is the projector number $1 \leq n \leq N$), and the contents of the memory at each address is read out. The input image pixels $\mathbf{I}(i)$ usually consist either of a single greyscale or an RGB triple, usually the output pixels $\mathbf{p_{ni}}$ are of the same format. The image warping function is performed by the information contained in this memory entry, the memory entry at location i is used to generate the intensity and colour for $\mathbf{p_{ni}}$. The entry contains a binary flag $b(i)$, one or more image coordinate sets (u, v) and one or more scalar multipliers. The binary flag indicates if the projector should output anything for that pixel or just remain black, the remaining information in the memory entry is ignored if the $b(i)$ value is 'off'. Otherwise, the entry contains either the image coordinates of a single input pixel or neighbhorhood of pixels around $\mathbf{I}(u_{ni}, v_{ni})$. In the simpler single pixel mapping case $\mathbf{p_{ni}}$ is generated (Eqn.1) and a scalar respresenting $\frac{r_{max}}{r_{ni}}$. Pixelating effects may be removed by creating each output pixel $\mathbf{p_{ni}}$ from several input pixels relative to (u_{ni}, v_{ni}) by a set of M coefficients $C_{mni} | 0 \leq m \leq M - 1$ and pixel values at $(\delta u_m, \delta v_m)$ (Eqn. refeqn:composite-neigh). For example, a neighbourhood of $M = 4$ would be used for bilinear interpolation.

$$p_{ni} = \frac{r_{max}}{r_{ni}} \cdot \mathbf{I}(u_{ni}, v_{ni}) \quad (1)$$

$$p_{ni} = \frac{r_{max}}{r_{ni}} \sum_{m=0}^{M-1} C_{mni} \cdot \mathbf{I}(u_{ni} + \delta u_m, v_{ni} + \delta v_m) \quad (2)$$

The spatial calibration requires that the image $\mathbf{I}$ coordinates (u_{ni}, v_{ni}) and binary flag $b(i)$ be found for each pixel position $i = (x_{ni}, y_{ni})$ in each projector/monitor image. The projection of points onto the composite display can be modeled by assuming a projection type and finding parameters, or by simply interpolating between the mesh of known correspondences that the ARTag pattern provides. If the projection is modeled, it is likely modeled as a perfect pinhole projection or a radial/tangential distorted version of a pinhole projection. In the experiment performed herein the projection was modeled as a pinhole model, in which case the mapping can be expressed simply as a *homography* [6]. Homographies can be calculated with a set of points as in the ARTag markers in this paper, or using lines as in the projector system of [1]. This projector-to-source-image homography is labeled $\mathbf{H_{pi}}$ in Eqn.3.

In order to use the camera homography one must assume a perspective (pinhole) projection. The calibrating camera should be calibrated itself, such as with Zhang's [10] and Tsai's [9] algorithms, the popular OpenCV software implements the former and was used in this paper.

$$\alpha \begin{bmatrix} x_{ni} \\ y_{ni} \\ 1 \end{bmatrix} = \mathbf{H_{pi}} \begin{bmatrix} u_{ni} \\ v_{ni} \\ 1 \end{bmatrix}. \quad (3)$$

$$\mathbf{H_{pi}} = \mathbf{H_{ic}^{-1}} \mathbf{H_{pc}} \quad (4)$$

$b(i)$ is set to false (display black) if (u_{ni}, v_{ni}) is outside the dimensions of the source image or if it is being displayed by another projector/monitor. At every point where projected regions overlap some heuristic must be applied, in our case we selected the projector/monitor with the highest resolution. Fig. 7 shows the projection space being divided up among cameras.

The projector-to-source-image homography matrix $\mathbf{H_{pi}}$ is obtained indirectly from the projector-to-camera homography $\mathbf{H_{pc}}$ and the image-to-camera homography $\mathbf{H_{ic}}$ as in Eqn. 4. $\mathbf{H_{pc}}$ is obtained from a least squares fit of the 48 or so ARTag corner points found in the camera image, each ARTag marker is unique and it is known *a priori* what the image coordinates are. None of the ARTag markers are repeated in the pattern. And finally, the image-to-camera homography $\mathbf{H_{ic}}$ is chosen simply from the four corners of the desired composite image window. These corners can be found by manually selecting points in the camera image, or by holding up special large ARTag marker so their corner sets the display corners.

Eqns. 1,2 contain the photometric calibration term $\frac{r_{max}}{r_{ni}}$ where the projector pixel intensity is scaled down to obtain a uniform image. r_{ni} represents the maximum image irradiance that the projector pixel p_{ni} would create in the calibrating camera's image, this can be estimated by fitting a surface to the camera image points known to be at the white level due to their place in the self-identifying pattern. To create uniform intensity, no projector pixel should create a brighter camera image pixel than r_{max}. r_{max} is the maximum intensity of any white point seen in the camera image, all projector pixels' brightness are scaled down to meet the brightness of the least bright pixel. r_{max} is found by finding the maximum level of camera pixels known to be from the projector according to the detected pattern. This can be performed once for the greyscale component of an image, or can be repeated for colour components to compensate for different hues between the projectors.

2.2 Self-identifying patterns: ARTag

Self-identifying patterns are special marker patterns that can be placed in the environment and automatically detected in camera images. Also known as *fiducial marker systems*, a library of these patterns and the algorithms to detect them help to solve the correspondence problem. Self-identifying marker systems such ARToolkit and ARTag are typically used for applications such as calculating camera pose for augmented reality and robot navigation.

ARTag [3, 4, 5] was chosen because of its availability (can be downloaded from[1] and added to user programs), its robustness to lighting variation, its very low false positive detection rate, and its very low inter-marker confusion rate (falsely identifying the marker ID). ARTag fiducials are square planar bi-tonal patterns which have a square border and internal 36 bit digital pattern (Fig. 2).

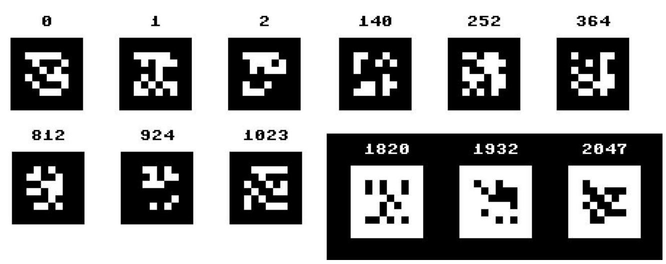

Figure 2: *ARTag markers. ARTag markers are bi-tonal planar marker patterns consisting of a square border and a 6x6 interior grid of cells representing logic '1' or '0'. 12 out of the library of 2002 markers are shown.*

ARTag markers rely only on a planar surface across each individual marker, the markers are detected invariant to affine transforms and homographies. With projector systems perspective distortion is introduced twice, once from the projector-to-screen and secondly from the screen-to-camera. With monitor display systems only one perspective distorion step occurs, however in both situations the geometric distortion can be modeled by a homography for which the ARTag markers are invariant. ARTag markers are also recognizable across a large range of scale, from a minimum of ¡15 pixels to almost the entire image in width.

[1]www.artag.net

ARTag is a bi-tonal system containing 2002 planar markers, each consisting of a square border and an interior region filled with a 6x6 grid of black or white cells. 1001 of ARTag markers have a black square border on a white background, and vice versa for the other 1001. The associated algorithm for detection first locates quadrilaterals which may be perspective views of the marker border, then the interior is sampled into 36 binary '1' or '0' symbols. Further processing is in the digital domain providing a non-linear response giving very low *false positive* and *inter-marker confusion* rates. With ARTag, the probability of falsely identifying one marker for another, or a piece of the background as a marker, is a probability of $< \%0.0039$.

The ARTag self-identifying pattern system was employed in our system to find correspondences in images of the display elements. The calibration image projected/displayed by the display elements is an array pattern of ARTag fiducials at known locations.

Fig.5 shows markers being automatically located in an image. ARTag has some robust features that allow it to detect markers when partly occluded, however this "incomplete marker detection" can be turned off so that only completely seen markers are used as correspondence points.

Fig. 3 shows the self-identifying patterns composed of arrays of ARTag markers.

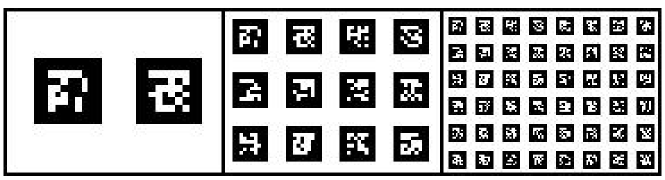

Figure 3: *Three different calibration images to display from each projector, with increasing density of ARTag markers.*

3 Photometric Calibration using the Self-Identifying Pattern

For the photometric calibration term in the equations of Section 2.1, r_{ni} represents the maximum brightness in the camera image that a projector point p_{ni} can create. The ARTag pattern identifies which points are at the full white level. The full black level can also be found. The ARTag markers have a very low false positive detection error rate so it is robust to use points inside the marker to sample the grey or RGB value (Fig.4).

4. Experiment

Four VGA projectors were set up facing a projection screen with their projection regions overlapping as in Fig. 1, and

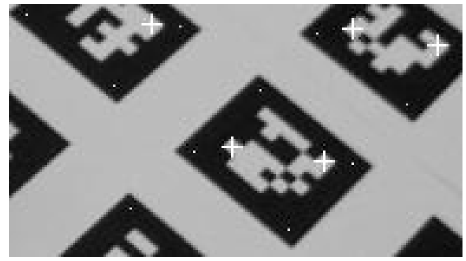

Figure 4: *Sampling of background points. Points inside the ARTag markers are known to belong to the background. Samples of the RGB values are taken for points corresponding to white and black points in the ARTag marker. White samples are shown as white crosses, black samples are shown as white dots. Only completely unoccluded markers are used for sampling.*

a colour digital camera of resolution 1280x1024 was positioned on a tripod viewing the screen. Each display element (projector) showed each of the three self-identifying patterns of Fig. 3 while the other three were projecting black images, these twelve images were processed to find the ARTag patterns

Three calibration images were shown by each projector (Fig. 3) and an image of the screen captured by the digital camera. Twelve images of this process taken by a second camera (not used in the computations) are shown in Fig.6. The images taken by the main calibrated camera were corrected for camera distortion using the OpenCV *cvUnDistortOnce()* function and input to the ARTag marker detection algorithm. An example of the camera image and the automatically detected ARTag markers is shown in Fig.

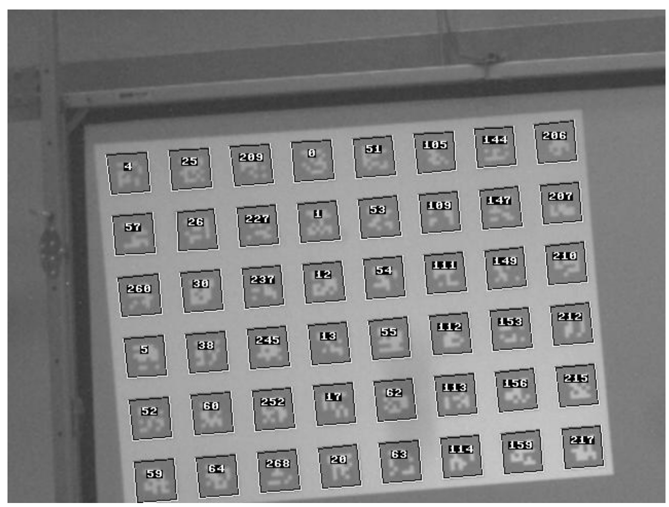

Figure 5: *Detecting ARTag markers in the camera image*

The experiment conducted did not implement the photometric calibration component of the proposed system Also, the projector non-linear distortion was neglected with the projector image to the undistorted camera image mapping assumed to be a homography. Only the camera was calibrated for radial and thin prism distortion.

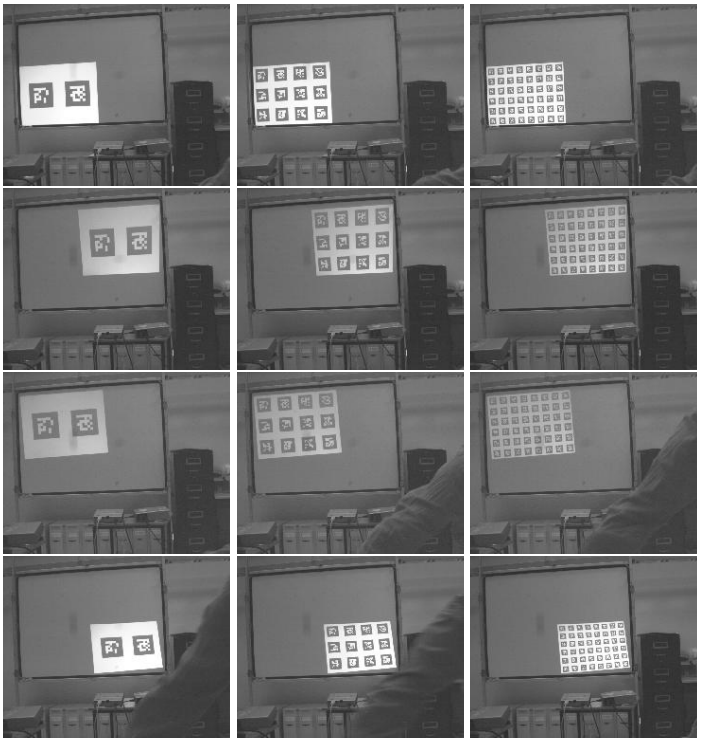

Figure 6: *Each projector displays each self-identifying marker image, three different densities are used in this experiment.*

A group of differing projectors were assembled for the experiment. Two of the projectors were Epson Powerlite 715C, one was a Plus DLP V3-8105F, and one was an Elmo 3LCD Data Projector. The camera was a IEEE-1394 Pixelink PL-A360 fitted with a 16mm lens. The camera was calibrated using the OpenCV *cvCalibrateCamera()* function (the correspondences for calibration were also found using ARTag, a printed pattern 30"x60" was mounted to a table used in our lab for camera calibration).

5. Conclusions

A new composite display system was proposed for creating a single display from a group of projector or monitor display elements. The system has only a few steps and allows for fast real-time compositing with simple hardware. Using several separate projectors together to create a seamless combined image is useful in a number of applications but this calibration can be a tedious and error prone task if performed manually.

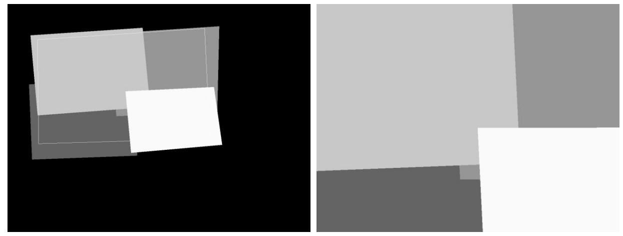

Figure 7: *Dividing the camera image (left) and the composite display image (right) into areas of responsibility for each projector. Black indicates no projector is displaying there, the three shades of grey indicate one of the four projectors are chosen to project to this region*

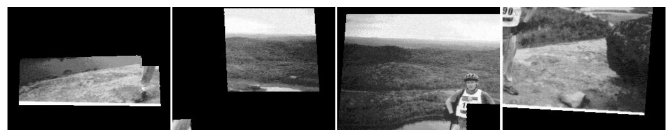

Figure 8: *The images projected by each projector, the areas where no picture is to display is black.*

Figure 9: *The final composite image with all 4 projectors displaying the images in Fig. 8.*

A proposed system was described where a user can very roughly arrange a number of monitors or projectors at a surface, define the desired display boundary, and achieve both geometric and photometric calibration automatically. A digital camera and computer vision was used to recognize patterns projected by the individual display elements. An array of *ARTag* markers was used to form a set of self-identifying patterns which are projected in turn from each display element. The ARTag markers are a reliable solution to finding correspondences and form the novel component of this system. This allows for a fast calibration only using a small number of calibration frames projected or displayed by each display element, as opposed to binary encoding techniques.

This system assumes a planar projection or display surface but can be extended to 3D. However, the planar display surface is a reasonable assumption, especially in projector systems. Using planar displays reduces the projection parameters to a homography which removes some of the ambiguity described in [7] when attempting to model a 3D surface.

An preliminary experiment was performed to validate the feasibility and showed reasonable results for this system still in early development. There are some inaccuracies in the border between the projected images but this is considered by the author to be a minor implementation error and not a flaw in the approach. This system is envisaged to be used in system with a *calibration* and *running mode* with the calibration implemented on a laptop computer and the running mode implemented with a low cost hardware engine. The laptop can be plugged in just for task of calibrating the composite display and can then be removed. In this way a complex set of display elements can be rapidly and robustly auto-calibrated in mere minutes or less.

References

[1] M. Ashdown, M. Flagg, R. Sukthankar, and J. Rehg. A flexible projector-camera system for multi-planar displays. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume II, pages 165–172, 2004.

[2] Y. Chen, D. Clark, A. Finkelstein, T. Housel, and K. Li. Automatic alignment of high-resolution multi-projector displays using an uncalibrated camera. In *Visualization 2000 Proceedings*, pages 125–130, Oct 2000.

[3] M. Fiala. Artag, an improved marker system based on artoolkit. In *National Research Council Publication NRC 47166/ERB-1111*, 2004.

[4] M. Fiala. Artag revision 1, a fiducial marker system using digital techniques. In *National Research Council Publication 47419/ERB-1117*, November 2004.

[5] M. Fiala. Fiducial marker systems for augmented reality: Comparison between artag and artoolkit. In *MIRAGE 2005 :Computer Vision/Computer Graphics Collaboration for Model-based Imaging, Rendering, image Analysis and Graphical special Effects*, INRIA Rocquencourt, France, Mar 2005.

[6] R. Hartley and A. Zisserman. Multiple view geometry in computer vision. Cambridge, UK, 2000. Cambridge University Press.

[7] R. Raskar, M. Brown, R. Yang, W. Chen, G. Welch, H. Towles, W. Seales, and H. Fuchs. Multi-projector displays using camera-based registration. In *Proceedings of IEEE Visualization*, pages 161–168, 1999.

[8] R. Surati. A scalable self-calibrating technology for large scale displays. PhD thesis, MIT, 1999.

[9] R. Tsai. An efficient and accurate camera calibration technique for 3d machine vision. In *CVPR*, pages 364–374, 1986.

[10] Z. Zhang. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *IEEE Trans. Pattern Analysis Machine Intelligence (PAMI)*, 22(11):1330–1334, Nov 2000.