2012 IEEE International Conference on Systems, Man, and Cybernetics
October 14-17, 2012, COEX, Seoul, Korea

# Color and Illumination Invariant Dice Recognition

Gee-Sern Hsu, Hsiao-Chia Peng, Shang-Min Yeh
Artificial Vision Lab., National Taiwan University of Science and Technology
Taipei, Taiwan

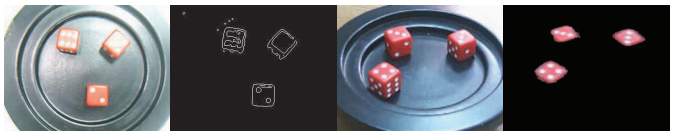
***Abstract*—A system is proposed for automatic reading of the number of dots on dice in general table game settings. Different from previous dice recognition systems which recognize dice of a specific color using a single top-view camera in an enclosure with controlled settings, the proposed one uses multiple cameras to recognize dice of various colors posed in a wide range of viewing angle and under uncontrolled conditions. It is composed of three modules. Module-1 locates the dice using the gradient-conditioned color segmentation (GCCS), proposed in this paper, to segment dice of arbitrary colors from the background. Module-2 exploits the local invariant features good for building homographies across multiple views and lighting conditions. The homographies are used to enhance coplanar features and weaken non-coplanar features, giving a solution to segment the top faces of the dice and make up the features ruined by possible specular reflection. To identify the dots on the segmented top faces, an MSER detector is embedded in Module-3 for its consistency in locating the dot regions regardless of illumination and viewpoint variations. Experiments show that the proposed system performs satisfactorily in various test conditions.**

***Index Terms*—Object recognition, foreground segmentation, invariant feature, local descriptor.**

## I. Introduction

Dice is a common table game in casinos, and most popular in the casinos in Asia. As automatic or computer-controlled games are becoming popular, many are interested in the technologies able to assist or even replace human bankers. A computer vision system is proposed in this paper for *dice recognition*, which refers to the automatic reading of the numbers of dots on dice, in generic table game settings.

Different from existing dice recognition systems [1], [2], [3], which use a single top-view camera and only work in an enclosure with controlled settings, the proposed system exploits stereo view and works in general uncontrolled settings, including dice of various colors, under different illumination conditions, and with various camera viewpoints. In controlled settings edges are the prominent features considered for dice recognition. But specular reflection, often observed in un-controlled illumination, make the approaches solely based on edges ineffective. Fig.1(a) shows a case with specular reflection and its edge map in Fig.1(b), along with the results obtained by the proposed system in Fig.1(c) and 1(d). In addition to the incapability of recognizing dice under various illumination conditions, the requirement that only the top surfaces of the dice can be considered also refrains the existing dice recognition systems from applications in general conditions. To capture the top surfaces, the camera must be made aligned to the dice. However it can be very difficult to maintain such an alignment in a table game, as the banker may roll the dice and place the dice roller back to a spot misaligned to the camera.



(a) (b) (c) (d)

Fig. 1: (a) shows a top view affected by specular reflection, with its edge map on (b); (c) shows a tilted view with its stacked image (d).

To have a broader application scope, the proposed system also considers various viewpoints to the dice captured by multiple cameras. Possible applications include the integration with table games and some new design of computer-controlled dice games. When integrated with table games, it can help or replace human bankers in offering services, such as counting and collecting wagers. Because non-top views reveal both the top and side surfaces of the dice, it is required to segment the top surfaces and remove or ignore the side surfaces. To meet this requirement, the proposed system exploits stereo vision with invariant features to extract the top surfaces and identify the dots on them.

The proposed system consists of three modules for dice detection, top surface segmentation and dots identification. To detect the dice, a scheme called gradient-conditioned color segmentation (GCCS) is proposed to locate and segment the dice of arbitrary colors from the background. The GCCS implements multi-scaled scanning windows running through the given image to search for the local regions with dense gradients, which are caused by dice edges, dot edges, and edges on the roller base. These edges can be categorized according to different orientation patterns. During the search the part of the background region without gradients is also identified. Given the local regions close to the dot edges and to the background without gradients, a classifier can be trained to distinguish the foreground, i.e., dice, from the background. Local invariant features are then applied on the dice regions for top surface segmentation. Eight local invariant feature detectors are evaluated in the performance in rendering reliable homographies across multi-views and various illumination conditions. The homographies are used to enhance the coplanar features on the top surfaces, leading to a solution good for top surface segmentation. To identify the dots on the segmented top surface, an MSER (Maximally Stable Extreme Region) [4] detector is applied for its consistency in rendering stable local features across illumination and viewpoint variations. Heuristic

978-1-4673-1714-6/12/$31.00 ©2012 IEEE

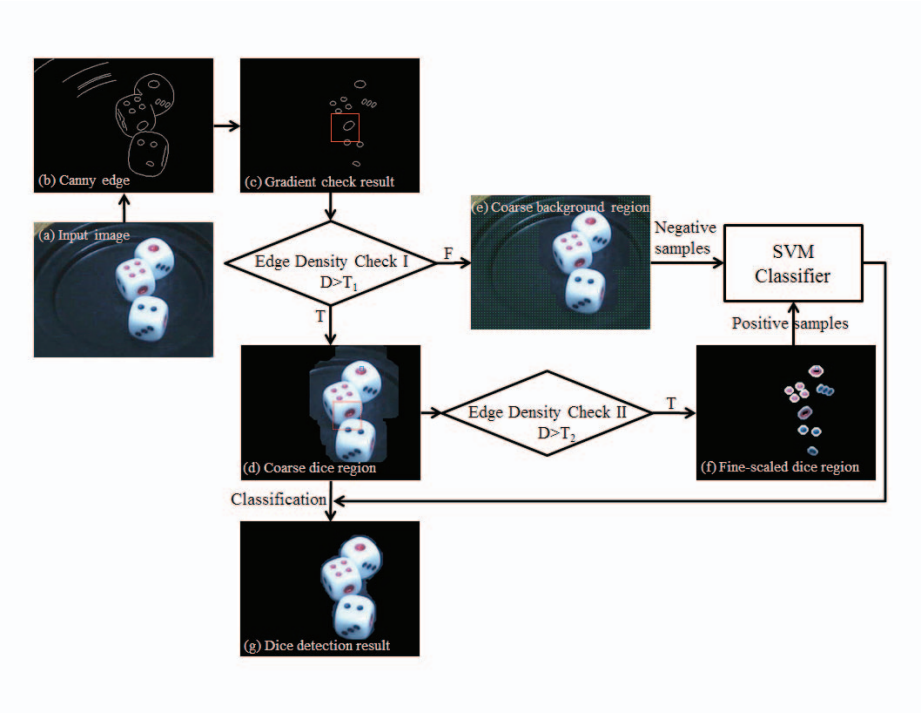



Fig. 2: HOG-based edge selection and training sample collection in GCCS (Gradient-Conditioned Color Segmentation).

rules on the spatial distribution of the dots are exploited to determine the number of dots on each of the dice.

The structure of this paper is organized as follows: dice detection by GCCS is presented in Section II, followed by the top surface segmentation in Section III, and then the dot identification in Section IV. An experimental study is presented in Section V, and a conclusion is given in Section VI.

## II. Dice detection by gradient-conditioned color segmentation

Because dice of arbitrary colors must be detected, we consider the segmentation using the gradients obtained from gray-scaled images. Although one can consider classifiers, such as that proposed by Viola and Jones [5], we are only interested in the approaches without the need of collecting a large number of training samples. Given an image processed by Canny edge filter, the proposed gradient-conditioned color segmentation (GCCS) is composed of the following steps:

- Obtain the 8-bins histogram of oriented gradients (HOG) [6] on each edge segment, and remove the edge segments with strongly unbalanced HOG. The strongly unbalanced HOG refers to the case that at least one out of the eight orientation bins in the HOG is larger than 20% of the mean. Because edge segments from the boundaries of dice and dots often give balanced or weakly unbalanced HOG, this step removes most of the edge segments from non-dice areas, and keeps the edge segments from the dice and dots, as shown in Fig.2(c).
- Apply a large scanning window to locate the areas of high gradient density as the *coarse* dice regions, and the areas of low gradient density as the coarse background regions (shown in Fig.2(d) and (e) with Edge Density Check I). A fine-scaled scanning window is then applied to the coarse dice regions only to refine the selection of regions of high gradient density, which are mostly the dice bodies around dots (shown in Fig.2(f) with Edge Density Check II).
- The fine-scaled dice regions and the coarse background regions offer two sets of pixel samples to train an SVM classifier with a linear kernel. Our experiments show that the linear kernel performs equally well as other types of kernels, but with a faster training speed. The SVM classifier is then applied on a given test image to distinguish dice from the background, as the result shown in Fig. 2(g). Note that the above Steps 1 and 2 are only performed at initialization, or when the dice are changed to different colors.

An experimental study on the segmentation performance across different color spaces shows that the HSV color components are the most appropriate features for the SVM classifier. We have compared the performance of the above GCCS to other approaches, it outperforms Mean-Shift and watershed segmentation when handling different colors of dice, illumination and viewpoint variations. It is believed that the GCCS can be extended to segment foregrounds from backgrounds when both have different gradient and color features.

## III. Top Surface Segmentation

Given the dice detected and segmented from the background, stereo vision with local affine invariant features are exploited to extract the top surfaces of the dice. Several affine

invariant features are compared in the performance rendering homographies robust to illumination and viewpoint variations.

### A. Homography Determined by Local Invariant Features

Many local invariant feature detectors were proposed in the last decade and applied in a broad range of applications. Reviews on these detectors can be found in [7], [8] and [9]. The invariant feature detectors can be generally categorized into three types [10]. One detects corner-like features, e.g., Harris-affine, Harris-Laplace, and multi-scale Harris corner detectors. One detects blob-like features, e.g., Hessian-affine, Hessian-Laplace, multi-scale Hessian and Difference of Gaussians (DoG) [11]. The multi-scale Harris-Hessian is a complementary detector which extracts both the corner-like and blob-like features. Different from the former two types, region detectors extract homogeneous areas, for example, the MSER detector [4] detects regions invariant to illumination variation with arbitrary contours.

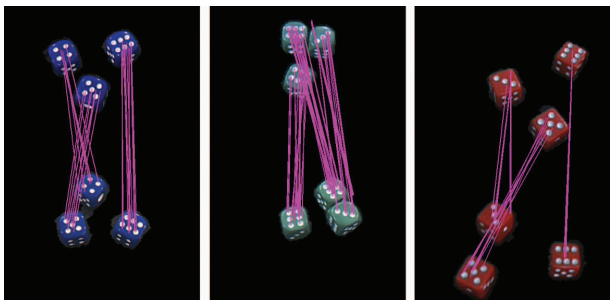


Fig. 3: Correspondences across two different views on the local invariant features detected by a multi-scale Harris-Hessian detector after dice detection.

Due to the limitation of Harris and Hessian detectors in handling multiple scales, both are modified with multiple scales and made scale-invariant in [12]. To determine the most appropriate scale for a local feature, Harris-Laplace and Hessian-Laplace both search for the characteristic scale with a Laplace operator added on top of the multi-scales. Harris-affine and Hessian-affine obtain the affine invariant corners or blobs by an iterative estimation of elliptical affine regions proposed by Lindeberg et al. [13]. The shape of the feature region is adapted to ensure that the same region is covered when extracted from a different viewpoint.

The performance of the aforementioned feature detectors in rendering accurate homographies between different viewpoints is evaluated by a comparison to the ground truth obtained manually. All of the features are represented in the form of the descriptor given in [11] as it is proven as one of the most effective descriptors in an extensive experimental study [7]. The match of the invariant features across different views is measured by the Euclidean distance between the feature descriptors, and a threshold on this distance measure is determined to select correspondences. Because a single dot on a die in a given view can appear quite similar to a different dot in another view, the scale factor in each feature detector must be first determined so that the correspondences of the features can be defined for local regions with multiple dots or even dots on multiple dice. We selected the scale factor as that comes with the maximum number of correct correspondences. RANSAC [14] is then applied to filter out outliers and determine the most appropriate homographies across different views with matched correspondences. Our experiments reveal that the multi-scale Harris-Hessian detector gives the best performance. Details are reported in Section V. Fig.3 shows an example of the correspondences across viewpoints.

### B. Segmentation by Multi-view Stacking

Given $N$ different viewpoints of dice images, $N(N-1)/2$ homographies can be obtained using the invariant feature correspondences. In most cases $2 \leq N \leq 4$ suffices. Each homography and its inverse define the transformation between a pair of different viewpoints, and such a transformation only works for the top surfaces of the dice as these surfaces are *coplanar*. This property motivates the stacking of coplanar surfaces to segment the top faces of the dice even when specular reflection appears in certain viewpoints. One can choose a dice image of any viewpoint as a reference image and transform the rest $N-1$ images of different viewpoints to the reference one using the corresponding homographies. Stacking of the reference image and $N-1$ transformed images does not just enhance the coplanar features but also weaken the non-coplanar features, as those on the lateral sides of the dice would be overlapped with features from different planes. As the specular reflection can be considered a view-dependent feature, different from the coplanar features observed from other views, it can be removed by imposing a threshold on a similarity measure.

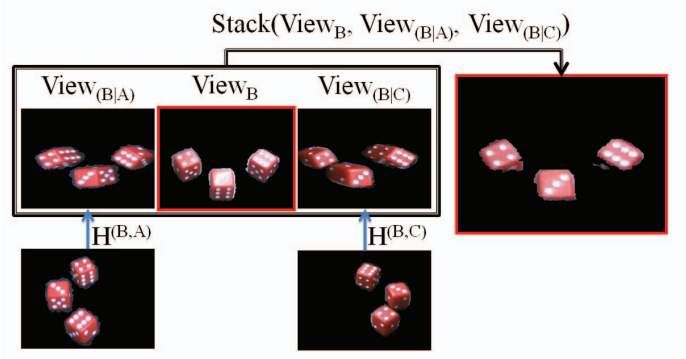
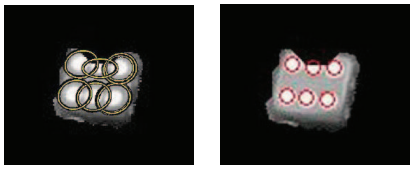



Fig. 4: Flowchart of the stacking process that enhances coplanar features and weakens non-coplanar features, giving a way for the extraction of the top faces of the dice.

An example with $N = 3$ on red dice is given and the details are illustrated by the flowchart in Fig.4, where $\text{View}_A$, $\text{View}_B$ and $\text{View}_C$ are the dice images from three viewpoints of the cameras. Assume that the homographies $\mathbf{H}^{(B,A)}$ and $\mathbf{H}^{(B,C)}$ transform $\text{View}_A$ to $\text{View}_B$ and $\text{View}_C$ to $\text{View}_B$, resulting in $\text{View}_{(B|A)}$ and $\text{View}_{(B|C)}$, respectively. In the right of Fig.4 shows the stacked image of $\text{View}_B$, $\text{View}_{(B|A)}$ and $\text{View}_{(B|C)}$ followed by post-process

Fig. 5: Two types of dice with five colors and varying sizes.



(a) MSER detected. (b) Merging of overlapped MSERs.

Fig. 6: MSER can detect incomplete or partial dot regions.

procedure labeled $\mathbf{Stack}(\text{View}_B, \text{View}_{(B|A)}, \text{View}_{(B|C)})$. One can obtain $\mathbf{Stack}(\text{View}_A, \text{View}_{(A|C)}, \text{View}_{(A|B)})$ and $\mathbf{Stack}(\text{View}_C, \text{View}_{(C|B)}, \text{View}_{(C|A)})$ in the same way.

## IV. Dot Identification and Dice Recognition

Given the top surfaces segmented from the dice, an MSER detector [4] is applied to identify the dots on them because of its stability in rendering persistent or slowly varying edges around the dots as illumination varies. The dots on dice are blob-like objects of similar size and on which the MSER usually anchors on the boundaries of such objects. where $t$ is the threshold. An MSER is a connected region in $E_t(\mathbf{x})$, with little change in its size for a range of thresholds $[t_a, t_b]$, extracted with a watershed like segmentation algorithm. The homogeneous intensity regions extracted are stable over a wide range of thresholds. Since the MSER only captures intensity-homogeneous regions, regardless of their contours, it can detect incomplete or partial dot regions, as shown in Fig.6.

The dots captured by the MSER are clustered by $k$-means ($k$ happens to be the number of dice). Depending on the number of dots in a given cluster, the spatial distribution pattern for that number is examined with size constraint. The size constraint refers to the fact that the size of each cluster must be between the maximum and minimum top surface of a die. The spatial distribution patterns can include the following: 6-dot must contain two parallel rows of dots and 3 dots each row. 5-dot must have two crossing rows of dots, 3 dots each row and crossing each other at the same central dot. 4-dot must have two parallel rows of dots and 2 dots each row. 3-dot is a row of 3 dots with an equal distance in between, and 2-dot has 2 dots parted by a distance close to the diagonal of the top surface. If found incompatible to the above spatial patterns, two possibilities would be verified. One is a non-dot spot falsely considered as a dot and the other is a valid dot failed to be identified as a dot. An extensive experimental study presented in Section V reveals that the combination of size-constrained clustering and spatial pattern confirmation yields a superb recognition performance.

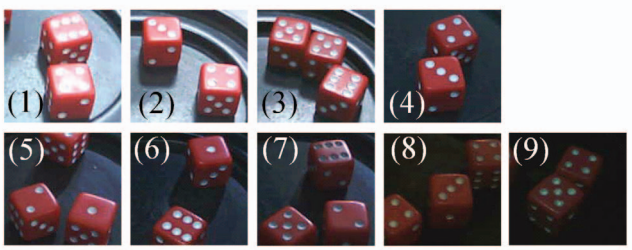



Fig. 7: Test set with 9 illumination conditions with viewing angle = 45°.

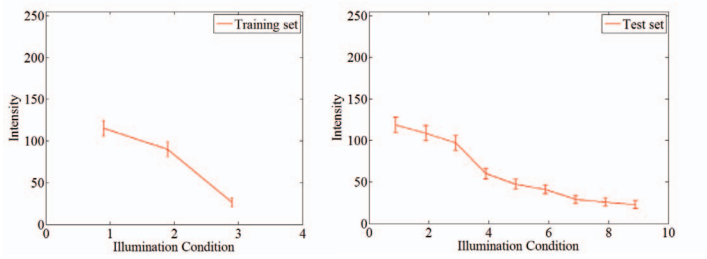



Fig. 8: 8-bit gray scale mean and standard deviation in 3 training conditions (left) and 9 testing conditions (right). Illumination conditions are shown in indices.

## V. Experiments

The experiments were designed to study the impacts made by the following variables on the dice detection rate, homography rendering accuracy and dice recognition rate.

1) Dice color: Two types of dice with five different colors were selected because they were popular in common table games, as those shown in Fig.5. Type I, known as the western style, has four colors, namely red, blue, green and brown with white dots on each die. Type II, known as the Asian style, has colors white and green, with dots in red, black and white.
2) Illumination conditions: 12 different conditions were selected that covered a wide scope of lighting intensity and directions. Three of them were used for training and the rest nine for testing. Fig.7 shows the test set under 9 illumination variation. The average intensity of red dice in training and test set ranges from 22 to 118 in the 8-bit scale. Fig.8 shows the average intensity and deviation for the two sets.
3) Camera viewpoints: three viewing angles, $\theta = 30°$, $45°$, $60°$, measured by the tilt from the horizon, are considered.

Our scenario follows Sci Bo, a popular dice game in Asia and US, where three dice are cast in each session. For each combination of the above variables, e.g., red dice with 60° camera viewpoint and under illumination condition 7, we ran 300 sessions in which 250 were random cast and 50 manually

placed. The manual placement was intended to imitate special cases, including two or three dice connected together and three dice with same outcome.

### A. Dice Detection Performance

The detection rates on five dice colors and under nine illumination conditions are given in Fig.9 (top), and the average is 98.73%. White and blue dice have the highest detection rates, followed by brown, red and green. The lowest detection rates on almost all dice colors occurred with Illumination-9, which is the darkest in our lighting settings, but still above 95%. This shows that the performance of the gradient-based GCCS only degrades marginally as the lighting intensity is substantially reduced. The significant drop on red dice at Illumination-7 was caused by the shadows inside the dots, and the shadow and the dice body had similar intensity values, as shown in Fig.10(d).

The GCCS appears to be robust to camera viewpoints, as shown in Fig.9 (bottom), where the best are still on the white and blue dice. Table I summarizes the overall detection rates associated with each different dice color. We also ran hundreds

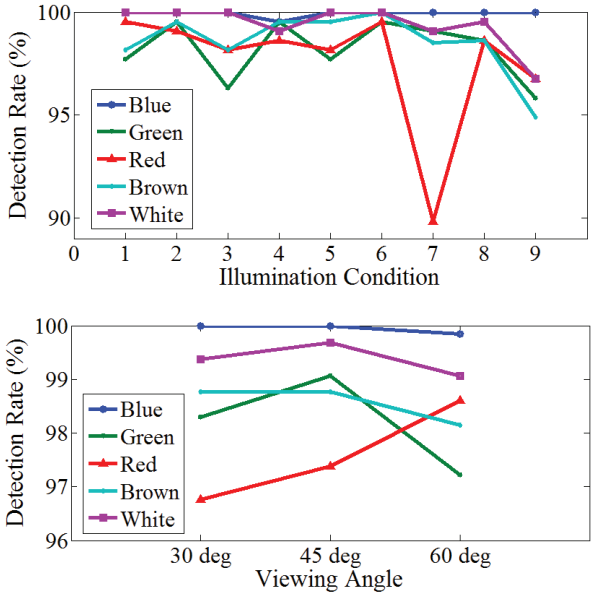



Fig. 9: The performance of the GCCS for dice detection. Detection rates in 9 illumination (top) and 3 camera viewing angles (bottom).

TABLE I: Average detection rates on dice of five different colors

| Dice Color | White | Red | Brown | Green | Blue |
|---|---|---|---|---|---|
| Accuracy (%) | 99.38 | 97.58 | 98.56 | 98.20 | 99.95 |

of sessions with multiple colors of dice, and a few typical segmentation results are given in Fig.10(a-c), with average detection rate 98.58%.

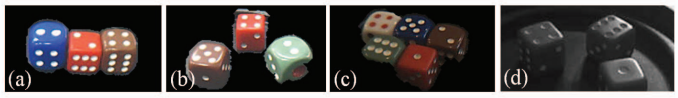



Fig. 10: (a-c) Performance on GCCS with different colors in every throw. (d) Red dice has lower detection rate on illumination condition 7 while the shadow inside the dots share similar intensity values with dice body.

### B. Invariant Features for Establishing Stereo Correspondences and Homographies

Because red dice gave the lowest detection rates, they were chosen in this experimental study for the determination of the invariant features best for setting up stereo correspondences and homographies. The invariant features were evaluated on $E_{F_i}$, which was the difference between the correspondences from the invariant-feature-based homography $\mathbf{H}_{F_i}$ and the ground-truth $\mathbf{H}_G$ obtained using manually selected correspondences, i.e.,

$$E_{F_i}^{(a,b)} = \frac{||(\mathbf{H}_{F_i}^{(a,b)} - \mathbf{H}_G^{(a,b)})\mathbf{x}_{F_i}^b||}{N_{F_i}} \quad (1)$$

where $\mathbf{H}_{F_i}^{(a,b)}$ is the homography that transforms the invariant features $\mathbf{x}_{F_i}^b$ detected by the invariant feature detector $F_i$ in the image $I_b$ to the corresponding ones in $I_a$; $\mathbf{H}_G$ is the ground-truth homography obtained by manual selected correspondences between $I_a$ are $I_b$, $N_{F_i}$ is the number of features detected by $F_i$, and $a, b$ denote two different viewpoints.

Additionally, it is also desired that the correspondences from the feature-based homographies can be consistent across different scales, as some features change with scales. To investigate what features were better than others in rendering desired homographies across illumination and scale, the original images in $320 \times 240$ pixels were scaled down to smaller sizes, and the error was computed in each size and averaged over the three illumination conditions in the training set. Fig.11 shows this comparison with viewing angle $60^o$, the smallest scale with $128 \times 96$ reveals relatively high errors, indicating that some details between the dice were lost in such a small scale and thus the accuracy in the homography estimation degraded.

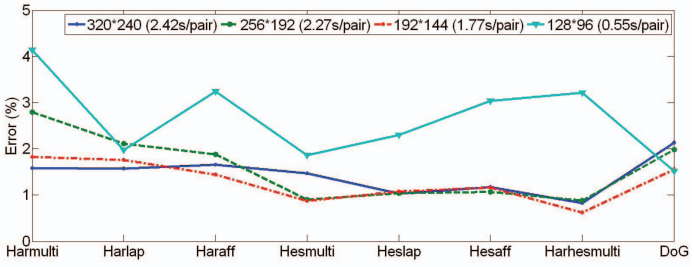



Fig. 11: Normalized error of feature-based homography across scales and three illumination conditions, with camera viewing angle $60^o$

Although many affine invariant detectors can handle the projection deformation caused by different viewpoints, it relies on an assumption that the camera is relatively far from the object, making the affine invariance applicable. In our tests the cameras were not far enough from the objects (dice), a few non-coplanar features were mistaken as coplanar ones by the affine invariant detectors, and therefore they were outperformed by other invariant detectors. Among the eight invariant feature detectors we tested, the multi-scale Harris-Hessian detector gave the lowest error at $0.87\%$, and it was about $1.7$ pixels in a $192 \times 144$ image. Our experiments also showed that the correspondence error could reach as high as $10\%$ when the viewing angle decreased to $45^o$, and $30\%$ when it reached $30^o$. This indicates that the stereo correspondences solely based on invariant features may be valid for viewing angle larger than $60°$.

### C. Dots Identification and Dice Recognition

The performance evaluation on the test sets with viewing angle $> 60^o$ reveals the following observations and results:

- As long as the correspondences from the feature-based homography are consistent over at least two scales, the average match error can be kept below or near $1\%$, and the top surfaces of dice can be accurately segmented in all tested conditions with different dice colors and illumination conditions.
- Illumination appears to be the major parameter that affect the recognition performance. Two identification rates are measured in each test illumination condition, one is the identification of the dots and the other is the identification of the dot number on each die as shown in Fig.12. Most show $> 99.5\%$ in dot identification rate. The illumination conditions 1 and 9, which are with the strongest and weakest intensity, appear with low rates on both dot identification and dice recognition, because both were with many blurred boundaries around the detected MSERs. The poorest performance, however, is given by illumination 7, which casts strong shadows in many dots. Strong shadows inside the dots blur the boundary between dots and specular reflection and degrade the performance of the MSER detection.
- The combination of size-constrained clustering and spatial pattern confirmation can effectively remove the false positives and yield accurate dice recognition rates in all tested conditions, as shown by the right bar at each indexed illumination in Fig.12.

## VI. Conclusion

A solution composed of three modules is proposed for dice recognition. The core part of Module-1 is the GCCS for dice detection and segmentation, which is proven effective in capturing dice of different colors, illumination conditions and viewing angles. Module-2 exploits the multi-scale Harris Hessian detector, which is experimentally validated to outperform many in rendering correct homographies under various test conditions and parameters. Module-3 exploits the fact that

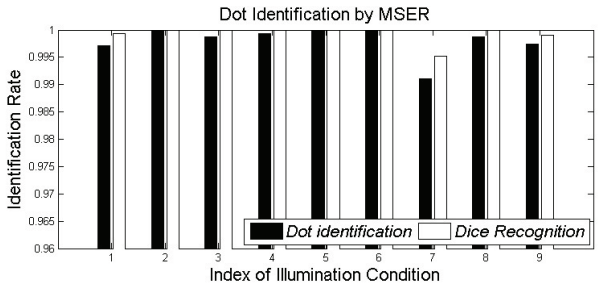



Fig. 12: Identification rates in 9 illumination conditions with camera viewing angle $60^o$. At each index the black bar on the left shows the rate of dot identification, and the white bar on the right shows the rate of dice number identification.

dots are made with high contrast on the surface of a die, and applies the MSER as the dot detector. Experiments reveal that, although false positives of dots are observed in few cases, as under strong or insufficient lighting conditions, the numbers of the dots on the dice can still be recognized accurately by the proposed solution, as long as the camera viewing angle is kept less than $60^o$. The next phase of this research would be the search for the boundary on the viewing angle, beyond which automatic stere correspondence can be hardly succeeded, and can only be performed manually.

## References


[1] Kuo-Yi Huang, "An auto-recognizing system for dice games using a modified unsupervised grey clustering algorithm," *Sensors*, 2008.

[2] Wen-Yuan Chen and Chin-Ho Chung, "Image automatic-recognition scheme for dice game using structure features technique," *Optical Engineering*, 2011.

[3] Y.-N. Lai, S.-T Hsu, C.-Y Wang, and M.-T Tsai, "Method for recognizing dice dots," U.S. Patent No. 2009/0263008 A1, October 2009.

[4] Jiri Matas, Ondrej Chum, Martin Urban, and Tomás Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in *British MachineVision Conference*, 2002.

[5] Paul A. Viola and Michael J. Jones, "Rapid object detection using a boosted cascade of simple features," in *CVPR (1)*, 2001, pp. 511–518.

[6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *CVPR (1)*, 2005, pp. 886–893.

[7] Krystian Mikolajczyk and Cordelia Schmid, "A performance evaluation of local descriptors," *PAMI*, vol. 27, no. 10, pp. 1615–1630, 2005.

[8] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A comparison of affine region detectors," *IJCV*, vol. 65, no. 1-2, pp. 43–72, 2005.

[9] Gee-Sern Jison Hsu, *Advances in Theory and Applications of Stereo Vision*, chapter 7: Stereo Correspondence with Local Descriptors for Object Recognition, pp. 129–150, InTech, 2011.

[10] Tinne Tuytelaars and Krystian Mikolajczyk, "Local invariant feature detectors: A survey," *Foundations and Trends in Computer Graphics and Vision*, vol. 3, no. 3, pp. 177–280, 2007.

[11] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, vol. 60, no. 2, pp. 91–110, 2004.

[12] Yves Dufournaud, Cordelia Schmid, and Radu Horaud, "Matching images with different resolutions," in *CVPR*, 2000, pp. 1612–1618.

[13] Tony Lindeberg and Jonas Gårding, "Shape-adapted smoothing in estimation of 3-d shape cues from affine deformations of local 2-d brightness structure," *Image Vision Comput.*, 1997.

[14] Martin A. Fischler and Robert C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, 1981.