Colour key-point detection

Hermine Chatoux, Noël Richard, Bruno Mercier; XLIM - University of Poitiers; Poitiers, France

Abstract

A lot of image processing tasks require key-point detection. If grey-level approach are numerous, colour and hyper-spectral ones are scarce. In this paper, we propose a generic key-point detection for colour, multi and hyper-spectral images. A new synthetic database is created to compare key-point detection approaches. Our method improves detection when the image complexity increases.

Introduction

Key-point detection is a recurrent step in image processing for image retrieval, registration, 3D modelling, etc. A key-point is a particular point or area. Identifying the right key-point is crucial to characterize a part of the image content. For the moment, the key-point performances are assessed only from multimedia databases without considering the chromatic frequencies of the images.

Currently, key-point detectors are developed for grey-level images or use a marginal approach in different colour spaces, therefore losing the channel inter-correlation. In this work, we propose a new colour key-point detector with a vectorial approach for the colour information. We want a detector that respects the physics of acquisition to obtain a measurement.

The next section proposes a brief state of the art underlining the limits of the existing approaches. Then, we present the Full-Vector-Key-Point detector explaining the novelty or necessity of each step. The third section presents the synthetic database used for the experiment presented in the fourth section. We assess the key-point extraction performance with a colour complexity scale thanks to a new database. Finally, we conclude.

The State of the Art

Detectors are classified in different categories: corners or blobs. The corner type (Fig. 1) focuses on extracting shapes like corner, cross or 'T'. Blob detector define a key-point as the centre of a uniform area contrasting with the background.

Corner Detector

One branch of the corner detector family is based on the proposition by Harris and Stephens [1]. It is based on the spatial auto-correlation matrix of the partial derivatives. Harris and

Figure 1. *Example of key-point detection.*

Stephens compare both eigenvalues to extract key-point: a corner is defined by two large eigenvalues. The corner scale is unknown prior analysis, Mikolajczyk and Schmid [2] proposed a multi-scale key-point detector.

In [3], Montesinos *et al.* proposed a colour adaptation based on the Di Zenzo gradient [4]. This expression consider the *RGB* colour space as orthogonal, hence, is similar to a marginal approach. Sebe *et al.* [5] or Everts *et al.* [6] transform the images in other colour space (colour invariants or colour opponent) prior to the detection.

The second branch of the corner family looks for pattern indicating a corner as presented by Smith and Brady [7]. The idea is to compare pixel values from the edge to the central pixel in an interest area. Mair *et al.* [8] adapt the comparison to speed up the detection. Their detector, named AGAST, compare the centre to the four cardinal extrema of the area.

A colour adaptation from Ishikura *et al.* [9] extracts key-points marginally in CIELAB colour space. Rassem and Khoo [10] measure quantified colour histogram to extract key-point. However, quantification lose information and so does not respect physical measurement.

Blob Detector

The extrema of the second derivative correspond to the centre of a uniform area. This is the idea beyond the blob family. The first proposition from Lindeberg [11] uses the Laplacian of Gaussian (LoG) to measure the centre to extract. This writing can be applied easily for a multi-scale analysis. To speed up computation time, Lowe [12] proposes to use Difference of Gaussian (DoG). Bay *et al.* [13] (Haar) extract key-points based on wavelet decomposition.

Laplacian of Gaussian have been tested marginally on *RGB* images by Corso and Hager [14] and Unnikrichnan and Hebert [15]. Tahery and Drew [16] and Khanina *et al.* [17] propose a colour adaptation of the Hessian approach. Everts *et al.* [6] use the Gabor filters marginally. These propositions are marginal or image content dependent which is not consistent with metrology.

Full-Vector-Key-Point Detector

We wish to produce a measure for key-point detection. To that end, we need to choose a colour space. Quantification is prohibited. To respect the physical properties, a vectorial construction is necessary.

This *Full-Vector-Key-Point* (FVKP) detector is based on the same steps as the Harris *et al.* [1] corner detector. Figure 2 presents these steps which are separated in three phases. First, the gradient needs to be measured, then the corner informations are extracted and finally the key-point decision is made.

Gradient Extraction

As an image contains corners at different scales, we used multi-scale derivatives filters. Mikolajczyk *et al.* [2] proposed multi-scale binomial derivative filters that are approximation to Gaussian filters. To be more generic, we construct our filter with a Gaussian and its derivative.
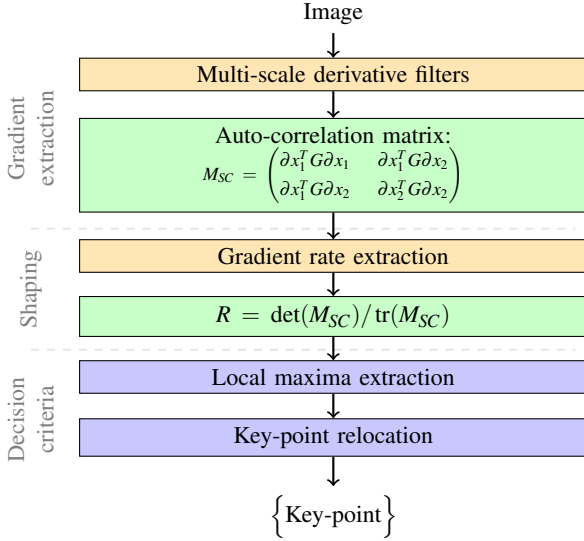
Image

↓

Gradient extraction	Multi-scale derivative filters

Auto-correlation matrix:
$$M_{SC} = \begin{pmatrix} \partial x_1^T G \partial x_1 & \partial x_1^T G \partial x_2 \\ \partial x_1^T G \partial x_2 & \partial x_2^T G \partial x_2 \end{pmatrix}$$

Shaping

Gradient rate extraction

$$R = \det(M_{SC}) / \operatorname{tr}(M_{SC})$$

Decision criteria

Local maxima extraction

Key-point relocation

↓

$\left\{ \text{Key-point} \right\}$

Figure 2. *Steps to obtain full-vector-key-point detector.*

These filters allows to measure, marginally, the gradient on every channel. The next step is to combine these gradients. Inspired from Di Zenzo [4], we proposed the *Full-Vector-Gradient* in [18] to extract gradients in the sensor space. We calculate a spatial correlation matrix (M_{SC}) depending on all the colour channels and their inter-correlations from the image I derivative along both axis x_i at a position x:

$$M_{SC} = \begin{pmatrix} \left\| \frac{\partial I(x)}{\partial x_1} \right\|^2 & \left\langle \frac{\partial I(x)}{\partial x_1}, \frac{\partial I(x)}{\partial x_2} \right\rangle \\ \left\langle \frac{\partial I(x)}{\partial x_1}, \frac{\partial I(x)}{\partial x_2} \right\rangle & \left\| \frac{\partial I(x)}{\partial x_2} \right\|^2 \end{pmatrix}, \quad (1)$$

with
$$\left\langle \frac{\partial I(x)}{\partial x_1}, \frac{\partial I(x)}{\partial x_2} \right\rangle = \frac{\partial I(x)}{\partial x_1}^T \cdot Gm \cdot \frac{\partial I(x)}{\partial x_2}, \quad (2)$$

$$Gm = \begin{pmatrix} \|s_0\|_2^2 & \langle s_0, s_1 \rangle_2 & \dots & \langle s_0, s_m \rangle_2 \\ \langle s_1, s_0 \rangle_2 & \|s_1\|_2^2 & \dots & \langle s_1, s_m \rangle_2 \\ \vdots & \vdots & \ddots & \vdots \\ \langle s_m, s_0 \rangle_2 & \dots & \langle s_m, s_{m-1} \rangle_2 & \|s_m\|_2^2 \end{pmatrix} . (3)$$

The Gram matrix Gm uses the scalar products $\langle s_i, s_j \rangle$ defined for the integrable functions (Eq.(4)). The functions correspond to the Spectral Sensitivity Functions (SSF) of each channel.

$$\langle s_i, s_j \rangle = \int_{\mathbb{R}} S_i(\lambda) S_j(\lambda) d\lambda. \quad (4)$$

These Equations are generic whatever the channel count.

Shaping

Before the response measure, an intermediate step is necessary to extract the strongest gradients represented by corners. The gradient rate extraction consist in a simple integrative filter. A corner area contains more edges than the same area crossed by a linear edge. Therefore after the integrative filter, the response from the corner area will be stronger than the edge one. This step is not about de-noising as it is often considered in other algorithms ([1, 2]).

The measure of the response we choose follows the principle given by [1]: both eigenvalues have to be strong to represent a corner. The response depends on both matrix invariants: the determinant (det) and the trace (tr). Harris and Stephens proposed

to estimate the response value with $R = \det(M_{SC}) - k \operatorname{tr}(M_{SC})^2$, k being an empirical constant. To free ourselves from the constant, we choose a new response:

$$R = \frac{\det(M_{SC})}{\operatorname{tr}(M_{SC})}. \quad (5)$$

Decision Criteria

Corners correspond to local maxima with high responses. Once a key-point has been selected, several informations are attached to it apart from its location: the response value and the orientation of the eigenvector associated to the strongest eigen-value.

By choosing a constant step between the derivative filters, we can follow the key-point shift of localisation with the filter size. This shift is a consequence of the gradient rate extraction step. Therefore we can relocate the large key-point to their original location. A corner can be detected at several scales. We hypothesize that the higher number of scale, the stronger the corner is. With the relocation, we have access to the number of scale where the key-point was detected. We choose to keep only key-point detected at three scales or more. The choice of three is consistent with other approaches comparing a scale with its previous and next.

Proposed Protocol

Mikolajczyk and Schmid [19] proposed evaluation criteria to assess the quality of a key-point detector. The first criteria is very dependent of the number of key-point detected: the more key-points, the better results are. The database used is constituted of real images with deformation from the original view. The images belong to the multimedia domain and has a low colour complexity.

Our goal is to assess the added value of colour in the detector. Therefore, we propose a synthetic colour images database. We choose to develop this assessment in the trichromatic colour domain, even if the generic writings allows it to be used in a multi or hyper spectral case.

The created database is similar to the one developed for the colour gradient assessment [18]. We assume it is easy to perform a corner detection if both regions are uniform, even with a low colour difference between both areas. But if the regions are
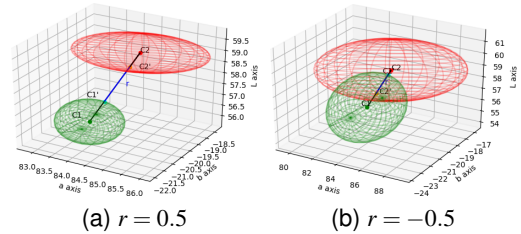

(a) $r = 0.5$ (b) $r = -0.5$

Figure 3. *Illustration of the separability rate (r) depending on the overlap between two colour distributions in CIELAB.*

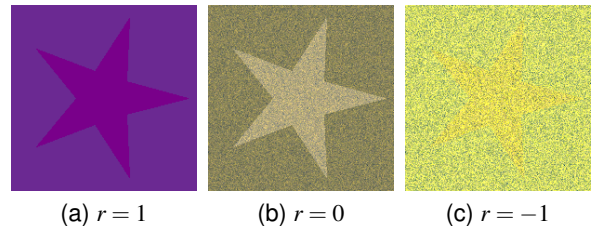(a) $r = 1$ (b) $r = 0$ (c) $r = -1$

Figure 4. *Example of synthetic images to detect key-point used to assess performances ($r = 1$: easy; $r = -1$: difficult).*

not perfectly uniform, they are defined by their respective colour distributions. The central idea of this dataset is to consider each side of a shape as a colour distribution. The corner extraction complexity is relative to the intersection of the two colour distributions (Fig. 3). The more they overlap, the more difficult it is for the corner to be extracted.

The proposed protocol is based on two $3D$ Gaussian distributions. Then, the colour average and the separability rate r define the colour distributions (Figure 3). When r decreases, the distributions overlap more, therefore the complexity increases. For $r = 1$, the background and foreground are uniform, we say it's without texture complexity.

The objective is to detect corner, therefore we choose a five branch star shape. One colour distribution is the background, the other is the foreground. We expect 10 key-points, one for branch extremities and one between each branch. A star shape offers the advantage of different orientations for the key-points. The Figure 4 presents three images with different complexity depending on the separability rate or the location of the selected colour pairs.

To assess the quality of the corner detection, we consider key-point as a binary classification test. A key-point found at a given distance of its true location is a true positive (TP). A key-point detected too far from a corner is a false positive (FP). A corner that is not detected is a false negative (FN). Then we measure precision and recall to estimate the detection quality:

$$Precision = \frac{\#TP}{\#TP + \#FP}, \quad (6)$$

$$Recall = \frac{\#TP}{\#TP + \#FN} \quad (7)$$

where # is the cardinal of the set.

Results and Discussion

We compare our detector (FVKP) with three state of the art approach: two from the blob family (DoG and Haar) and one using pattern (AGAST). For these methods, we used the *OpenCV* implementation with default parameters. To compare colour detectors, we have extracted key-points marginally from each *RGB* channel. The combination of the different channel responses is processed by keeping only one occurrence of key-point detected on several channels.

Prior to precision/recall results, we have to choose a distance for the distribution average and the maximum distance for a key-point to be considered as a true positive.

Choosing a Colour Distance

For this selection we use a simple image (similar to the case from Fig. 4b) where $r = 1$ for a hundred pairs of colours per ΔE. the Figure 5 present the percentage of image where key-point are detected.

The detection threshold depends on the response function maxima, therefore when $\Delta E > 0$ our approach (FVKP) detects key-points in every images. For the other key-point detector, the threshold is a parameter to adjust, so no key-point is detected for colour difference below 3. Above 3, only some images allow the extraction of key-point. When the luminance difference is high, grey-level approaches are efficient to extract key-points. But, even for larger $\Delta E = 15$, 20% of images have no key-point extracted from the DoG approach. An adjustment of parameters might change the results.

The validity of ΔE is below 20. To maximise the number of images with detected key-point for grey-level approach, we choose $\Delta E = 15$.
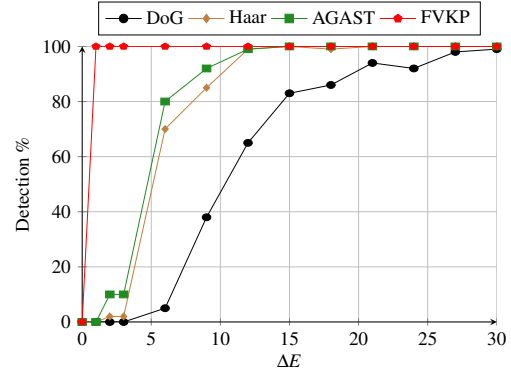

Figure 5. *key-point detection depending on the colour difference without texture complexity.*

Choosing the Distance to True Key-Point

To evaluate the detectors quality, we wish to count the key-point detected at a given distance from a true key-point. To select one, we compute the number of TP at a detection distance D_{max}. The Figure 6 gives the recall according to D_{max}. This is the average of 100 images without texture complexity. Only images with detected key-points are considered for the state of the art approaches.

The pattern detector (AGAST) is the most efficient with all key-points detected at less then 2.5 pixels of the theoretic key-points. It is true for around 99% images while our method finds key-points at a maximum of 4 pixels for all images. The Haar detector finds easily the first five key-points, the last five need a larger distance. It indicates only 5 key-points are detected. When the detection distance increases, the inner and outer branch belong to the same detection area. The other two approaches respond with an increasing distance for each new key-point. Only one detected key-point could explain such a results.

To be able to compare all approaches we select a medium distance allowing all detectors to extract key-points. We choose $D_{max} = 30$. It corresponds to 15% of the image width and is inferior to the distance between the inner and outer branches which exclude a key-point to be counted twice as true positive. A smaller distance would drastically decrease the results from the DoG approach which one of the most used actually.

Complexity Impact on Precision and Recall

In the previous experiments, we focuses only on the true positive without information on the false positive. For the chosen colour distance $\Delta E = 15$ and detection distance $D_{max} = 30$,
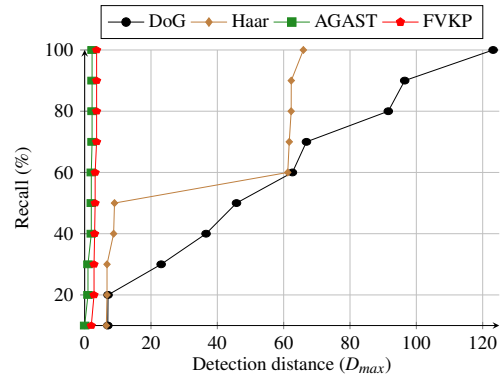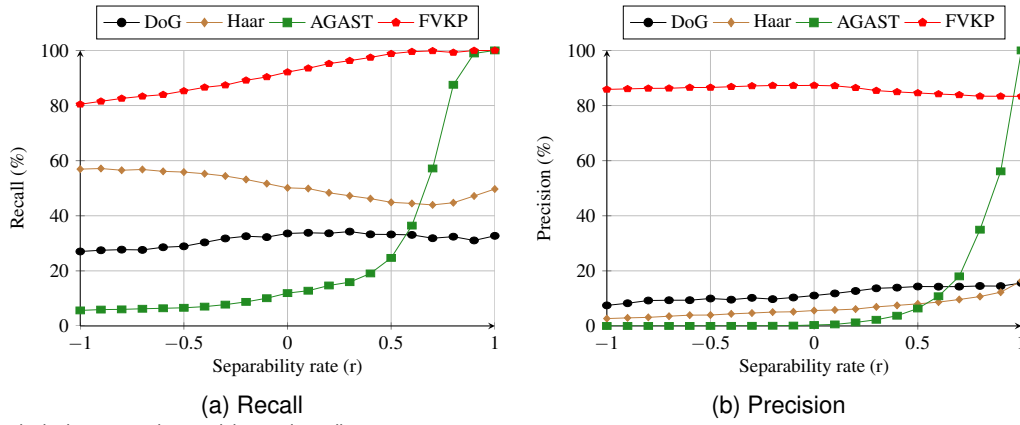

Figure 6. *Recall depending on the detection distance without texture complexity.*

Figure 7. *Complexity impact on the precision and recall.*

we measure precision and recall presented in Equations (6) and (7). 400 images are simulated, the results from Figure 7 are the average of images where key-point were detected. The measure of precision and recall depend on the separability rate: a measure of the image colour complexity.

The Figure 7a presents the number of key-point detected closely to a true key-point in regard to key-points to detect. The FVKP has a good recall whatever the complexity: 100% for simple images, 80% in the more complex one. AGAST shows good results for high separability rate but crashes when the colour complexity increases. The other detectors presents the same tendency for any complexity, around 30% for DoG and around 50% for Haar.

The Figure 7b shows the number of key-point detected closely to a true key-point in regard to the detected key-points. Precision decreases with the number of key-points when this detected key-point number is superior to the key-point number to detect. The key-point number detected by the AGAST detector increases rapidly with the complexity explaining the fall when $r < 1$. The noisy aspect of the images reduces the ability of the AGAST detector to find patterns. The number of key-point detected by DoG and Haar increases slightly with complexity inducing a small decrease of the precision with complexity. Overall, these two detectors are impervious to colour complexity. By their nature, they are not adapted to detect corners which can explain their poor results. The colour detector keeps a high precision above 80% whatever the complexity. Our method preserves only key-points detected at several scales. It reduces the number of detected key-points, keeping the strongest ones even when the complexity increases.

The FVKP detector gives better results whatever the complexity for both precision and recall. Moreover, our approach detects key-point for every images. Indeed, these results show the ability of the AGAST approach to find patterns in non-noised cases. By their nature, DoG and Haar (blob type detector) are not adapted to detect corners which can explain their poor results.

Conclusion

After showing the limits of previous key-point detectors, we proposed a new approach. It considers the sensors characteristics, and respects the metrology constraint. It is generic and can be applied to any sensors given its characteristics (colour, multi, hyper-spectral).

We proposed a new synthetic database to measure the added value of our proposition. The database is designed to extend the limits of data-driven approach. The key-point detectors are assessed in regard to the spatio-chromatic complexity of the image content.

Based on a precision/recall analysis we have shown our method is more robust to change in colour complexity compared to state of the art approaches.

References

[1] C. G. Harris and M. Stephens, "A combined corner and edge detector." in *Alvey vision conference*, vol. 15, no. 50. Citeseer, 1988, pp. 10–5244.

[2] K. Mikolajczyk and C. Schmid, "Indexing based on scale invariant interest points," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, vol. 1. IEEE, 2001, pp. 525–531.

[3] P. Montesinos, V. Gouet, and R. Deriche, "Differential invariants for color images," in *Proceedings. Fourteenth International Conference on Pattern Recognition (Cat. No. 98EX170)*, vol. 1. IEEE, 1998, pp. 838–840.

[4] S. Di Zenzo, "A note on the gradient of a multi-image," *Computer vision, graphics, and image processing*, vol. 33, no. 1, pp. 116–125, 1986.

[5] N. Sebe, T. Gevers, J. Van de Weijer, and S. Dijkstra, "Corner detectors for affine invariant salient regions: is color important?" in *International Conference on Image and Video Retrieval*. Springer, 2006, pp. 61–71.

[6] I. Everts, J. C. Van Gemert, and T. Gevers, "Evaluation of color spatio-temporal interest points for human action recognition," *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1569–1580, 2014.

[7] S. M. Smith and J. M. Brady, "SUSAN: a new approach to low level image processing," *International journal of computer vision*, vol. 23, no. 1, pp. 45–78, 1997.

[8] E. Mair, G. D. Hager, D. Burschka, M. Suppa, and G. Hirzinger, "Adaptive and generic corner detection based on the accelerated segment test," in *Computer Vision ECCV 2010*. Springer, 2010, pp. 183–196.

[9] K. Ishikura, N. Kurita, D. M. Chandler, and G. Ohashi, "Saliency detection based on multiscale extrema of local perceptual color differences," *IEEE Transactions on Image Processing*, vol. 27, no. 2, pp. 703–717, 2018.

[10] T. H. Rassem and B. E. Khoo, "Performance evaluation of new colour histogram-based interest point detectors," *Multimedia Tools and Applications*, vol. 74, no. 24, pp. 11 357–11 398, 2015.

[11] T. Lindeberg, "Feature detection with automatic scale selection," *International journal of computer vision*, vol. 30, no. 2, pp. 79–116, 1998.

[12] D. G. Lowe, "Object recognition from local scale-invariant features," in *Computer vision, 1999. The proceedings of the seventh*

IEEE international conference on, vol. 2. Ieee, 1999, pp. 1150–1157.

[13] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, Jun. 2008.

[14] J. J. Corso and G. D. Hager, "Coherent regions for concise and stable image description," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2. IEEE, 2005, pp. 184–190.

[15] R. Unnikrishnan and M. Hebert, "Extracting scale and illuminant invariant regions through color." in *BMVC*, 2006, pp. 499–508.

[16] S. Tahery and M. S. Drew, "A novel colour hessian and its applications," *Electronic Imaging*, vol. 2017, no. 18, pp. 171–176, 2017.

[17] N. Khanina, E. Semeikina, and D. Yurin, "Scale-space color blob and ridge detection," *Pattern Recognition and Image Analysis*, vol. 22, no. 1, pp. 221–227, 2012.

[18] H. Chatoux, N. Richard, F. Lecellier, and C. Fernandez-Maloigne, "Full-vector gradient for multi-spectral or multivariate images," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2228–2241, May 2019.

[19] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.

Author Biography

Hermine Chatoux *obtained her PhD in image processing at the University of Poitiers (France - 2019). She, currently, works at the XLIM institute. She is specialized in colour science. Her research activities requires to merge mathematics and physics to develop metrological colour image processing.*

Noël Richard *is an associate professor at XLIM laboratory. He is the technical chair of the CIE TC8.14 working on the definition and assessment of the spatio-chromatic complexity. He developed a new image processing based on distance functions for metrological purposes extended to hyper-spectral images for full-band vector processing.*

Bruno Mercier *received Ph.D. degree in Computer Science from University of Poitiers (France - 2006) on relighting real objects from photographs: he developed the complete chain. He is currently a Research Engineer at XLIM lab and his main research interests include computer vision, industrial vision and 3D visualization.*