# AUTO CROPPING FOR DIGITAL PHOTOGRAPHS

*Mingju Zhang*[1,2] *Lei Zhang*[1] *Yanfeng Sun*[1] *Lin Feng*[2] *Weiying Ma*[1]

[1] Microsoft Research Asia, Beijing(100080), P.R.China
[2] Institute of University Students' Innovation, Dalian Univ. of Tech., Dalian (116024), P.R.China

## ABSTRACT

*In this paper, we propose an effective approach to the nearly untouched problem, still photograph auto cropping, which is one of the important features to automatically enhance photographs. To obtain an optimal result, we first formulate auto cropping as an optimization problem by defining an energy function, which consists of three sub models: composition sub model, conservative sub model and penalty sub model. Then, particle swarm optimization (PSO) is employed to obtain the optimal solution by maximizing the objective function. Experimental results and user studies over hundreds of photographs show that the proposed approach is effective and accurate in most cases, and can be used in many practical multimedia applications.*

## 1. INTRODUCTION

With the rapid development of digital cameras and scanners, family photographs are being accumulated rapidly and automated tools for organizing and editing these photographs become extremely desirable. In such applications, automated photo cropping is one of the important features to automatically enhance photographs. Auto cropping targets at seeking an optimal rectangle for printing or auto-printing task (to adapt a photograph to a specified paper), auto slideshow task (to automatically generate panning, zoom in and zoom out effects), and photograph enhancement task (to get a better composition). Such applications all require to crop photographs into a specified aspect ratio. Though general composition templates and rules have been thoroughly discussed in photography theory in many textbooks or tutorials on the web, automatic solution is still lacked. As it is generally a very difficult problem to automatically give a good looking cropping result, most commercial software either simply output a cropping rectangle from the center of the photograph, which may ignore important subjects or even destroy the photos, or leave the tedious task to users.

To obtain an optimal composition given a specified cropping aspect ratio, image content analysis needs to be conducted to analyze which regions are salient or attractive. For home photographs, faces are usually the focuses in photographs and can be automatically detected by practical face detection algorithms [1]. However, due to the large variations in pose and illumination existing in home photographs, not all faces could be detected. Also, many scenery photographs do not contain any face. To complement this drawback, attention model [2] based on contrast provides a great support to the applications. Based on psychology principles, attention model gives a way to detect salient subjects (attended areas) other than faces in photographs, and it also provides a way to know the most informative region (attended view) in a photograph. Though the attended areas do not reflect semantics, which is still an open problem of image understanding, faces and attended areas do provide rich information in image composition.

Based on such information, Liu *et al* proposed a smart browsing method which aims at browsing large pictures in a small screen. However, the goal of this method is to generate a dynamic browsing path to browse subjects in a picture dynamically. Thus in most cases, it puts the subject in the center of the view rectangle (cropped rectangle in our application), which is against photography composition rules for still photographs [3]. In the field of robot-photographer [4], templates with photography rules are employed, and the robot-photographer takes photos in the composition of such templates. It is indeed a similar work, but an important difference is that it is to create photographs other than to deal with existing photographs. The latter task needs to comprehend what the photo-takers want to show in a photograph, and thus is much more difficult. As well, the robot-photographer only works with photographs with people.

Generally, the requirement of auto cropping is to crop a photograph into a specified aspect ratio, improve the composition of the photograph if possible, and give different solutions for different conservative coefficients. There are some widely accepted rules in photography composition, such as rule-of-thirds, empty space, no-

[1] This work was performed when the author was visiting Microsoft Research Asia.

middle [5,6], which give the base rules for auto cropping. But to the best of our knowledge, we didn't see any work being done to use them automatically into the photograph auto cropping.

In this paper, we propose an auto cropping model and formulate the auto cropping problem to an optimization problem, in which the objective function is defined as the sum of three sub model energies, i.e. energy of composition sub model, energy of conservative sub-model and energy of penalty sub model. Particle swarm optimization (PSO) is then utilized to obtain the optimal solution by maximizing the objective function [7]. The candidate solution which maximizes the objective function will be the final cropping result. The proposed auto cropping model has several advantages. Firstly, the conservative sub-model can avoid the face missing and the limitation of the attention model to some extent. Secondly, photography rules and templates are explicitly utilized, which could lead to an artistic cropping result, as well as improve the composition of the original photographs in many cases. Additionally, it can deal with photographs without faces. Experimental results and user studies over hundreds of photographs show that the auto cropping results are satisfactory in most cases and are effective for printing, slideshow and photo enhancement applications.

## 2. AUTO CROPPING

In this section, we present the auto-cropping model, which defines an objective function to obtain an optimal cropping solution. The input to the model is the width, height of the original picture, the conservative coefficient, the faces detected by face detection [1] and the region of interest (ROI) result detected by attention model [1] [2]. The model consists of three sub models, i.e. composition, conservative and penalty. PSO is then employed to seek the optimal solution by maximizing the objective function. The objective function is formally defined as follows:

$$E(R_c) = E_c(R_c) + E_s(R_c) - E_p(R_c) \quad (1)$$

where $R_c$ is the cropped rectangle. $E_c$ indicates the composition sub model to describe how good the composition of the cropping solution is. $E_s$ indicates the conservative sub model to prevent the photograph from being cropped too aggressively. The bigger $E_s$ is, the smaller the effect of the composition sub model is. And $E_p$ is a penalty factor to prevent faces or ROIs being cut off.

### 2.1 Composition sub model

Usually, photographers implicitly or explicitly utilize different templates when they are dealing with different subjects. For example, when they are dealing with a portrait photograph, the face may need to be at the center of the photograph. Or, when they are dealing with some single person photograph cases, they may use the *rule-of-thirds* [5] to put the body at the vertical line of one thirds.

To deal with general cases, in this paper, we propose 14 templates, which are, **IP** (identification passport photograph), **BPT** (big face portrait), **SPT** (small face portrait), **TB** (two big faces), **TM** (two mid-size faces), **TS** (two small faces), **MB** (3-5 small faces), **MM** (3-5 mid-sized faces), **MS** (3-5 small faces), **CB** (>5 faces, big *human area*), **CS** (>5 faces small *human area*), **SC** (scene, no face and attended areas), **SCA** (no face, one attended area) and **SCM** (no face, >1 attended areas). Just as their names imply, we can categorize photographs according to the face count, attended area count and several thresholds separating small, mid-size and big faces.

Three types of rectangles are used for crowd group photographs, namely, *face area*, *body area* and *human area*. *Face area* is the minimum rectangle containing all faces. *Body area* is the extension of the *face rectangle*. And the *human area* is calculated by the attended view. These areas are illustrated in Figure 1.

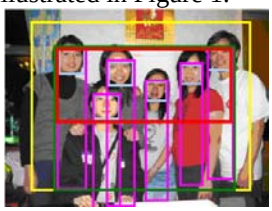

**Figure 1. Special rectangles:** light blue rectangles, red rectangle, pink rectangles, yellow rectangle and green rectangle respectively indicate the face rectangles, face area, body rectangles and human area.

Reference points are defined to describe the rules for templates, as shown in figure 2. $P_A$ , $P_B$ , $P_C$ , $P_D$ and $P_E$ are the cross points of the middle lines, golden division lines and lines of thirds of cropped rectangle $R_c$.

For each template in 2.1, if the main subject center $P_M$ is near to the reference points $P_{cp}$, $E_C$ will get a bigger value. But because of the limited information of the input data, we must also consider a conservative point, $P_{csv}$.

For single face pictures IP, BPT and SPT, $P_M$ is the center of the only face rectangle. For 2-5 faces pictures TM, TS, TB, MM, MS, MB, CB and CS, $P_M$ is the center of face area. $P_M$ is ignored for SCM and SCA. $P_V$ is the center of the attended view.

Reference points $P_{cp}$s are the most important points for the templates. They are regarded as the best location for $P_M$. For IP, TB, MB, $P_{cp}$ is $P_A$, the center of $R_c$, which means that if the subject(s) is (are) large, the center of the face rectangle should be close to the center of the cropped rectangle. $P_{cp}$ for SC is also $P_A$ . $P_{cp}$ for BPT, TM is $P_B$, $P_{cp}$ for SPT and TS is $P_C$. $P_{cp}$ for MM and CB is $P_E$. $P_D$s are reversed points for some out-of-rule conditions.

As the ROI detection results do not really reflect semantics which are very important in photographs, simply using the common photography rules [5,6] may destroy

photographs aggressively in many cases. To prevent the photo-taker's original composition from being ignored, $P_{csv}$ is created for conservative. For BPT, TM, TS, CB, CS and SCA, $P_{csv}$ is $P_E$, and $P_A$ for others, which means that if the subjects are close to the center, give them the trend to go to the center.

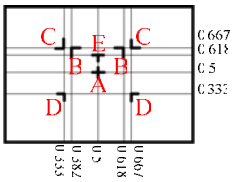


**Figure 2. the Reference Points**

The $E_C$ for each template has a uniformed equation:

$$E_1(R_c) = \alpha \cdot \max(\exp(-\frac{\|P_{cp} - P_M\|}{d}), \exp(-\frac{\|P_{csv} - P_M\|}{d})) \quad (2)$$

$$E_2(R_c) = \frac{w}{n_a} \sum_{areas} \exp(-\frac{1}{d} \min_L \|L_{thirds} - P_{area}\|) \quad (3)$$

$$E_c(R_c) = \mu_1 \cdot E_1(R_c) + \mu_2 \cdot E_2(R_c) \quad (4)$$

where $E_1$ indicates the energy for the main subject of the photo, so $\mu_1$ is set to 1 when the photo has main subject such as face or human area, otherwise $\mu_1$ is set to 0. Similarly, $E_2$ indicates the energy for the assistant subject of the photo, so $\mu_2$ is set to 1 when the photo has attended area(s), otherwise $\mu_2$ is set to 0. $d$ is the length of the cropped rectangle's diagonal. $\alpha$ is an empirically determined constant. Because the confidence of the assistant subject detected by ROI is not very high, $w$ is a small empirically determined coefficient. $L_{thirds}$ are the horizontal and vertical lines located at one-third position in the cropped rectangle. Note that $P_{cp}$, $P_{csv}$, $L_{thirds}$ depend on $R_c$.

Intuitively, Eq.2 means that if the main subject is close to either the reference point or the conservative point, the energy will be larger. And if the assistant subject is close to the line-of-thirds, the energy will be larger. Main subject has more effect then the assistant subject. This sub model generally makes the solution to be more artistic.

### 2.2 Conservative sub model

Though the composition sub model provides an approach to making the result artistic, there is still risk to destroy the photograph composition. Because the former sub model may result in a small cropped rectangle, we need to adopt some methods to prevent the photograph from being cropped too aggressively. The area of cropped rectangle and the distance from the center of cropped rectangle to the center of the attended view are good terms to constrain the result. Eq.5 gives the formal definition of the conservative sub model.

$$E_s(R_c) = \phi \cdot ((\frac{\beta}{S_v} + \frac{\lambda}{S_t}) * (S_i + S_c) + \eta \cdot \exp(-\|P_c - P_v\|)) \quad (5)$$

where $S_v$ is the area of the attended view, $S_t$ is the total area of the photograph, $S_i$ is the intersection of cropped rectangle and the attended view, and $S_c$ is the area of cropped rectangle. $\beta$, $\lambda$, $\eta$ are empirically determined constants. $\phi$ is the conservative coefficient.

Eq.5 means that, if the intersection of cropped rectangle and the attended view is larger, the energy will be larger; If the area of cropped rectangle is larger, the energy will be larger. If the center of cropped rectangle, $P_c$ , is close to the center of the attended view, $P_v$, the energy will be larger. $\phi$ is an input parameter to weight the effect of the conservative sub model in Eq. 5.

### 2.3 Penalty sub model

To prevent faces from being cut off in the final cropped rectangle, the penalty factor $E_p$ is defined as follows.

$$E_p(R_c) = \sigma \cdot \sum_{f \in F} (S(R_f \cap \overline{R_c}) / S_{total}) \quad (6)$$

where $R_f$ is the one of face rectangles in the photograph. $S(.)$ tells the area of the input rectangle. $\sigma$ is an empirically determined constant to weight the penalty energy. This sub model makes the result more reasonable by forcing the cropped rectangle to contain more information (faces).

### 2.4 Optimization

The cropped rectangle is represented as (left, top, width, height). Given a specified aspect ratio, the solution can be reformed into a vector of ($l$, $t$, $w$) ($h$ can be calculated by $w$ and the aspect ratio), the optimal solution can be obtained by globally searching the maximum candidate of Eq. 1 in the solution space. In this paper, we adopt PSO to seek the optimal solution.

In a PSO system, multiple candidate solutions coexist and collaborate simultaneously [7]. Each solution candidate, called a 'particle', flies in the problem search space, looking for the optimal solutions. A particle adjusts its position according to its own 'experience' and the whole system's 'experience'. So PSO system combines local search methods with global search methods. A particle's status is recorded by two factors: the position and the velocity. A standard PSO system is described as updating each particle's status constantly until the system converges to a stable point or iterates for a specified iteration count.

The global search looks like a time consuming process. But for auto cropping problem, the solution space is not very complicated, and the size of the cropped rectangle is usually not very small. Therefore the solution space can be efficiently limited. On a common PC with a Pentium IV 1.8G CPU, it typically takes 0.013s for PSO to converge and output the optimal result on average.

## 3. EXPERIMENTS

Auto cropping result is generally very subjective and is difficult to objectively evaluate the performance. Thus we conducted two user studies to evaluate our work. 100 pictures randomly selected from 600 home photographs are used in the study. Some cropping examples are shown in Figure 3.

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

**Figure 3. Some examples of auto cropping.** (a) The original photographs, (b) the photographs imposed with face and ROI detection results, (c) the inputs to the auto-cropping algorithm, (d) the cropped rectangles in the original regions, (e) the cropped photographs.

Five users are invited to participate in the user studies. The first user study is to evaluate the auto cropping result in different aspect ratios. And the second user study is to evaluate the improvement of the picture composition after cropping. An example result of different aspect ratios is illustrated in figure 4.

(a) (b) (c) (d) (e) (f)

**Figure 4. Cropping result for different aspect ratios**
(a) original photo, (b) 2:1, (c) 4:3, (d) 1:1, (e) 3:4, (f) 2:3

In the first user study, all users are required to mark "good", "acceptable" or "bad" to the auto-cropping results in 5 different aspect ratios and 2 conservative coefficients to the same photograph. Table 1 shows the result.

**Table 1. Auto cropping result evaluation 1**

| Cons. Coef. | Good | Acceptable | Bad |
|---|---|---|---|
| 0.0 | 157 (31.4%) | 273 (54.6%) | 70 (14%) |
| 1.0 | 51 (14.2%) | 436 (83.2%) | 13 (2.6%) |

In the first user study, we can see that the algorithm exhibits a satisfactory score of cropping a photograph into different aspect ratios. Also, it is shown that a large conservative coefficient leads to a small bad score. Figure 5 shows an example result for different conservative coefficients.

(a) (b) (c)

**Figure 5. Different conservative coefficient.** (a) the original photo, (b) cons. coef .= 0.0, (c) cons. coef. = 1.0

In the second user study, all users are required to mark "better" "acceptable" or "worse" to auto cropping results in the same aspect ratio as that of the original pictures, which can evaluate the improvement of the picture composition after auto cropping. Table 2 shows the evaluation result.

**Table 2. Auto cropping result evaluation 2**

| | Better | Acceptable | Worse |
|---|---|---|---|
| **Auto-cropping** | 41(41%) | 43(43%) | 16(16%) |

Also we can see that in the second user study, the considering of the artistic rules leads to a good score of the improvement of the picture composition.

## 4. CONCLUSION

We have presented a photograph auto cropping algorithm in this paper. Taking face detection result and ROI detection result as input, 14 templates are proposed to explicitly utilize artistic composition rules. To obtain an optimal result, we formulated auto cropping as an optimization problem by defining an energy function for each template, and then use PSO to obtain the best solution. It can be used to output a good cropping rectangle with different aspect ratios in printing and auto slideshow applications, or to enhance pictures by improve their compositions. Experimental results show that the proposed approach is effective and accurate in most cases, and can be used in many practical multimedia applications.

## 5. REFERENCES


[1] Xiao, R., Li, M.J., Zhang, H.J., Robust Multi-Pose Face Detection in Images. *IEEE Trans. On CSVT, Special Issue on Biometrics*, 2003

[2] Y.F. Ma, Zhang, H.J., Contrast-based Image Attention Analysis by Using Fuzzy Growing, *ACM Multimedia 2003, Berkeley, USA*, pp. 374-381, 2003.

[3] Liu, H., Xie X., Ma, W.Y., Zhang, H.J., Automatic Browsing of Large Pictures on Mobile Devices, *11th ACM International Conference on Multimedia, Berkeley, CA, USA*, 2003.

[4] Z. Byers, et al., An Autonomous Robot Photographer *International Conference on Intelligent Robots and Systems*, 2003.

[5] T. Grill and M. Scanlon, *Photographics Composition*, American Photographics Book Publishing, 1990.

[6] B. Gooch, et al., "Artistic composition for image creation" *Eurographics Workshop on Rendering*, 2001.

[7] Kennedy J, Eberhart R, Particle swarm optimization, *IEEE Int Conf on Neural networks, Perth, Australia*, 1995.